



## Technical White Paper

### A Growing Problem

Today's information age is accelerating at quantum speed. Advances such as the Internet and high-speed networks have propelled the never-ending quest for information. An infinite amount of information is being created and accessed by people around the globe everyday. We live in an information intensive society.

Unfortunately there remains a significant amount of information inaccessible to those most in need. Pertinent information is stored in varying systems, in various media, employing various technologies, all across the globe. What is lacking is a content management framework – highly scalable information architecture capable of managing infinite amounts of data of all types.

A clear example of this is the healthcare sector. Healthcare today remains one of the most information intensive and least automated of all industries. It remains a "paper world", in which most institutions have implemented a central registry, or vault-like process, in which the "official" records are maintained.

In fact, much of what is maintained in the healthcare institution's central registry is generated by computer based systems, yet the printed version is the official record. The business problem here is often the time delay experienced in the continued efforts to maintain an accurate paper based record, such as a patient's medical file in an environment of constant updates.

Furthermore, this approach is being replicated by each institution, with no opportunity for electronic sharing of clinical data. Consider this question: "Where is your medical record?" The answer: "Your medical record is scattered across the country". It is composed of little pockets of information that were formed wherever you interacted with the healthcare system.

Electronically generated information covers traditional desktop information, such as word processing documents and spreadsheets, but also document imaging, desktop publishing documents, voice, video and other industry specific multimedia data types, such as X-rays. Until the advent of the MAINSOURCE™ technology there has been no efficient way to manage either these new types of data or the volume of the legacy data.



## A Unique Solution

MAINSOURCE's™ state-of-the-art distributed object database architecture is capable of acquiring, securing and delivering information. This unique architecture provides unmatched scalability and an unlimited capacity while streamlining management of information over the web. The MAINSOURCE™ approach represents the first time that object database techniques have been applied in any consistent scalable manner to Information Technology and the Internet, and promises to revolutionize the way information is managed and delivered.

Developed inside Digital Equipment Corporation (now part of Compaq Corporation), with a significant investment, MAINSOURCE's™ platform independent software kernel utilizes strategic technologies to deliver the next generation content management framework.

MAINSOURCE's™ next generation information management solution includes the following features:

- Content management for any digital data type
- Transaction processing
- Distributed architecture with dynamic caching
- Utilizes existing web servers and browsers
- Unlimited user and data storage capacity
- Built-in version control
- Support for renditions
- Automatic customization
- Dynamic profiling
- Comprehensive object security
- Easy integration via open interface through an API
- Data replication and backup
- Performance/statistics collection
- Simplified change control



## The MAINSOURCE™ Goals

The MAINsuite™ product family was developed to address the following specific goals:

- Implement a central registry that maintains a hierarchical filing paradigm.
- Handle multimedia BLOB data (Binary Large Object)
- Support distributed user data, maintained close to users who need to create and retrieve it
- Extend the paper document paradigm to an electronic multi-media document
- Implement effective security of the retrieval of organizational data, permitting access to only those users with the appropriate privileges
- Employ “enabling” technologies such as optical storage, client-server architectures and high-speed networks to alleviate the paper problem faced by organizations
- Integrate to new and existing applications

## Technical Challenges

There are significant technical issues to be faced when implementing a solution to solve these goals. The major challenges are listed below, with a brief description on how they are solved within the MAINsuite™ product family:

- Management of complex data
- Scalability
- Security
- Control
- Integration to existing applications

## Management of Complex Data

Information management solutions must be able to handle a wide range of complex data types, ranging from traditional data such as word processing, to multimedia data such as voice or video. The MAINSOURCE™ software addresses this problem through the use of *object oriented methodology*. The platform stores the data it is passed in a set of objects which are used to identify and manage the information. These objects include pages, documents and folders.



In the MAINSOURCE™ software, all objects are maintained in their native form, that is they are not altered in any manner. When an object is requested by a client service, desktop or application, the object is retrieved and presented to that service which knows what methods can be applied to that particular object. For example, a business application requesting a document can view image objects (bit-map) or to view and edit word processing objects (file). Users can also launch the appropriate native word processing application (e.g. WordPerfect) to support the editing operation.

Complex data objects have many characteristics, many of them unique to each class of object. One attribute that is quite universal is the large size of these objects. The common comparison used is one that compares the size of a single sheet of typewritten paper, created with a word processing package, to a scanned bitonal document image. This word processing created document would be roughly 2-2.5 KB in size. This same page, when scanned as a bitonal document image, would be roughly 50 KB in size, after compression. Suffice to say that the treatment of document images is quite different than simple word processing documents.

### **Scalability**

Many organizations, when implementing an information management solution, first want to implement in a controlled pilot environment, with a limited number of users. Unless the solution has been designed to grow to support an enterprise implementation, limitations in the network bandwidth can be a serious problem. In the scenario of potentially hundreds to thousands of users simultaneously requesting access to complex (large) object types, the network will quickly become the bottleneck, resulting in unacceptable response times to users. Often, this same network is shared with existing applications, so the impact will be felt in other areas as well.

Other areas which are critical in an enterprise information management implementation include managing disk access contention on the information manager. Like the network contention described above, hundreds of users can paralyze the information manager server databases, if not designed appropriately.

The MAINSOURCE™ software solves these problems through the use of a centralized Information Server combined with distributed Data Servers, across the network. The basic premise is that the management of the objects



(Information Server) is distinct from the actual objects themselves (Data Server).  
A

transaction processing architecture is used to manage high volume requests to the Information Server and Data Servers.

The Information Server acts as the central registry function, and maintains all the relevant information on the objects maintained within the information management framework. It performs the following functions:

- Maintain the page / document / folder (folders within folders) hierarchy
- Support multiple copies of a complex object data, distributed across the network, close to users, thus limiting network traffic
- Implement security, access controls
- Transaction processing

Complementing the Information Server are distributed Data Servers. These are distributed across the network, and actually manage the physical objects themselves. The Data Servers:

- Manage storage in both on-line and near-line media environments
- Handle magnetic to optical storage migration requirements
- Optimize access to near-line optical storage
- Support an unlimited combination of near-line and online storage

The MAINSOURCE™ architecture supports the scalability goal by allowing the addition of Data Servers as the number of users and data volumes increases, and by utilizing a transaction processing based Information Server.

## Security

Just as a central registry provides for a secure vault in the paper world, the same must be provided for in the electronic environment. The MAINSOURCE™ software:

- Breaks down the information class hierarchy to a page / document / folder structure
- Implements security at the object level
- Treats users of the system as objects (exactly as it treats data)
- Controls access by a combination of user object and data object attributes

Designed to handle thousands of simultaneous users, an administration utility provides a grouping capability allowing one or more roles to be established for



each user. This utility identifies assigned (or inherited) privileges, storage rights and other account parameters as well as provide for the encoding of data ownership and access rights.

Through the Application Program Interface ( API), the MAINSOURCE™ platform provides access to the user profile enabling applications to tailor their functionality as well as manipulate data object attributes according to the security rules. MAINSOURCE™ security features are a layer on top of the running network and operating system layer, thereby leaving the 'system level' security features intact.

## Control

As information is managed in a solution, there is a need for control of the updating of information within the central registry function. Besides security, which is described briefly above, two key features are supported within the MAINSOURCE™ architecture:

- An *object repository* maintains version control of objects. A user, given privileges, can request either the current version of an object, or a previous one
- A check-in / check-out capability manages cases where an object is requested by a user to be updated. While the original is checked-out, all other subsequent requests for that particular object will present a read only version

## Integration

Information management solutions can effectively exist as stand alone systems within organizations. In many cases, the information maintained in such an environment is vital to other applications.

The MAINSOURCE™ software features a suite of modular subsystems, all accessed through public APIs. Through these APIs, this platform can be integrated with the customer's existing applications. It has been a design goal to present the document management system as a "black box" to both the developer and/or application.



## The MAINsuite™ Architecture

In order to satisfy these requirements, MAINSOURCE™ has created a powerful, scalable solution for managing complex information content, and intelligently delivering it to the consumer in a customized, optimal fashion. It is a solution that is well-positioned and architected for Internet-based content delivery. The MAINsuite Tag is a core concept of this solution, and represents a differentiating factor.



Within the MAINSOURCE™ architecture, data is represented as objects and managed using logical references called MAINsuite Tags. This concept of *logically referenced* content is what gives MAINSOURCE™ the unique ability to manage any type of data, across any type of data server or storage device connected to the internet – and make it all appear to be a well structured, consolidated database.

Advanced indexing technology incorporated into the MAINsuite Tags support the reality of managing information on a truly global scale - with the capacity to cross-reference over 4 billion information servers, each of which is capable of identifying trillions of data objects. This distributed multi-server architecture also means that applications can be scaled to service millions of simultaneous users in a high-performance, high-availability configuration.

Within this distributed architecture, the MAINSOURCE™ *Level 7 Information Switches* act as intelligent 'content selectors', identifying the most appropriate source of the requested information managed by the object repository, and routing the request to one or more distributed Data Servers.

Information Switch selection criteria can include such things as the requestor's language, location, line speed, preferred density of graphic renditions and source data format. Data Servers manage the actual "data bits" and special media handling capabilities such as optical jukeboxes as well as links to legacy systems, external data repositories and dynamic, distributed data caches.



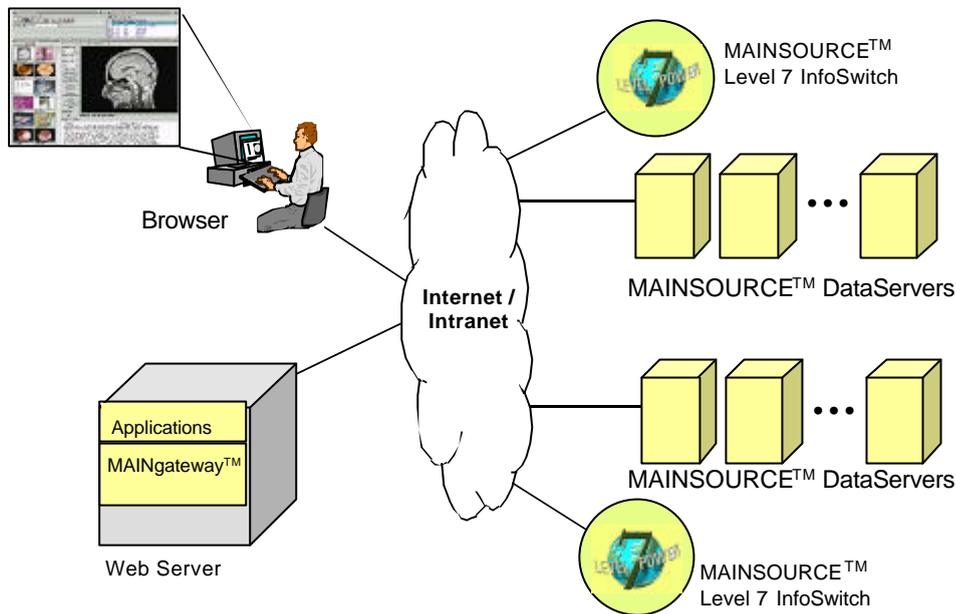


Through use of the MAINsuite Tag, the capabilities of MAINSOURCE™ can also be seamlessly integrated into third party applications, while ensuring that security

is consistently applied. Each and every object is wrapped with a standard security layer and MAINSOURCE's™ architecture provides for the inclusion of client specific security features such as encryption and additional passwords.

In short, the MAINSOURCE™ architecture provides a robust information management solution that:

- Scales to meet the heavy demands of today's information intensive environment
- Enforces the bullet-proof security that is demanded by organizations, consumers and legislation
- Simplifies planning and protects investments through unlimited growth capability
- Provides efficient management of complex data types including video, audio and image
- Supports geographic distribution of data
- Performs transaction processing
- Handles content revisions and rendition control
- Facilitates long-term archiving of data
- Integrates to new and existing applications



First level object indexes in MAINSOURCE's™ design, the MAINsuite Tags, allow for managing information on a truly global scale with the capacity to reference over 4 billion information servers each capable of identifying over 36 thousand trillion objects per year. Each and every object is wrapped with a standard security layer and MAINSOURCE's™ architecture provides for the inclusion of client specific security features such as encryption and additional passwords.

Data managed by this architecture can be transparently stored on near-line devices such as optical disk jukeboxes or tape silos to support an unlimited volume of storage, as well as retained off-line for long-term storage requirements.

In summary, our mission at MAINSOURCE™ is to provide the complete software solution to manage and consolidate enormous amounts of complex multimedia information and deliver it securely over the web. That is what our MAINsuite™ products are all about –providing the right information to the right person at the right time.