

International Federation of Library Associations and Institutions
**UNIVERSAL DATAFLOW AND TELECOMMUNICATIONS
CORE PROGRAMME**

OCCASIONAL PAPER

2

**OVERVIEW OF
DOCUMENT MANAGEMENT
TECHNOLOGY**

Gary Cleveland
National Library of Canada



June, 1995

**International Federation of Library Associations and Institutions
UNIVERSAL DATAFLOW AND TELECOMMUNICATIONS
CORE PROGRAMME**

The IFLA Core Programme on Universal Dataflow and Telecommunications (UDT) seeks to facilitate the international and national exchange of electronic data by providing the library community with pragmatic approaches to resource sharing. The programme monitors and promotes the use of relevant standards, promotes the use of relevant technologies and monitors relevant policy issues in an effort to overcome barriers to the electronic transfer of data in library fields.

CONTACT INFORMATION

Mailing Address:

*IFLA International Office for UDT
c/o National Library of Canada
395 Wellington Street
Ottawa, CANADA
K1A 0N4*

UDT Staff Contacts:

Leigh Swain, Director
Email: leigh.swain@udt.ifla.org
Phone: (819) 994-6833

or

Louise Lantaigne, Administration Officer
Email: UDT@udt.ifla.org
Phone: (819) 994-6963

Fax: (819) 994-6835

Email: UDT@udt.ifla.org

URL: <http://www.ifla.org/udt/>

Occasional papers are available electronically at: <http://www.ifla.org/udt/op/>



Overview of Document Management Technology

Gary Cleveland

National Library of Canada
gary.cleveland@udt.ifla.org

June, 1995

INTRODUCTION

Current document management technology grows out of the business community where some 80% of corporate information resides in documents. The need for greater efficiencies in handling business documents to gain an edge on the competition has fueled the rapid development of Document Management Systems (DMS) over the last two years. Document management has replaced data management—the focus of computing for the last twenty years—as the latest challenge facing information technologists. This paper provides a general overview of document management, its associated standards, trends for the year to come, and prominent vendors of document management products.

WHAT IS DOCUMENT MANAGEMENT?

Document management is the automated control of electronic documents—page images, spreadsheets, word processing documents, and complex, compound documents—through their entire life cycle within an organization, from initial creation to final archiving. Document management allows organizations to exert greater control over the production, storage, and distribution of documents, yielding greater efficiencies in the ability to reuse information, to control a document through a workflow process, and to reduce product cycle times. The full range of functions that a document management system may perform includes document identification, storage and retrieval, tracking, version control, workflow management, and presentation.

Traditionally, there have been two classes of document management: 1) management of fixed images of pages (the class that seems to be most familiar to librarians); and 2) management of editable documents, such as word processing files and spreadsheets. These two classes differ largely in the fact that images are static, while editable documents are dynamic and changing. The functions associated with the two classes differ as well. Systems supporting images focus on access, with input, indexing and retrieval as important functions, while systems supporting editable documents focus on creation, with joint authoring, workflow, and revision control at the center.

The barriers between these two classes of systems, however, are breaking down. Vendors are moving away from specializing exclusively in one class or the other, with a trend toward creating larger, integrated document management systems that incorporate a full range of document management functions. Such systems control the creation and use of documents throughout their life span—across platforms, applications, and company organizational units.

It is important to note that document management is not yet a single technology, but several. The major challenge at this time is the integration of several software packages—those for image storage and retrieval, workflow management, compound document management, and document presentation—into a single integrated system. To facilitate this process, vendors of DMSs are forming alliances and creating common standards to provide an open approach to the technologies (see Standards section below for more information).

In the longer term, experts in the field predict that document management functions will cease to be implemented as separate, dedicated applications. They instead will be incorporated as basic tools of operating systems, much like current file access mechanisms.

ELEMENTS OF A DOCUMENT MANAGEMENT SYSTEM (DMS)

The elements of a DMS include software to perform all functions necessary to manage the document across an organization from cradle to grave. Each element is described below.

Underlying infrastructure. While not part of an application per se, an appropriate underlying infrastructure is nevertheless a prerequisite to supporting a DMS. The infrastructure is the set of desktop computers, workstations, and servers that are interconnected by LANs and/or WANs. It must have characteristics such as network operating system independence, file format independence, location independence, long file names, and link tracking.

Authoring. Authoring tools support document creation. Some more sophisticated tools support structured or guided authoring, where authors are constrained by the system to enter data in specified ways. Typically, they are interfaced with DMSs in order to capture document metadata at the time of creation and revision.

Workflow. Workflow is defined as the coordination of tasks, data, and people to make a business process more efficient, effective, and adaptable to change. It is the control of information throughout all phases of a process. The path of a particular document is determined by the document type (e.g., press releases, manuals, policy papers, memos), the processes governing a document, and organizational roles (i.e., who has the authority to see what?). It supports functions such as authoring, revising, routing, commentary, approval, conditional branching, and the establishment of deadlines and milestones.

Workflow is a central aspect of document management because it allows organizations to get control of, and increase the efficiency of, the flow of documents that support their business. Typically,

workflow has been implemented in separate software packages, but it is beginning to be incorporated into large integrated DMSs.

Storage. The core of the DMS is the database and search engines supporting storage and retrieval of documents. Traditionally relational, DMSs are moving toward object-oriented databases. However, most vendors are now using mixed databases with relational databases used to point to information objects. Such databases are called Object-Relational Database Management Systems (ORDBMS).

Library services. Not to be confused with what librarians consider to be library services, this is a term used specifically by the document management community to refer to document control mechanisms such as checkin, checkout, audit trail, protection/security, and version control.

Presentation/distribution services. Presentation and distribution concerns the form and manner in which users are provided with information. DMSs should allow "multipurposing" where information can be distributed in different formats, such as viewed on a network (e.g., the Web), distributed on CD-ROM, or printed on paper. Businesses can reuse information, putting it into a format determined by the target market or business function. On-demand printing, where a document is printed when it is needed from a document database, is growing in popularity and importance.

TRENDS IN ELECTRONIC DOCUMENT STRUCTURES

The trend in document management is away from the management of static documents toward complex, compound documents. In general, electronic documents fall on a continuum with static documents at one end and complex, compound documents at the other. Static documents, such as digital images, are the least flexible—they cannot be edited or made machine readable without further processing (i.e., optical character recognition). Next along are static, but editable documents, such as word processing documents and spreadsheets. While modifiable, these documents are: a) tied explicitly to one application (e.g., WordPerfect); b) considered to be "dumb" in that they contain little or no information about themselves; and c) typically flat

files, or information blobs, prohibiting access of specific information elements within them.

Complex, compound documents at the far end of the continuum, however, exhibit none of the above characteristics. They are not tied to one application or platform; they are dynamic, constantly in a process of change; and they are "intelligent", carrying information about their content and structure. In this way, documents are reflecting the trend toward object-oriented architectures, where information is contained in objects—units of information of a finer granularity than traditional documents—which also contain information about themselves and their originating applications.

Documents in this new object-oriented conceptualization are considered to be containers of a wide variety of information, rather than single flat files or blobs. Instead, they are simply a collection of pointers to external elements that are dynamically assembled as they are retrieved. For example, they contain pointers to:

- **information objects**, the actual content, assembled temporarily for a specific purpose and which can be later reused in other documents;
- **information about document behaviours**, such as who is allowed to see the document, who must approve it and in what sequence—essentially, information about the workflow process of the document;
- **metadata**, such as its author(s), revision history, and status (e.g., draft or final copy);
- **links to other external elements**, such as datasets, graphics, images and fonts.

Another way of looking at a compound document is as a "bill of materials" used to assemble a specific representation of information on the fly. The bill of materials contains information objects, and information about how the objects are put together. The resulting compound documents can be delivered as one of two types: 1) pre-specified documents that are constructed in the same way each time; or 2) virtual documents that are assembled for a specific purpose, that exist temporarily, and which may change from one viewing to another.

WHY USE DOCUMENT MANAGEMENT SYSTEMS?

Vendors and representatives of companies that have implemented DMSs, repeatedly stress the need to examine business-critical objectives before embarking on DMS projects. For businesses, such as engineering, insurance, and pharmaceutical firms, the goal of automating document management is to get control of and increase the efficiency of the flow of documents that support their business—producing aircraft or automobiles, processing insurance claims, creating and getting approval for new drugs. The benefits of document management for businesses include:

- leveraging intellectual capital such that knowledge is created once, then reused many times
- managing workflow, controlling the flow of information through all phases of a process
- fostering more effective teamwork to accelerate business-critical applications
- delivering products faster with better customer service (reduce cycle times)
- allowing rapid response to events

ISSUES FOR LIBRARIANS

The trend toward compound documents and large, integrated document management systems raises serious issues for librarians. Specifically, how are librarians to deal with virtual documents, and what is the proper use of a document management system in libraries?

For librarians, the idea of a virtual document—a document that exists only briefly, and which changes with each particular viewing, and with each particular viewer—presents a nightmare scenario. Yet, it is a good example of where current thought on document management in business does not map directly onto library conceptions of what a document or a document database might be. For business, the goal is not necessarily to preserve a particular document instance, but to obtain, on demand, information needed to support business-critical objectives. The emphasis is placed on time-sensitive information provided on-demand. In the high-end DMS world, a document is not the static entity with which librarians are familiar, but a particular view of

an information base, at a particular time, to a particular user.

When considering the use of DMSs in libraries, questions that arise include: how well does the prevailing business model of document management – controlling proprietary documents as a means to an end, not the end in itself—map onto library models? What are the business-critical objectives of libraries in general, and of the National Library in specific? What are the documents to be managed? Who are the clients? Are the documents internal documents (e.g., memos and reports), publications emanating from the National Library, or documents of which we are the custodians but for which we do not necessarily hold copyright? With the failure rate of document management projects estimated at around 50%, these are extremely important questions to answer before embarking on any document management project.

DOCUMENT MANAGEMENT STANDARDS

As with any system for information management, there is a large number of complementary and competing standards that describe the architecture of DMSs. Some of the more prominent standards and standards initiatives include:

ODMA (Open Document Management API). ODMA is an industry-lead effort to ease cross-platform and cross-application communication by standardizing access to document management clients through a published Application Programming Interface (API). The ODMA API allows applications (word processors, spreadsheets) to communicate with a DMS in a flexible manner without the need of a hard-coded link between the application and the DMS. Thus, ODMA allows multiple applications to access the same DMS.

Partners in the ODMA endeavor include Borland, Documentum, Interleaf, Novell, Oracle, PC DOCS, SoftSolutions, Sybase, WordPerfect, and XSoft.

Shamrock. Another industry-lead effort is the Shamrock Document Management Coalition. The Shamrock Coalition will specify a software layer – called “middleware”—to be inserted between applications and document servers, providing a common set of tools for document security, administration, and access. It essentially provides a

gateway between applications and multiple, disparate document repositories. Where ODMA allows multiple applications to access the same DMS, Shamrock allows applications to access multiple DMSs.

The central partners of Shamrock are Saros and IBM, with most of the technology for Shamrock based on Saros’ proprietary DMS engine. Additional partners include Adobe, EDS, Frame Technology, Hewlett-Packard, Microsoft, Verity, ViewStar, Wang, as well as ODMA members Documentum, Interleaf, PC DOCS, Sybase and XSoft.

DEN (Document-Enabled Network). A competing standard of Shamrock is DEN. DEN is a partnership between Novell and Xerox to create middleware to make it easier for developers to create applications for networked document management. There is currently talk underway between the developers of Shamrock and DEN to combine their efforts and merge the two standards.

OPENDOC is a “component software architecture” that provides interfaces allowing independently-written software to work together in a single document.

OLE (Object Linking and Embedding). This proprietary standard from Microsoft allows objects in one application to be linked to objects in another. For example, a graph in a word processing document can be linked to the original data in a spreadsheet application. When the data in the spreadsheet changes, the graph in the document is automatically updated to reflect those changes. OLE allows applications to share data as well as the functionality of the originating applications.

OGM (Object Management Group). This industry consortium is producing a framework for specifying object-oriented environments. The framework itself is called OMA, or Object Management Architecture. It specifies, in part, a reference model that classifies the components, interfaces and protocols of an object-oriented system.

STATE OF THE DOCUMENT MANAGEMENT INDUSTRY

Much technology has appeared in the last two years to support document management. In fact, so much

has appeared that the current problem facing organizations is the integration of document management technologies (authoring, workflow, presentation). The overall trends in document management, according to Frank Gilbane, editor of The Gilbane Report on Open Information and Document Systems, are:

- Most managers seeking document solutions are looking for compound document management, electronic and paper delivery, document image management, and workflow.
- Object-orientation is the trend, but most DMSs are relational.
- Advanced document viewers are appearing.
- On-demand printing is growing rapidly.

Predictions for 1996 include:

- Continued convergence of technologies and suppliers of DMSs.
- Growth of "general purpose" DMSs. Typically, DMSs are designed for specific applications. The trend, however, is toward general-purpose DMSs that can be used equally well by different sections of one company and by companies in different sectors.
- SGML will become more mainstream due to the widespread use of HTML and the support for HTML in Microsoft and WordPerfect products.
- Multimedia will not yet become part of corporate business documents. A catalyst is needed (such as voice converging with email).

DOCUMENT MANAGEMENT SYSTEM VENDORS

A growing number of companies are providing DMSs. The companies vary widely in the number of DMS functions they implement and in the manner in which the functions are implemented. Major companies providing DMSs include:

- Documentum
- Electronic Book Technologies
- Folio Corporation
- Frame Technology
- Fulcrum Technologies
- Information Dimensions, Inc.
- Interleaf

- Open Text Corporation
- Oracle Corporation
- PC Docs, Inc.

NOTES
