THE

GILBANE
REPORT ™

on Open Information & Document Systems

OBJECT-ORIENTED DOCUMENT DATABASE SYSTEMS — WHAT ARE THE BENEFITS?

In a previous issue (Vol. 1, No. 6) we covered how document-centered interfaces and object-oriented programming affect document management implementation strategies. In that article we did not go into depth on how document management systems could use object-oriented database technology — partly because there was a lot more talk than activity. Things have changed. In our feature article in this issue John Chelsom takes a detailed look at how object-database features can enhance document management.

WELCOME CALS JOURNAL SUBSCRIBERS

We were sad to hear the *CALS/Enterprise Integration Journal* would cease publication. The Journal was a highly professional and valuable publication that played a major role in the growth and understanding of CALS. We wish Jeff and Bill and their families the best — they deserve the gratitude of all involved in CALS. As part of an arrangement with the publishers, *CALS Journal* subscribers who have paid subscriptions will receive *The Gilbane Report* in place of the Journal for the remainder of their subscription period. While we are not a CALS publication, this is not an unlikely match — document management the way we view it is arguably the central focus of CALS to begin with. We will also expand our coverage of CALS activity. In our next issue we will provide some thoughts on the CALS market and the state of CALS. In the meantime, see our CALS Update on page 16.

CONTENTS

OBJECT-ORIENTED DOCUMENT DATABASE SYSTEMS — WHAT ARE THE BENEFITS?

EXECUTIVE SUMMARY

Strategic Overview

- Sales of object-oriented databases are expected to be in the multi-billion dollar range by the turn of the century.

- Documents fit well into the object-oriented paradigm.

- This year, object-oriented document management systems are hitting the market.

Objects, Documents and Databases

- Currently, almost all document management systems use a relational database even though documents do not fit well into the RDBMS view of the world as consisting of rows and columns.

- Documents lend themselves particularly well to the object-oriented model.

- An object is a data structure that has a set of properties and a set of behaviors. An object is an instance of an object type or class.

- A property or behavior can be defined once for the object type, and all instances inherit it.

- Readers "naturally" take an object-oriented view of documents, understanding them as being composed of distinct elements that can be easily modeled as object types.

Object-Oriented Database Management System (ODBMS)

- An ODBMS can address many of the issues important for managing documents.

- Document objects can be of arbitrary lengths and may contain many different types of data. An object-oriented document management system can handle this.

- It can also manage the relationships among the different document objects.

- An ODBMS uses internal pointers to objects. Different systems use different methods, which makes it impossible to share objects across systems (until the vendors adopt a standard such as CORBA).

- The ODBMS seems to be poised for the same sort of explosive growth witnessed by the relational database management vendors in the mid 1980s.

- For a relational database to model a document, it has to use non-intuitive workarounds, introducing overhead.

- The object-oriented model can handle all the data types and relationships required for a document system.

- It can track and manage documents to any level of granularity, including paragraphs, graphics, tables and even individual phrases and words.

- It can provide sophisticated controls for workgroups simultaneously working on a document.

- It can track versions not only of a document of each of its components.

- It can provide configuration management, understanding the interdependencies among all the objects it's tracking.

- It takes advantage of the client/server architecture.

- The user interface to an ODMS tends to be under-developed.
- An ODBMS provides the same sort of back-up and security features as its relational counterparts.

Documents and Information Management

- Document objects can be viewed within the wider context of an organization's information and workflow.
- An example is Numerical Algorithms Group which use an ODBMS to manage product information with a particular emphasis on the generation of product documentation.

Objects, SGML and The Net

- An ODBMS is a natural choice for implementing an SGML-based document management system because of SGML's natural object-orientation.
- SGML vendors are releasing repository systems based on ODBMS technology.
- The current products require a high degree of customization.
- By linking an ODBMS with a WWW server, users of WWW browsers can send Object Query Language queries across the Internet to find HTML documents.

Relational Database Management System (RDBMS) and ODBMS

- RDBMS vendors are finding ways to make their products more object-oriented.
- One approach is to extend the relational data model to handle objects as well.
- Another approach is to preserve the relational data model but put in an additional interface layer between it and the users.

Risks and Costs

- There are the risks associated with being a pioneer.
- There is a risk that prototype systems will not scale up.
- Poorly-implemented systems may have poor performance.
- It can be tricky to integrate an ODBMS with an existing relational database manager.
- There are costs in switching to an object-oriented authoring system, and in converting legacy data.
- Moving to an ODBMS may require extensive training.
- It may also require upgrading to client/server hardware.

Conclusion

- Object-oriented document management systems are the new generation and are likely to be here to stay.
- If you are thinking about implementing a document management system, you should be considering an ODBMS.

STRATEGIC OVERVIEW

In recent years, almost every area of information technology has felt the growing influence of object-oriented systems and thinking. Object technology has survived the early hype to emerge as a robust solution to real business needs and is bringing about a revolution in software applications for both users and developers. Global sales of object-oriented software are expected to be in the multiple billions of dollars by the turn of the century, with object-oriented databases accounting for almost perhaps a third of that.

Documents fit well with the object-oriented paradigm and many users of popular word processing and desktop publishing software have become accustomed to manipulating documents and document components (such as tables, lists and figures) as objects. However, mainstream document management systems have been built almost exclusively on top of relational database systems - despite the mismatch between real-world documents and the relational data model.

For several years there was no real alternative to using relational database management systems (RDBMS) for handling documents and when object-oriented databases began to appear in the late 1980s, many viewed the technology and the vendors who supplied it as unstable and unproven. Five years later, the picture has changed with the emergence of established vendors and products and the introduction of industry standards for object database management systems (ODBMS).

Against this background, recent months have seen the release of a clutch of new document management systems, based on ODBMS. With many similar products and applications in the pipeline, there are strong signs that the object revolution has come to the field of document management. This article examines the developments in object technology that have made this revolution possible and assesses the advantages and disadvantages of object-oriented document database systems.

OBJECTS, DOCUMENTS AND DATABASES

Currently, almost all document management systems that have a document repository at their heart use a relational database management system to manage the repository. An RDBMS thinks of all data as fitting into tables; the rows are the records and the columns are the fields. While this works extremely well where the data is highly regular (e.g., personnel records), documents do not map very well to the table model. First, the contents of documents are highly irregular; they are of varying lengths and of highly diverse types (i.e., text, graphics, multimedia elements). Second, the contents of documents have much more complex

ACME COMPUTER CO. - PERSONNEL

ID	Name	DOB	Sex
E123	James Burns	03.07.62	M
A927	Maggie may	12.10.68	F
D295	Ronnie Clark	10.04.76	M

OBJECT-ORIENTED DATABASES

by

Albert Collins

CONTENTS

Introduction

Mixing Data Types Numbers and Strings Text and Graphics

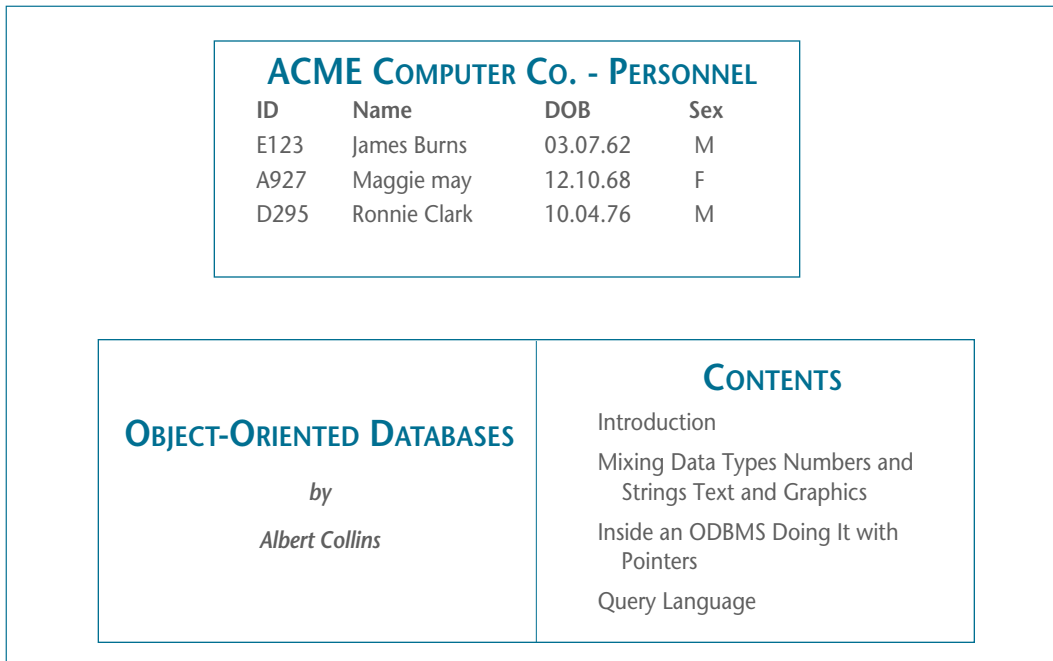Inside an ODBMS Doing It with Pointers

Query Language

Figure 1: Relational tables are good at modeling data such as personnel records (top), but not good at modeling the hierarchical structure of a document (bottom).

interrelationships than can be expressed easily in a two-dimensional table. Documents are, after all, hierarchical — they consist of a succession of nested chapters, sections, subsections, etc. — and a tree, not a table, is the natural way to express such a structure.

Introducing Objects

In the context of software engineering, an object is a data structure that has a state and a behavior. The state is determined by a set of properties and the behavior by a set of operations. Types of objects are defined such that objects of the same type have the same set of properties and operations; different instances of objects of the same type have different values for their properties.

Note: The terms 'method' and 'class' are commonly used in place of 'operation' and 'type', which are the preferred terms in the ODMG-93 standard (see below).

An example of an object type in a document management system might be a 'book' with the properties 'author', 'title' and 'chapters'. A particular book might have 'John Steinbeck' as the value of its author property, 'East of Eden' as its title and an appropriate set of chapter titles as the value of its chapters property. The book type could have a 'print' operation to generate a paginated hardcopy and a 'view' operation to display it on a computer monitor.

If we now create a new type of object — a chapter object — in addition to the book object, then the chapters property of the book could contain a list of chapter objects instead of chapter titles. This is an illustration of a special type of property that expresses a relationship between two object types (i.e. a book contains a set of chapters). Figure 2 shows some objects and relationships for a small document management system.

The 'is_a' relationship between the document type and the book, article and memo types is a special one that defines books, articles and memos as being subtypes of the document supertype. The subtypes can inherit properties and operations from the supertype. This means that properties such as 'author' and operations such as 'print' that apply to books, articles and memos could be defined once for the general document type and they would be automatically applied to (inherited by) all the document's subtypes.
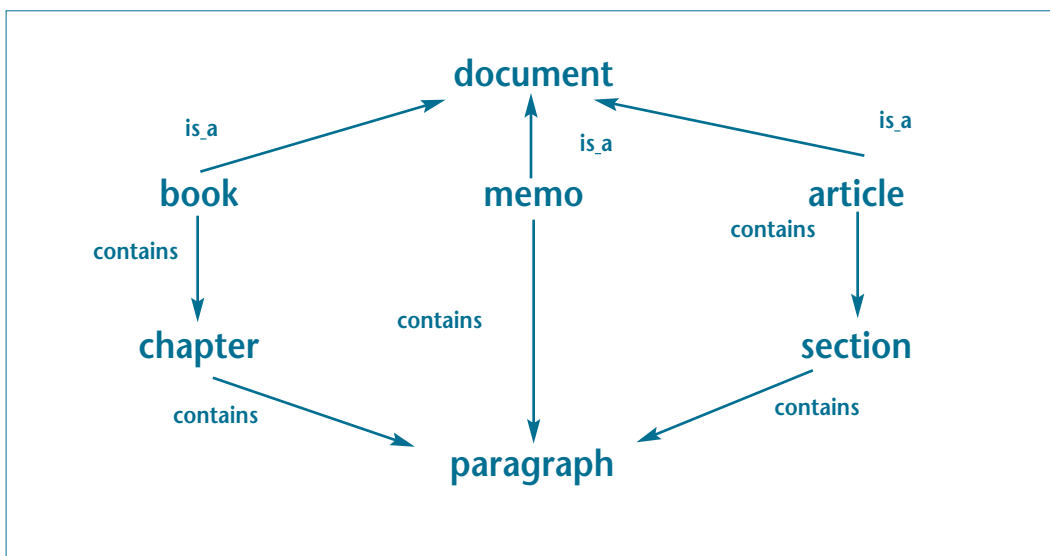

Figure 2: An example of objects and relationships in a small document management system.

Objects and Documents

It can be seen from the examples given above that documents lend themselves particularly well to the object-oriented model. A typical document is composed of a set of distinct elements such as sections, paragraphs, tables and figures that can be modeled easily as object types. The model fits with the reader's view of the document as a mixture of data elements of varying length and type. It also fits with the view of word processor users who expect the elements to have different operations associated with them: they expect to be able to adjust the line spacing of a paragraph but not of a graphic, and they expect to be able to pixel-edit a raster image but not a paragraph.

Object-Oriented Databases

ODBMS address many of the issues that are important for managing documents. Instead of storing data as tables, ODBMS store objects that can contain any arbitrary amount or type of data and can capture arbitrary and complex interrelationships. Because of this, ODBMS offer several key advantages when used to store document objects:

- They can manage not only documents but also the elements composing the documents and the links between them.

- They can manage unstructured data such as arbitrary lengths of text and graphics, and multimedia document components such as video and sound clips.

- They enable applications to share very rich data.

- By being aware of the structure of documents they can assist authors in creating documents to specification.

- Because they manage document components, complete documents can be assembled on the fly in response to user queries, thus enabling customized print-on-demand.

- They have a natural fit with SGML, the ISO standard for structured documents.

How Do ODBMS Work?

Applications written in object-oriented programming languages such as C++ or SmallTalk create and manipulate objects that exist in the computer's volatile main memory while the application is running; when the application stops, the objects cease to exist. An ODBMS provides the means of storing such objects as persistent data in the computer's permanent filestore and accessing them when applications are run. To do this, each object is assigned a unique identity or pointer. When the object is manipulated by the application at runtime the pointer indicates a location in main memory; when the application is not running the pointer indicates a location in the permanent filestore. A key feature of any ODBMS is the way in which object pointers are handled and mapped between main memory and permanent filestore locations. Each ODBMS product uses its own method and the performance of different products varies quite widely. Because of the different methods it is not possible to share objects between different ODBMS running on the same system, although this should become possible once vendors adopt the CORBA standard for identifying objects (see side bar on page 9).

It should now have become apparent that ODBMS products tend to be 'low level', in that they require applications to be programmed in languages such as C++ or Smalltalk. This fact, combined with the tendency of early products to use proprietary data models and interfaces to the programming languages, has impeded the adoption of ODBMS. However, these impediments have largely been removed with the upsurge of C++ as a programming language and the adoption of the ODMG- 93 standard, which introduces

an industry-wide data model and an SQL-like query language for ODBMS (see side bar on page 9).

A Growing Market

The ODBMS market, though still small in comparison with its relational counterpart, now appears to be exhibiting the same explosive growth that was witnessed by RDBMS vendors a decade ago. In addition, the market has begun to polarize around a small number of well-established vendors, as many of the early innovators have dropped away. Again, this mirrors the pattern seen previously in the RDBMS market.

Relational vs. Object-Oriented Document Models

The differences between the relational and object-oriented models of a document are illustrated by the example in Figure 3. This shows the tables of contents for two books which share one common chapter. The object model contains two types of object for books and chapters. The relational model contains two tables for the books and chapters and a third to model the relationship between them.

Not only is the relational model shown in Figure 3 less intuitive and succinct than the object-oriented model, it also introduces more overhead when the application is running. In the object-oriented model, the contents of a book are explicitly available in the set of chapters stored with the book object; in the relational model the contents must be found by executing a query.

O-O DOCUMENT MANAGEMENT FEATURES AND CAPABILITIES

Inherent in the design of most commercial ODBMS are many of the features required to implement a successful document management system. Consequently, many organizations are now looking towards ODBMS to implement their next generation of document management systems, either directly or through document database products built on top of ODBMS (see sidebar on page 9).

Managing Disparate Data Types

The object-oriented data model is a very rich one, capable of handling all the data types and relationships required for document management applications. Furthermore, an ODBMS can store these data types directly, without having special layers of application code for packing and unpacking permanently stored objects.

Object Granularity and Sharing

An object-oriented document management system can track and control document components — paragraphs, graphics, tables, and even individual phrases and words — as well as documents as a whole. Because links between objects are maintained as pointers, it's easy for multiple documents to share component objects. This not only reduces storage space but also makes it easy to update all documents containing an object when the object itself is updated.

Group Working and Access Control

If two or more users or processes are working with the same objects it's important to make sure that they don't try to edit an object simultaneously. This situation often arises if two authors are working on documents that contain shared elements. ODBMS solve such problems by creating read or write locks on data retrieved from the database server.

When specifying an ODBMS, it is important to check that the data locking mechanism is suitable for the intended application. Some systems lock pages of memory as they are

DOCUMENT MANAGEMENT SYSTEMS
by Peter Snell

1. What Are Documents?
2. Features of a DMS

OBJECT-ORIENTED DATABASES
by Shelly Keating

1. What Are ODBMS?
2. All Done with Pointers
3. O-O Document Databases

RELATIONAL MODEL

TABLE: book

BOOKID	AUTHOR	TITLE
100	Peter Snell	Document Management Systems
200	Shelly Keating	Object-Oriented Databases

TABLE: chapter   TABLE book-chap

CHAPID	TITLE	BOOKID	CHAPID
101	What Are Documents	100	101
102	Features of a DMS	100	102
201	What is an ODBMS	100	301
202	All Done With Pointers	200	201
301	O-O Document Databases	200	202
		200	301

OBJECT-ORIENTED MODEL

INSTANCE:	book		INSTANCE:	chapter
BOOKID:	100		CHAPID:	101
AUTHOR:	Peter Snell		TITLE:	What Are Documents?
TITLE:	Document Management Systems			
CHAPTERS:	(101, 102, 301)		INSTANCE:	chapter
			CHAPID:	102
INSTANCE:	book		TITLE:	Features of a DMS BOOKID: 200
AUTHOR:	Shelly Keating		INSTANCE:	chapter
TITLE:	Object-Oriented Databases		CHAPID:	201 CHAPTERS: (201, 202, 301)
TITLE:	What is an ODBMS			
			INSTANCE:	chapter
			CHAPID:	202
			TITLE:	All Done With Pointers
			INSTANCE:	chapter
			CHAPID:	301
			TITLE:	O-O Document Databases

retrieved from disk, others lock individual objects. If an application uses small-grained objects and these are stored on the same page of memory, then a system that locks pages can restrict access to objects in a seemingly unnecessary way. So, for example, two documents that share one small paragraph of text could not be edited simultaneously. Such a situation does not arise in a system that uses object level locking.

Version Control

ODBMS can track not just different versions of a document but also different versions of its components. For example, version 3 of a document might contain version 10 of a graphic.

The level of support for version control of objects varies between ODBMS products. The best allow versions to be stored as complete copies (quick to access but demanding on storage space) or as deltas (i.e., only store the changes between versions). Access to delta versions is slower since they must be reconstructed from the current version. Another feature that can be useful is the ability to maintain parallel versions that can be merged into a single version later on.

Document Configuration Control

In a compound document database system there must be some way to configure the contents of a particular document in terms of the components it contains. There are several ways in which this can be achieved using an ODBMS.

Because an OBDMS manages object elements and understands the relationships between them, it can view a document as a 'bill of materials' that specifies the objects it contains. This view means that documents can be created as database 'reports' (using reporting tools similar to those used in RDBMS) with the contents determined either by retrieving a fixed set of components as stored in the database or by evaluating a query expressed in the Object Query Language and constructing the set of components on the fly.

Client/Server Architecture

ODBMS are well equipped to take full advantage of the trend towards distributed client/server hardware platforms. ODBMS products generally have a client/server

---

ODMG-93 and CORBA Standards

The ODMG-93 standard was published in the Autumn of 1993 by the Object Database Management Group, a consortium with representatives from most of the major ODBMS vendors. Their aim has been to produce a standard that will allow source code to be ported easily between different ODBMS products, with the intention that ODMG-93 will play a similar role to that of SQL for RDBMS. The main features of ODMG-93 are:

- A clearly defined Object Model which underpins the standard.

- An Object Definition Language (ODL) that can be used to define objects and their associated properties and operations.

- An Object Query Language (OQL).

- Bindings to the C++ and Smalltalk programming languages.

Publication of the ODMG-93 standard and the commitment to support it made by all ODMG member vendors has countered earlier criticism of the lack of standards for ODBMS.

The Common Object Request Broker Architecture (CORBA) is a standard published by the Object Management Group, a vendor consortium which includes several ODMG members. An object request broker allows objects to be registered and accessed by multiple applications on a system-wide basis. The principle use of CORBA in connection with documents is in the OpenDoc standard — a platform independent competitor to Microsoft's Object Linking and Embedding, that allows applications to share document components such as text, graphics and tables.

For ODBMS, CORBA could provide a common means of identifying objects so that they can be accessed by different ODBMS' running on the same system.

architecture in which the server manages the storage and access of objects on disk and supports multiple client processes that manipulate objects locally. The server and client processes can be distributed around a network. ODBMS performance depends heavily on the manner in which processing is split between the client and server and the way that clients maintain a local cache of data retrieved from the server. Both factors should be considered carefully during application development so that the database can be tuned for optimal performance.

User Interface

The user interface is one area of ODBMS in which application programmers can expect to do most of the work themselves. The richness of the object-oriented data model means that graphical browsers of objects or networks of object relationships must operate at a low level if they are to cater to all general cases (i.e., the browsers are more useful for application developers than end users). Also, the range of integrated tools for GUI development is more limited for ODBMS than for RDBMS — a situation that is likely to change as the ODBMS market matures.

Backup and Security

ODBMS provide the same sort of backup and security facilities as their relational counterparts. This includes the ability to 'roll back' changes made to the database during a session and to keep changes local to a particular client process until committed to the server and broadcast to other users.

DOCUMENTS AND INFORMATION MANAGEMENT

For most organizations, documents play a key role in storing and exchanging information and hence the document management system must often be viewed within the wider context of an organization's information and workflow. Reflecting this, ODBMS have been used in several product information and workflow management products.

Performance and Scalability Issues

Two vital questions to be asked of any database management system are:

Does it deliver the performance required for a given application?

Does it maintain this performance as the number of users increases?

Both questions must be considered very carefully for ODBMS. High levels of performance can be achieved with ODBMS but badly designed or badly tuned applications can run very slowly. Performance benchmarks also vary considerably between different ODBMS products, so it's worth shopping around. Some factors to consider in relation to performance are:

- A good data model will maximize the amount of information that can be accessed by following pointers (fast) rather than executing queries (slower).

- If objects that are used together in the application are stored together on disk then they can be accessed in one trip to the server and stored in the client cache, thus improving access times.

- Some ODBMS can boost performance by having multiple servers running on a network — but the cost of buying additional servers must be taken into account.

- Performance can be tuned by building indexes for commonly executed queries or searches.

- Performance is lowered when features such as object locking and version control are enabled.

Documents as Business Objects

In such systems, documents are just one of a number of types of 'business objects' that are involved with a product development lifecycle. The system manages the status of all the objects and the relationships between them. Typical features of a product informa- tion or workflow management system include:

- Business objects are stored in regular system files created by any number of software tools involved in the product development process.

- Business objects might include documents, CAD drawings, spreadsheets, mail mes- sages, etc.

- The ODBMS is used to store meta data about the business objects.

- Meta data supports version control, access control and progression of objects through the development lifecycle (e.g. from design to delivery stages).

Implementation: 'Off the Shelf' vs. 'From the Ground Up'

Having decided to take the object-oriented path to document management, an organization then faces the choice of using an ODBMS to build an application from the ground up or of purchasing a document management product off the shelf. Naturally each approach has its advantages and disad- vantages.

Generally, implementation from the ground up gives greater control over the application but at the cost of greater development effort. Specific advantages of this approach are:

- The application can be designed to match requirements exactly.

- Direct support for standards such as an ODMG-93 gives the flexibility to port the application or change ODBMS vendor at a future date.

- Full control to enhance the application with future releases of the ODBMS.

- The ODBMS can be optimized precisely for the application.

The disadvantages are:

- Extensive application development is needed, including the user interface and basic features such as version control and configuration control.

- Good implementations require talented and well trained developers.

- Investment in ODBMS tools and development staff must be made before a working version of the application can be evaluated.

Buying a dedicated document management system reduces development effort but some design goals may need to be sacrificed. Specific advantages are:

- Most of the development work has already been done user-interface, version control and configura- tion control should all be part of the product.

- A working version of the product can be evaluated before the decision to purchase is made.

Disadvantages are:

- Requirements may be compromised by the capabilities of the product.

- The ODBMS is hidden within the product and cannot be reprogrammed or used independently.

- Standards such as ODMG-93 are only supported indirectly in the embedded ODBMS - there may be an additional, proprietary layer of data model or application interface.

- Future product support and upgrades rely on both the ODBMS vendor and the document database vendor.

Example - Numerical Algorithms Group

The Numerical Algorithms Group (NAG) is a leading supplier of software libraries for numerical and statistical analysis. Originally NAG produced a single library of FORTRAN routines but by the late 1980s they were shipping libraries in many different programming languages and embedding algorithms in third party products. In 1991, realizing that the task of maintaining links between related information in different products would rapidly become uncontrollable, NAG began to develop a database to manage product information with a particular emphasis on the generation of product documentation.

After developing and evaluating an initial prototype, a full implementation of the system went ahead using C++ and a mainstream ODBMS running on a network of UNIX workstations.

Documents can be generated from text fragments and basic product information stored in the database. Links are maintained between related items of information in the database and between documents and their components. When an information item is edited, the links are activated so that documents can be regenerated and related items reviewed to see if they require amendment.

The system is currently used by the publications department to assist in the production of a wide range of technical product documents in a variety of output formats (e.g. printed, on-line, hypertext).

SGML AND O-O DATABASES

The Standard Generalized Markup Language (SGML) has begun to play a significant role in the management of structured documents in many sectors of commerce and industry.

Two important benefits of SGML are its capacity to deliver multiple output formats (e.g., printed, on-line, hypertext) from a single source document and the opportunity it provides for sharing elements between documents. Careful management of SGML documents and document fragments is necessary if these advantages are to be exploited fully and hence there is a growing market for SGML text repositories. The inherently object-oriented nature of SGML makes ODBMS a natural choice for implementing such repositories.

Object-Oriented SGML Repositories

Most major vendors of SGML editors and on-line viewers have released, or announced plans to release, SGML element repository systems, based on ODBMS technology. The key features offered by such systems are:

• Management of SGML documents and document elements, including version and access control.

• Management of elements that are shared between documents.

• Facilities to configure complete, valid SGML documents from stored SGML elements.

• Tight integration with SGML tools such as editors and on-line viewers.

• Integration of SGML and non-SGML data.

• Usual database backup and security features.

The products currently available require a high degree of customization for individual applications, by developers with ODBMS and SGML skills. They will probably appeal to large organizations who have already invested heavily in SGML and who are willing to make substantial further investment in a specialized SGML document management system.

World Wide Web Servers and O-O Databases

Several products are now available which link ODBMS with World Wide Web servers. Such systems allow users of WWW browsers to send Object Query Language queries across the Internet to the server, which executes them using the local ODBMS and returns the results back across the Internet as an HTML document. This capability greatly enhances the scope of a normal WWW browser from a hypertext document viewer to a document database search engine.

BRINGING OBJECTS TO RELATIONAL SYSTEMS

The rapid rise of ODBMS has led many RDBMS vendors to search for ways of making their products more object-oriented. Generally, there have been two approaches - rewriting the kernel of a relational database so that it can also handle objects or adding an object-oriented layer to a database that remains fundamentally relational.

Several products are now available which link ODBMS with World Wide Web servers.

Hybrid Object/Relational Databases

A significant share of the ODBMS market has been captured by products that extend the relational data model to handle objects as well. The 'database engine' in such systems is capable of storing both relational tables and objects. The advantages of object/relational databases are:

• They provide a comfortable migration to objects for developers familiar with relational databases.

• They can store relational and object-oriented data within a single system.

The disadvantages are:

• They do not adhere to standards for either RDBMS (SQL) or ODBMS (ODMG-93).

• By catering for both relational tables and objects it is possible that neither will be implemented optimally.

• Document databases are unlikely to use the relational part of the system.

O-O Layers for Relational Document Databases

Vendors of document management products based on RDBMS have begun to offer object-oriented features by adding an additional interface layer between users and the underlying relational database system. This allows documents to be manipulated as compound objects whilst retaining the relational database for storage. The main advantages of this approach are:

• Users can exploit the benefits of object-oriented documents, such as maintenance of compound documents and sharing of components.

• The stable foundation of the existing relational system remains untouched (in contrast to the hybrid systems featured above).

The disadvantages are:

• The additional layer makes application development more complex.

• The object model for documents is proprietary and fixed.

• Document objects must be mapped into relational views before being stored, introducing a performance overhead compared with direct storage as objects.

Example - Electricite de France, Electronic Library System

The Electronic Library System at Electricite De France (EDF) is a project started in 1993 with the aim of managing the technical documentation produced by the 2,500 employees of the Research and Development Division. Some of these documents are already stored directly in a relational database, others are held as document images with bibliographic details held in the database. However, the RDBMS is limited to storing complete documents (not components) and does not support the vast web of links that exists between them.

The Electronic Library System uses an ODBMS to store individual components of structured, multimedia documents and to maintain the links between components. The system will be required to manage tens of thousands of documents, accessible by many users (potentially several hundred).

RISKS AND COSTS

For some years, ODBMS have been viewed with some suspicion by many systems developers who have regarded the technology as being largely unproven. Indeed, the choice between object and relational technology for a document database project often comes down to weighing the inherent suitability of the object-oriented approach against the proven track record of relational systems in large scale projects. With this in mind, it is worth considering carefully the risks and costs involved in any object-oriented document database project, such as:

- The risks associated with being a technology pioneer.
- The risk that emerging object models (in object- oriented operating systems, for example) may not be compatible with the chosen ODBMS.
- The risk that prototype systems will not scale up satisfactorily.
- The risk that performance will not be adequate (particularly in relation to scaling up).
- The risk that the ODBMS will not interface properly with existing RDBMS.
- The cost of switching to an object-oriented document creation process (e.g. authoring in SGML).
- The cost of converting legacy data.
- The cost of acquiring properly trained staff or retraining existing staff.
- The cost of upgrading to client/server hardware.

Object-oriented systems can be radically different from existing technology.

To some extent, the costs of implementing an object- oriented system are no different from implementing a new system based on any other technology — any new system is likely to involve new hardware, new document creation processes and re-engineering of existing data. However, the cost of ensuring that developers and users are well trained is an important factor to consider, since object-oriented systems can be radically different from existing technology.

The risks of adopting ODBMS technology have certainly reduced in recent years, as evidenced by the number of new ODBMS-based products now emerging. However, it is still worth taking careful steps to minimize the risks listed above, including:

- Looking for an ODBMS system that supports the ODMG-93 standard.
- Ensuring that the system interfaces with RDBMS if this is a requirement — for example, some ODBMS support the Open Database Connectivity (ODBC) standard.
- Finding reference sites that have already implemented similar ODBMS-based projects.
- Testing the scalability of the ODBMS solution at an early stage (preferably before a commitment to purchase it is made!).

CONCLUSIONS

Object database systems are a new generation of database systems and they are here to stay. They meet many of the requirements of document database systems which the previous generation of relational databases have struggled with. This includes the ability to handle document components and multimedia data and to capture the complex interrelationships required in many document management applications.

The market for ODBMS is growing rapidly, boosted by the introduction of the ODMG-93 standard and the increasing popularity of the C++ programming language. This rapid market growth reflects the shift of ODBMS from innovative research projects to a commercially viable technology, with stable products, well established product vendors and industry-wide standards. Object-Oriented database management systems should now be considered as a serious option for the implementation of document databases.

John Chelsom

John Chelsom runs CSW informatices Ltd., a company dedicated to developing document management solutions using SGML and database technology.

CALS Update

There is renewed activity on the CALS standards front. MIL-STD-1840 (the CALS standard that describes how to package various electronic files for delivery) is being substantially revised. The new revision (Revision C) will take account of some of the tremendous advances in the variety and use of electronic files, standards, and commercial product capabilities. This standard will be of interest mainly to defense contractors (and subcontractors). It *should* be of significant interest to them, and *should* be closely tracked. As a follow-up to meetings in both March and April, there will be a meeting June 21-23. Below is the contact information and some background taken directly from the meeting notice. For further information contact George Touchette (513) 427-5888 toucheg@wpdis01.wpafb.af.mil, or Alan Peltzman (703) 735-3568 peltzmaa@cc.ims.disa.mil.

Background

Revision C to MIL-STD-1840 will update data format specifications and add new media types, as well as incorporate format and philosophy changes desirable under the new DoD standardization paradigms. Special attention is being paid to accommodating COTS tools, Internet, and electronic transfers; and to future migration to a non-Military Standard.

1840 Rev C "Class I" is the subject of this meeting. "Class I" will retain the maximum practical compatibility to MIL-STD-1840B. "Class II" will be an architectural restructure of the standard for future extensibility. The schedule objective is to complete a Revision Team draft of "Class I" by 23 June 1995, with a technical definition for "Class II" by the end of 1995.

Objectives

The objectives of this meeting are to:

1. Coordinate and/or close all "1840 Class I" Action Items.

2. Reach a Revision Team consensus on the 1840C draft document (structure, contents, and comments).

3. Establish milestones for completing "1840 Class II."

SHOW WRITE-UPS

DOCUMATION '95

Documation '95 was a conference not of surprises, but of confirmations: the industry has found its footing and is on a steep upward climb.

Held in Long Beach, March 7-9, it showed a very positive growth in its quantifiable metrics; seminar attendance was up 20%, vendor participation in the exhibition was up 180%, and attendance at the exhibition hall was up almost 200%. But, perhaps the most important growth "statistic" was the increased agreement on what the central issues are: new ways of distributing documents (including the Internet), the need for standards, and building systems upward from the management layer. This was reflected in the tutorials, panels, general sessions, and product offerings in the exhibition hall.

Each morning was given over to a general session, chaired by Frank Gilbane. At the first, on the state of the industry, Frank discussed trends during the past year and made a number of predictions for the next year. The make-up of the panel demonstrated the range of the show. Jeff Miller of Documentum laid out the repository-based document-management architecture that undergirds many implementations. Dean Cruse of Recognition Software discussed the role of workflow systems in compound document management architectures. Rory Cowan of RR Donnelley & Sons defended the continuing importance of print and pointed to the "downstream revolution" in digital print that, in some ways, dwarfs the changes we're seeing in authoring and compound document workflow systems. Peter Lamb of Andersen Consulting gave an implementor's view of the emerging market and technology with his combination of folksiness and incisiveness.

The next day's general session presented the views of senior management of three large users of document management systems (Jim Jensen of McDonnell Douglas, Barbara Faucette of Glaxo, and Ewart Newton of Citibank) and then four implementers of such systems (Sandy Emerson of Sybase, Paul Jensen of Thomson Publishing, Jeff Baeckler of GE Nuclear Energy, and Cesare Del Vaglio of McGraw-Hill). All seemed to agree, at least implicitly, with Jim Jensen who warned vendors not to try to be the best at everything, but to provide the hooks to enable "best of breed" systems to be built. Barbara Faucette cautioned that this requirement for openness means more than "I've published my API." Paul Jensen, in a talk provocatively entitled "Cultural Failures of SGML," also pointed to the "class warfare" the introduction of SGML to the workplace can cause — the "deprofessionalization" of editors who now are required to do new tasks, the alienation of the technical staff, the fears the new system will eliminate jobs.

On the third day, the general sessions opened with a panel on the key technologies that will shape the direction of document management. Bob Epstein of SyBase talked about the "integrated enterprise" we are quickly moving towards, with documents being the principle way companies communicate with customers. The rapid growth of bandwidth will allow documents targeted at an audience of one and other heretofore unrealizable capabilities. Curt Allen of Folio talked about the coming "Information Marketplace" where electronic document transactions need to be managed on a large scale (including security, transaction costs, and copyright management).

The general sessions closed with something akin to an historic occasion. The topic was "Document Interoperability Standards," and together on the stage, ODMA (Alvin Tedjamulia, Novell), Shamrock (Frank Dawson, IBM), and DEN (Ira Scharfglass, XSoft), presented a common vision of how the three standards will interoperate

Frank Dawson and Ira Scharfglass even hinted that DEN and Shamrock would announce the merger of the two organizations a few weeks later.

At the same panel were represented SGML (Mary Laplante, SGML Open), OpenDoc (Neil Kapin, CI Labs) and the Object Management Group (Chris Stone, OMG). All three of these claimed to be relatively neutral about the other standards, although Chris Stone strongly intimated that OpenDoc has the inside track on being recognized by the OMG as the standard model for document objects, primarily because OpenDoc has gone through the process of applying for that status while its major competitor, OLE (Microsoft), has not. In addition, OpenDoc is built on the OMG's CORBA object model, whereas OLE uses Microsoft's COM model.

The afternoon sessions reflected these general themes. There was a marked seriousness about SGML at the show, a topic that pervaded most panel discussions. There were a solid number of presentations by users, as well as some "mopping" up of the new view of document management, on topics such as how product data management, workflow and groupware relate to document management; the prevalence of such topics of con-solidation is a sign that the new view of document management has firmly taken hold.

There was, of course, a huge interest in publishing on the Internet, reflected in the tutori-als, panels and — most of all — on the show floor. Rare was the booth that did not have some form of Internet publishing offer.

Several vendors made announcements at the show, and many others showed new prod-ucts. Some of the announcements were:

Odesta Manages Work Processes

Odesta Corporation launched Odesta Livelink, a "workprocess" management system that combines workflow, project coordination and document management. It allows users to establish visual workflows for any set of objects, and tracks and manages them through-out their life cycles. It also gives a project view that enables a workgroup to share infor-mation easily. An extensive development environment makes the software highly config-urable and extensible. The client runs on top of virtually any database.

OpenText Indexes the Web

Open Text Inc. has announced it is building a full text index of all the documents on the World Wide Web which it will be making available as a free service to anyone connected to the Web. It has Web crawlers out finding every document they can, and will continu-ously update the index. Users can search for information by typing in words or phrases, using Boolean searches. The service is available (in beta) through Open Text's home page: http://www.opentext.com/

The company also announced that MKS Inc. will provide users with easy access to the Open Text Web Index in their Internet Anywhere product.

Texcel Manages SGML Information

Texcel announced Texcel Information Manager, an SGML-based workgroup authoring system. It provides rigorous support of SGML and manages the elements of documents, to any level of granularity. A sophisticated electronic review capability distributes drafts, gathers comments, and enables the author to automatically update the original. It also provides workflow management and automation, and comes with tools to administer the repository. Information Manager's collaborative authoring and electronic review modules integrate with ArborText ADEPT.

Xyvision Updates Parlance

Xyvision announced the release of the new version of its compound document management system, Parlance Document Manager. It has a new API and a host of new features.

Other new products demonstrated included:

• Document Sciences Corporation's new document automation architecture - Compuset.

• InContext demonstrated their new developers tool kit.

• XSoft showed their new object based document management system, now christened "Astoria".

• Novell showed their new WordPerfect SGML edition, as well as their HTML authoring tools.

• Microsoft showed their Internet Assistant and the newest version of SGML Author for Word.

• ArborText showed new HTML creation tools.

• MicroStar showed new authoring tools as well as support for FrameBuilder document definition files.

Documation will be held in Long Beach next year on March 10-13.

AIIM

If you want big and you want imaging, AIIM is for you. But as the imaging folk — vendors and users — recognize that managing the images of documents is only part of a broader process, the show increasingly is adding sessions and exhibits on the topics familiar to readers of this newsletter, including electronic delivery, the Internet, document workflow and repository-based document management.

This year AIIM, in San Francisco, was bigger than ever with over 44,000 attendees. The exhibit floor occupied all of the Moscone Center, north and south. Among the announcements:

Wang and Microsoft Improve Their Image, Workflow

Microsoft has made a $90 million investment in Wang Laboratories. It will bundle the Wang image viewer in with Windows 95 for viewing TIFF images. They also agreed to collaborate on a workflow API, and did not promise that it would be compatible with the API being put forward by the Workflow Management Coalition. In addition, Wang will provide support services to Microsoft.

Acrobat Captures Paper

Adobe has announced Acrobat Capture, a "page capture" program that converts paper documents into PDF, the format used by Acrobat for viewing files. Capture preserves formatting information about the pages scanned in, including complex layouts, mixed fonts, graphics and multiple columns. It performs optical character recognition on the text so that it can be copy and pasted, spell-checked and indexed. If Capture isn't sure about a word that it has processed, it replaces it with a bitmapped image of it and indexes its best guess at it, reducing the need for human intervention. It can also output in RTF and popular word-processing formats so it can be used even when Acrobat is not in the picture.

Fulcrum Surfs the Net

Fulcrum Technologies has announced its Surfboard, a search and retrieval server engine for the Internet. Surfboard creates full text indexes of document collections, supporting all major formats; the searches can be performed over the Web so that those visiting the

site can find the information they need. Non-HTML documents can be converted on the fly to HTML for viewing with a Web browser. It is scheduled to ship in May and will be priced at US$15,000.

INTERNET WORLD

If not much was new at the spring Internet World, in San Jose, it was perhaps a sign of an industry that has not yet come into its own. There's at least another year of infrastructure development work to be done before the business model becomes obvious. In the meantime, every major vendor seems desperate to make its mark by offering something — anything — Internet-related.

About 20,000 people attended the show. Vendors seemed to be offering network access services, make your own home page in five minutes, or browsers.

Some new products:

Active Information Management

Active showed E-Mall, enabling ordering consumer goods over the Net. It consists of server software that contains a database of product information and client software for browsing that information. The consumer chooses a catalog for a supplier or category, and then specifies the types of goods he wants to see. He can then get further details including multimedia demonstrations. If something strikes his fancy, he can put it into his "grocery cart" with the click of a button, and then can order the goods, having it charged to his credit card. Users can specify the hours in which they want it delivered and save carts for frequently-ordered sets of goods such as groceries. It is designed to be international, with a built-in VAT option and currency conversion. The client is free to users; the server is not free to suppliers.

NLIGHTN

NLIGHTN showed its unannounced product, a "universal index" of the Internet and on-line databases. The user can search by full text search, by database, and by some structure. Users can see an extract of the information, automatically generated, before downloading the article. The viewer is free. Downloading is not. The company sees its competitors as Dialog, Nexus/Lexus, Dow Jones and others. NLIGHTN did not announce when it will disclose the product.

Sun

Sun Microsystems showed its recently-announced Hot Java. Java is a C++-like language for initiating and controlling multimedia events. Java travels with HTML files and is executed by Hot Java, Sun's Web browser. Hot Java is under-featured compared with the current generation of over-featured browsers; you would use it specifically to view Java-enabled sites.

Microsoft

Microsoft showed its HTML Assistant for Word, a Word add-in that provides HTML conversion and editing. They also showed their NT Web server.

Interleaf

Interleaf announced that Cyberleaf, its Web publishing product, that takes in files from multiple word processors and builds them into linked Web collections, is available to academic institutions at cost. The product is normally priced at $795 for Unix. The company also showed a beta version of the software running on NT Windows.

Quarterdeck

Quarterdeck showed its version of Mosaic which can download multiple documents simultaneously, display them as they download, and tack them through a drag-and-drop tree structure displayed in a separate pane. They also showed their add-in to Word for authoring HTL documents. Pre-release versions of the software are available at http://www.qdeck.com/

SkiSoft Publishing Corporation

SkiSoft introduced its Web Publisher, a conversion tool that shapes RTF files into HTML documents adapted to the Web. The conversion is template-driven, and will create HTML tags based on explicit mappings, guesses at mappings based on names (e.g., a Word "Heading1" it will guess should be an HTML "h1"), or on formatting cues. It can automatically insert graphical buttons for navigation and a hyperlinked table of contents at the front of the document. A trial version is available at http://www.skisoft.com/skisoft

SoftQuad Inc.

SoftQuad announced that HoTMetaL and Panorama are shipping with Spyglass's Enhanced Mosaic 2.0. Enhanced Mosaic is the commercial version of the original Mosaic created by the NCSA. HoTMetal is an HTML editing tool and Panorama is a dynamic SGML viewer. By integrating Panorama, Mosaic now can display any SGML documents it comes across in addition to the standard HTML documents.

SPRING SEYBOLD SEMINARS

The Spring Seybold Seminars in Boston was bigger and brassier than ever, but also increasingly focused on its roots in electronic prepress and output in the graphic arts.

Until a few years ago, exhibitors at the spring seminars were supposed to show new technology and were forbidden from trotting out the usual, gaudy booth fare. The result was a low-key event that felt a bit like a science fair. Now, however, the booths and lures are unbridled so it looks and feels like the normal trade show.

The content, however, remains distinctive. Whereas the show a few years ago looked like it might be shifting its focus to the new arena of compound document management, it now only had one seminar session on that topic. Electronic viewing and Internet publishing still have a good presence at the show, however because the graphic arts providers who are its backbone understand that they need to move in that direction.

It's a big show. About 17,700 people attended, up over 20% from the previous year. Approximately 2,500 people attended the seminars, and there were 225 exhibitors.

Among the announcements at the show:

Acrobat Works with a Net

Adobe announced Netscape Navigator will support Acrobat Weblink, a free add-on that allows Acrobat documents to link to other Web documents. A future version of Navigator will "seamlessly" view PDF (Acrobat) documents. The Netscape server will be able to download randomly-accessed portions of PDF files — for example, a page from the middle of a long document — to avoid having to squeeze an entire PDF document over the net. Adobe also says that a future release of PageMaker will output HTML.

Adobe also announced it will move Acrobat onto OS/2 on Intel and PowerPC platforms, and will support IBM's AFP architecture so that AFP can be used to manage PostScript files for on-demand printing.

Jouve Views SGML

Jouve announced GTI-Publisher, an off-the-shelf version of an SGML viewer creating using version 5 of the GTI Software Developers Kit. The server module builds a full text index of SGML documents, sets styles, and establishes links. The result is a set of files that can be viewed with their client software. The user can create bookmarks, but cannot alter the formatting of the document. The company said the pricing will be "competitive" — and they view EBT as their competition— when the product ships in the fall of this year.

InContext Spider Spins Webs

InContext announced a Windows-based "web processor" called InContext Spider. This takes the user interface of InContext 2, an SGML editor, and packages it for creating HTML documents destined for the WorldWide Web. The system lets a user create HTML elements and type in text, ensuring that any document that comes out of it is legitimate HTML. At any point during the editing session, the user can launch the document into a Web browser (one will be bundled, but InContext was not ready to announce which it will be) to see the outcome and to check the links. In addition, the URLs within any document currently in the browser can be extracted and pasted into the document under edit with a click or two of the mouse. It will cost $99. A development kit will also be available.

Open Text Gives Some Latitude

Open Text introduced Latitude, a document distribution system that can handle both SGML and WYSIWYG documents. It uses Open Text 5 to build a full text index of documents that may reside on various servers throughout an organization. The user searches for documents and the system automatically launches the appropriate viewer, including an SGML viewer based on Panorama, Common Ground for WYSIWYG documents, and the Spicer viewer for engineering drawings. The company demoed the product on a one-gigabyte collection of documents from Catepillar, residing on PC, Macintosh and Unix servers, and on a CD-ROM. Latitude also includes a button that uses the Open Text Web Index to search for documents across the Internet.

Microstar Assists Word to HTML

NEAR & FAR Author is an add-on to Microsoft Word that turns Word into a guided editing word processor that outputs SGML. It presents a graphic picture of the DTD tree in a separate window that also shows the author's current position. While editing in Word, the only paragraph styles presented to an author are those that are legitimate SGML elements. The document can be saved as an SGML file or as a Word file. It is due to ship in June. It will cost $249 per copy, with volume pricing available.

EBT Announces Partnerships and Giveaway

EBT made several announcements at the show.

EBT is integrating DynaBase and ArborText's ADEPT SGML editor so that documents and document fragments when checked out from DynaBase immediately launch ADEPT.

They also announced they have licensed Belise from AIS for use in DynaPage. Belise transforms SGML into Framemaker, Interleaf or RTF pages. It uses whatever formatting information there is in the DynaText style sheet. The software is AWK-like, but is more object-oriented and has more C++ constructs, according to the company.

EBT has also established a university grant program whereby the company provides software in return to rights to whatever gets developed.

Finally, the company announced a partnership with Phoenix Software which is building documentation viewers based on EBT software which will be bundled on the hard disks of

some major PC manufacturers to be named at a later date. Phoenix provides a graphical front end — a drawing of library shelves, in one example — to make the documentation easier to use.

FrameMaker Goes to 5

Frame announced release 5 of Framemaker, available in June on Windows, Macintosh, PowerMac, Sun and HP; the company announced it will also support OS/2. The new release adds some additional pagination functionality such as straddle-column heads and automatic runarounds, and the ability to share text across unrelated documents (text by reference).

Frame documents can be saved as HTML, requiring only that the user map the Frame elements to the standard HTML elements via a dialogue box. The system can also generate PDF for Acrobat viewing, adding the required structural information required for Acrobat to show a hyperlinked outline. It also supports more types of links (including pull-right).

The company also announced the next version of FrameBuilder, now named Framemaker+SGML. The release aims to increase ease of use while increasing support of SGML. The new release can display attributes in the structure-view window. Attributes can be edited, and can affect the format, so that changing a list attribute can change the list from numbered to bulleted, which is reflected in the WYSIWYG document window. The WYSIWYG window can now optionally show tags.

The product automatically manages ID attributes, using a cross-reference metaphor. It can automatically generate reference text such as in an appendix.

The tables in the product now are fully structured and use the CALS model.

The EDD, which expresses the mapping of the DTD to Frame formatting, can automatically import a DTD and allow an author or manager to set formats relative to to the document context by stating the formatting information directly, by referring to existing paragraph styles, or by creating paragraph style groups.

Navisoft Browses and Serves

NaviSoft, Inc., a subsidiary of America On-line, introduced NaviPress and NaviServer, a client/server network publishing system. NaviPress lets you create Web pages and work with the links without first having to save the document and load it into a browser. It is able to create "mini-webs" — sets of interlinked documents — and shows them in a visual map. The system can map any Web document loaded into it, including those encountered elsewhere on the Net.

NaviServer is based on a relational database (from an unnamed vendor) and enables a publisher quickly to create a relational table and a set of HTML forms for end-users to interact with in order to populate the table. The database tracks permissions and accounting, including the ability to accumulate per-page charges (as opposed to merely charging for connect time). It also performs versioning, so that an HTML document that includes graphic elements can be "rolled back" to a previous version, including the previous version of the graphic. In addition, NaviServer uses the database to provide search and retrieval functions, and to provide some intelligent guesses, based on semantic analysis, to authors about where they may want to link a highlighted phrase to.

XSoft Astoria Manages SGML

XSoft, a division of Xerox, introduced Astoria, a structured document management system. Although it can handle any type of data, it can manage the individual elements of an SGML document.

Using ObjectStore from Object Design, Inc., Astoria builds a database of document elements. The individual accesses ther personal view of the database through a "workbook" which shows the hierarchy of a document set, using cabinet and folder icons in a File Manager-like display. It also shows the content model of the DTD, the first couple of hundred characters in an element, if an element is checked out, and if you have permission to access it. The database can be searched through full text, Boolean or proximity queries.

When an element (or set of elements) is checked out, the system "fabricates" a sub-DTD and creates an SGML document which is then edited either in InContext or SoftQuad's Author/Editor. The material is re-parsed when checked backed in, and any invalid elements are flagged as such. Versioning information is applied. A user can declare a new "edition," which is a permanently-saved set of documents at their current revision level.

The system provides a viewer which is intended not as a delivery vehicle to end users but as a browser into the dynamic database.

Astoria is a client/server application, and comes with a set of server tools, some from ODI and some specific to Astoria, including tools for adding users, changing permissions, etc.

The Solaris version of the Server will ship first, with an NT Client, to be followed within a quarter with a Windows client, and then a Unix client.

INDUSTRY NEWS

INSTANT TRENDS

If all you had to go on were the news since the previous issue, you would spot the following "trends":

All products are either moving up to new versions or being made available for free on the Web, or both.

Businesses seem to be in the business of making business partnerships with other businesses, in order to sell products to the businesses who are not (yet) their partners.

FACTOIDS

According to the Software Publishers Association, sales of software on CD's reached $648 million in 1994 on unit sales of 22.8 million. Home education products grew 258% and games grew 222%.

The Net grew over 31,000% in the past three years in the U.S., and by over 400% in 1994 in some areas. And it's not slowing down: the growth rate in the fourth quarter of 1994 was 26%, the largest in recent history; the fastest growth of hosts was in commercial (.com) domains. (Source: Internet World magazine)

The average small business spends 10 hours per week faxing documents, eight hours copying them, and seven printing them, according to Impulse Research in a survey of companies with fewer than 50 employees done for Okidata. (Source: USA Today)

A survey by Intelliquest for Microsoft showed that 63% of children ages 11-17 would rather use a computer than read a book and 59% would rather use it than watch TV. (Source: USA Today)

The Playboy home page received 659,448 visits on March 2, but was beaten by Penthouse the next day when it received 853,000 visits.

Video is now more popular than computers for training, according to a survey of 2,000 companies by Anderson Soft-Teach; human instructors are number one. (Source: Information Week)

E-mail traffic is increasing, but physical mail volume, at 787 billion pieces in 1994, was up 5 billion (most of them catalogs mailed to my house in the week before Christmas). Faxing has reduced the paper traffic among businesses, but business-to-consumer mail has more than made up the difference. (Source: Inter@ctive Week)

OMNI GOES WEB AND ONLY WEB

Omni, the science magazine, is moving from print to on-line as its primary delivery mechanism. The monthly version will be on-line only, with a quarterly "super" edition appearing in paper on newsstands (i.e., no subscriptions). The magazine expects to save about $4 million a year in printing and distribution costs. It has a circulation of 750,000 in its hardcopy form.

PHOENIX PUTS SGML ON-LINE ON PC'S

Phoenix Publishing Systems, Inc.'s VIRDOX software enables PC hardware companies to publish their SGML-based end user documentation on-line with an easy to use graphical user interface. It uses EBT's DynaText as its display engine and incorporates search capabilities. Phoenix anticipates a minimum of five million PC's to be shipping with on-line documentation using VIRDOX by the end of the year.

InfoAccess's GUIDE Passport for On-line Publishing

InfoAccess has introduced GUIDE Passport, an electronic publishing system that can import word processing files, create links and a full text index, and enable designers to create a customized user interface. The product is template-driven for production publishing, and includes a macro programming language. It is schedule for release in May for US$1,875 for Windows.

WYSIWYG Web Authoring from GRIF and INRIA

GRIF S.A. and INRIA have announced Symposia which they claim is the first WYSIWYG authoring tool for HTML documents. (Isn't WYSIWYG HTML an oxymoron?) The software enables many authors to work on the same document through a document-locking mechanism. The software is free and can be downloaded from http://symposia.inria.fr.

DOCS Open for the Enterprise

PC DOCS is now shipping DOCS Open V2.5, an enterprise version of its document management software. The product now runs on the Macintosh, in addition to PC's and Unix. (The company has dropped the "PC" in the product's name to reflect the presence of other types of computers in enterprises.) Enhancements include a new, optional application development environment, a tool that uses Lotus Notes to replicate DOCS data collections (so users can publish to anyone on a Notes network), enhanced NT security, and a new workflow module.

Intergraph Announces Document Management Solution

Integraph has announced DM2, a modular set of off-the-shelf products for document management. DM2 is built on object-oriented technology from Metaphase Technology, Inc. It consists of a server with industry-specific data modeling, a client, viewing and redlining, full-text retrieval, and graphical workflow. It runs on Unix, Windows and Macintosh. Intergraph is targeting the process, utilities and federal industries during the first half of 1995, and will add transportation and manufacturing in the second half.

InTEXT Announces Web Publishing Tools

InTEXT Systems has announced Precision, a toolkit that automatically converts documents to SGML and HTML formats, determines their key words and phrases, and creates condensed documents. These "thumbnail" documents produce indexes the company says are five to ten times smaller than traditional full-text indexes. The toolkit is available now for Windows and Unix.

FileNet and Meta Do BPR Software

FileNet and Meta Software Corp. announced they will jointly develop and market a software product that provides an interface between Meta's business process reengineering software (Workflow Analyzer) and Filenet's Visual Workflo software. Integration of the two products will enable uses to document, simulate and optimize a mission-critical business process, and export the resulting model into Visual Workflo. The software will ship to "initial users" in June, with greater availability later this year.

FileNet also announced that it document-imaging and WorkFlo software will support Windows NT Server 3.5, with general availability in the third quarter.

KEYFILE ADDS SQL GATEWAY

Keyfile Corp. has announced it has developed a gateway for its products that will enable users to access SQL databases. With the Open SQL Gateway, Keyfile users can query any standard RDBMS storing Keyfile documents. It will be available in the second quarter for US$2,495 for a single unit.

ACTION DOES WORKFLOW FOR OTHERS

Action Technologies, Inc., has announced ActionWorkflow DocRoute engine will be used by Watermark, LaserData and Imagery Software as the workflow component for their document imaging solutions. PC DOCS and Saros also announced they are incorporating the workflow management engine. The DocRoute engine provides integrated workflow capabilities for client/server enterprise document management systems.

EXCALIBUR ANNOUNCES NEW RELEASES

Excalibur Technologies Corp. has announced version 2.0 of their TRS Text Retrieval Server that supports concurrent access to multiple document database indexes on a LAN or WAN. The company also announced an optional WWW server integration kit that interfaces to Web servers through the standard Common Gateway Interface so that TRS can work with an Web server. The company also announced EFS 3.6, its client/server document management and retrieval software, adding native file format support, color image support, and annotation. The new version of ESF will be available in the second quarter; a 10-user concurrent-use license is US$51,000.

OMG TO INTEGRATE OLE AND CORBA

The Object Management Group has issued a Request for Proposal to provide interworking between Microsoft's OLE object model (Common Object Model) and the OMG's Common Object Request Broker Architecture (CORBA). This would enable users of OLE — virtually all Windows users — to use applications that support the OMG object model, which in future releases will include the OLE-to-CORBA gateway. Microsoft supports this approach. OpenDoc, the everyone-but-Microsoft rival to OLE, will be CORBA-compliant when it ships.

IDI ANNOUNCES WEB SERVER

Information Dimensions, Inc. has announced a Web server designed to handle huge collections of HTML documents. BASIS WEBserver provides document management features and navigational tools. It automatically converts documents in a BASISplus database into HTML; it supports documents in popular word processing formats as well as ASCII and SGML. Users can navigate using graphical forms and "Search Assistants." The system can also generate a Virtual Table of Contents for HTML files.

KALEIDA MEDIA PLAYER FREE ON WEB

Kaleida Labs Inc. is making available free copies of its Media Player on its WWW site. The software is available for either Windows or the Macintosh. The aim is to encourage developers to use ScriptX, a device independent-language for creating multimedia scripts and objects. The Web site is http://www.kaleida.com/

COMMON GROUND LAUNCHES 2.0, GAINS HANDS

Common Ground Software, Inc. has changed its name from No Hands Software, and has announced shipment of Common Ground 2.0 for Windows. The new version adds advanced search, navigation, annotation and publishing features, such as: resolution-independent fonts; on-the-fly anti-alaising of text, Verity search and retrieval; PostScript support, thumbnail views; JPEG and text compression, and DES encrypted security.

The company also announced Internet Publisher for publishing documents on the Web while preserving all formatting, graphics and fonts. The software can also download a page from the middle of a document, rather than having to download the entire document.

APPLE INTRODUCES INTERNET SERVER

Apple Computer, Inc. has introduced its Internet Server Solution, a PowerPC loaded with Web server software. Apple is marketing it explicitly as a way to build and maintain a Web site without having to know Unix. The software bundle includes: MacHTTTP Web server from BIAP Systems; BBEdit HTML editing from Bare Bones Software; Netscape browser from Netscape Communications; AppleSearch indexing software from Apple, and Adobe Acrobat Pro. The system begins at US$3,000. (Apple reports that there are an estimated 1.7 million Macintoshes connected to the Internet and over 2,900 Macintoshes being used as WWW servers.)

SAROS ANNOUNCES UPGRADE

Saros Corp. has announced Saros Document Manager version 1.5, their document management system based on Saros Mezzanine. New features include mobile computing, more platforms (Windows NT Workstation and Windows 95, with Macintosh and Motif versions slated for the third quarter), UI enhancements, Adobe Acrobat integration, and integration of workflow and imaging products from Watermark, Action and Recognition. The price per workstation is US$495 and is available now.

KEYFILE ANNOUNCES 3.0

Keyfile Corp. has announced the new version of its document management and workflow system. Keyfile 3.0 enhances version control, workflow, and user interface customization. The new version provides multilevel version control so that folders and their contents can be tracked together and independently, dynamic and graphical workflow mapping, and API enhancements. Keyfile 3.0 costs US$795 per client. Version control is a $995 option. Availability was not announced.

Xyvision Launches Partners Program

Xyvision, Inc. has announced a partners program to help corporate and third party integrators use their products to build document management solutions. The program will help the partners identify opportunities, and includes training, technology transfer (assisted by a "relationship manager") and continuing support.

Workflow Coalition Provides Documentation

The Workflow Management Coalition, with 135 members, will enable the industry to review its core set of workflow standards, including the API it is developing for interaction with workflow systems.

News flash: Hardly anything changes

"A device for transmitting handwritten numerals into computer systems" has been "developed by IBM. The experimental model identified 98.5 per cent of the 100,000 numbers written by 150 persons and put the numbers into punched-card form."
— Entry under "Computers" in the 1963 World Book Year Book

Buzzer of the Month

This month's coveted award for the best use of buzz words in a press release goes to Xyvision:

"Xyvision's own Parlance Document Manager (PDM) is a client/server information management system that stores document components as information objects inside a central database, allowing these elements to be shared, reused, and recombined in multiple documents and other 'information products."

People News

Jon Kannegaard has been appointed president of SunSoft. He had been vice president and general manager of the company's developer products business unit.

George LeBlanc has joined Workgroup Technology Corporation as vice president of marketing.

Fulcrum Technologies has hired **Joseph Nardi** as vice president of Asia Pacific sales and **Suanne Day** as vice president of customer services.

CALENDAR OF EVENTS

Below is a selection of key events covering open information and document system issues. There are many other conferences and shows covering related topics. We will attempt to keep this list to those events that focus on areas most directly related to the topics covered in our report.

On Demand Digital Printing and Publishing Strategy Conference and Exposition. June 27-29, New York City, NY. The only commercial conference and expo devoted to on demand printing technology and applications. Sponsored by Expocon and CAP Ventures.
Call (203) 256-4700, ext. 139, Fax (203) 256-4730.

Electronic Publishing Special Interest Group Tutorial Series: World Wide Web, Internet, and SGML. July 10-11, Chicago, IL and July 13-14, Austin, TX. Revolutionize the way your compnay communicates with the world. Sponsored by Electronic Publishing Special Interest Group.
Call (703) 548-2867 for more information.

CAP Document Management & Electronic Delivery Seminars. Fall dates TBA, London, UK. These two day seminars are conducted by Gilbane Report staff and are managed by Technology Appraisals. Call +44 81 893 3986 or (617) 547-2929, Fax +44 81 744 1149 or (617) 547-8811.

Seybold Seminars '95. September 26-29, San Francisco, CA. The annual conference where the publishing technology elite gather. Focus is on pre-press, color, newspaper, and magazine applications with some corporate application coverage. Call (415) 578-6990, Fax (415) 525-0183.

Document Management and Imaging Expo Tutorials. October 23, New York, New York. Introduction to document management and imaging as well as workflow and business process redesign. Call (617) 837-7200, Fax (617) 836-8856.

Document Management and Imaging Expo. October 24-26, New York, New York. The premier document management and imaging event on the east coast. Call (207) 236-2524,
Fax (207) 236-6452.

CALS Expo '95. October 23-26, Long Beach, CA. The annual expo and conference covering CALS activity in the U.S. and internationally. Heavy defense industry emphasis. Call (202) 775-1440, Fax (202) 775-1309.

SGML '95. December 4-7, Boston, MA. Sponsored by the Graphic Communications Association. Call (703) 519-8160 for more information.

Documation France Conference & Workshops. December 11-14, Paris, France. Sponsored by Techno-Forum SARL. These two day seminars are moderated by Frank Gilbane, Yves Stern, and Guy Fermon and cover compound document management and electronic delivery technology and trends. Call +33 1 43 48 57 92 or (617) 547-2929, Fax +33 1 43 48 55 43 or (617) 547-8811.

TOPICS COVERED IN PREVIOUS ISSUES

TOPICS COVERED IN UPCOMING ISSUES

Order Form

○ Please start my subscription to: The Gilbane Report on Open Information & Document Systems (6 issues). Back issues available for $45 each.

U.S.A.: $225   Canada: $232   Foreign: $242

Additional copies and site licenses are available at reduced rates. Call for information.

Please send me additional
information on:

○ Documation '95   ○ Market Research
○ Strategic Consulting Services
○ Compound Document Management Consulting Service
○ Document Management & Electronic Delivery Seminars

○ My check for $ _____ is enclosed ○ Please bill me
○ Please charge my credit card   ○ MasterCard   ○ Visa   ○ American Express

Name as it appears on card _____ Number _____

Signature _____ Expiration date_____

Checks from Canada and elsewhere outside the U.S. should be made payable in U.S. dollars.
Funds may be transferred directly to our bank: Please call for details.

Name _____ Title _____

Company _____ Department_____

Address_____

City _____ State _____ Zip _____ Country _____

Telephone _____ Fax _____ E-mail _____

Mail or fax this form to:
CAP Ventures, One Snow Road, Marshfield, MA 02050
Fax: (617) 837-8856 • To order by phone call: (617) 837-7200

MARK YOUR CALENDAR!

ON DEMAND DIGITAL PRINTING & PUBLISHING STRATEGY CONFERENCE AND EXPOSITION

Only One Event Will Show You the Future of Printing

June 27 - 29, 1995

**Jacob K. Javits Convention Center
New York, New York**

For additional information please contact:
Expocon Management Associates, Inc.

Call, Fax, Or E-mail Us To Find Out More About Events & Companies Mentioned In This Issue

The Gilbane Report on Open Information & Document Systems is published 6 times a year.

The Gilbane Report is an independent publication offering objective analysis of technology and business issues. The report does not provide advertising, product reviews, testing or vendor recommendations. We do discuss particular pieces of product technology that are appropriate to the topic under analysis, and welcome product information and input from vendors.

Letters to the editor are encouraged and will be answered. Mail to Editor, The Gilbane Report, CAP Ventures, One Snow Road, Marshfield, MA 02050 or frank_gilbane@capv.com

(203) 256-4700, ext. 139
(203) 256-4730 FAX