

# EXHIBIT B



US006236994B1

(12) **United States Patent**  
**Swartz et al.**

(10) **Patent No.:** **US 6,236,994 B1**  
(45) **Date of Patent:** **May 22, 2001**

(54) **METHOD AND APPARATUS FOR THE INTEGRATION OF INFORMATION AND KNOWLEDGE**

(75) Inventors: **Ronald M. Swartz**, Dresher; **Jeffrey L. Winkler**, Collegeville; **Evelyn A. Janos**, West Chester, all of PA (US); **Igor Markidan**, Cherry Hill, NJ (US); **Qun Dou**, North Wales, PA (US)

(73) Assignee: **Xerox Corporation**, Stamford, CT (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/106,335**

(22) Filed: **Jun. 29, 1998**

**Related U.S. Application Data**

(60) Provisional application No. 60/062,933, filed on Oct. 21, 1997.

(51) **Int. Cl.**<sup>7</sup> ..... **G06F 17/30**

(52) **U.S. Cl.** ..... **707/6; 707/101; 707/102; 707/104**

(58) **Field of Search** ..... **707/101, 6, 102, 707/104; 706/50, 59, 61**

(56) **References Cited**

**U.S. PATENT DOCUMENTS**

5,418,943 \* 5/1995 Borgida et al. .... 707/10  
5,644,686 7/1997 Hekmatpour ..... 106/45  
5,745,895 \* 4/1998 Bingham et al. .... 707/10

**OTHER PUBLICATIONS**

IEEE publication, "Enterprise Knowledge Management" by Daniel E. O'Leary, pp. 54-61, Mar. 1998.

IEEE publication, Knowledge-management Systems: Converting and Connecting, by Daniel E. O'Leary, pp. 30-33, May 1998.

Computer (Internet Watch), "Web-Based Knowledge Management" by Hermann Maurer, Austria, pp. 122-123, Mar. 1998.

IEEE publication, "Using AI in Knowledge Management: Knowledge Bases and Ontologies" by Daniel E. O'Leary, pp. 34-39, May 1998.

IEEE publication, "Web-Based Knowledge Management for Distributed Design" by Nicholas H.M. Caldwell, pp. 40-47, May 2000.\*

Villiers; "New Architecture for Linkage of SAS/PH-Clinical Software with Electronic Document Management Systems"; Revised Jun. 19, 1997; pp 1-7.

FileNet; FileNet's Foundation for Enterprise Document Management Strategy White Paper; pp 1-27, No date.

\* cited by examiner

*Primary Examiner*—Thomas Black

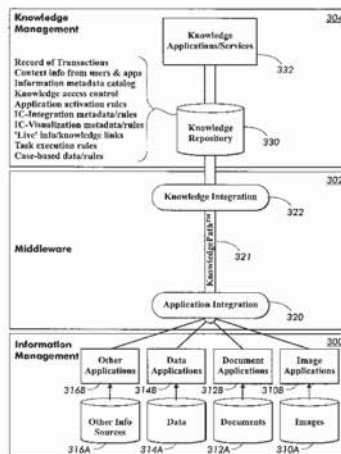
*Assistant Examiner*—Diane D. Mizrahi

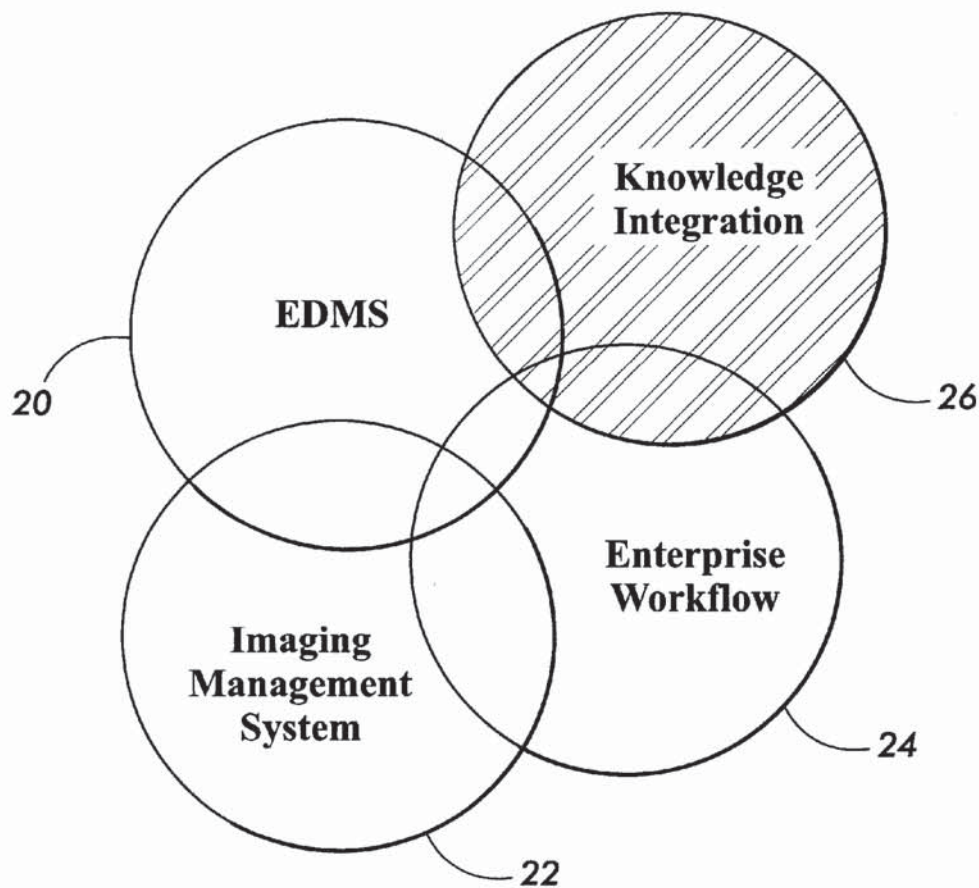
(74) *Attorney, Agent, or Firm*—William F. Eipert; Duane C. Basch

(57) **ABSTRACT**

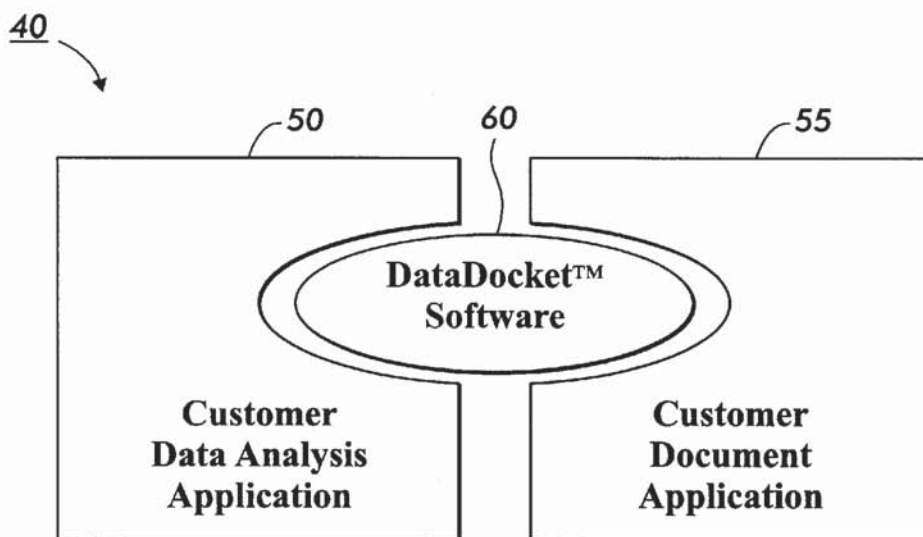
The present invention is a method and apparatus for first integrating the operation of various independent software applications directed to the management of information within an enterprise. The system architecture is, however, an expandable architecture, with built-in knowledge integration features that facilitate the monitoring of information flow into, out of, and between the integrated information management applications so as to assimilate knowledge information and facilitate the control of such information. Also included are additional tools which, using the knowledge information enable the more efficient use of the knowledge within an enterprise, including the ability to develop a context for and visualization of such knowledge.

**20 Claims, 17 Drawing Sheets**

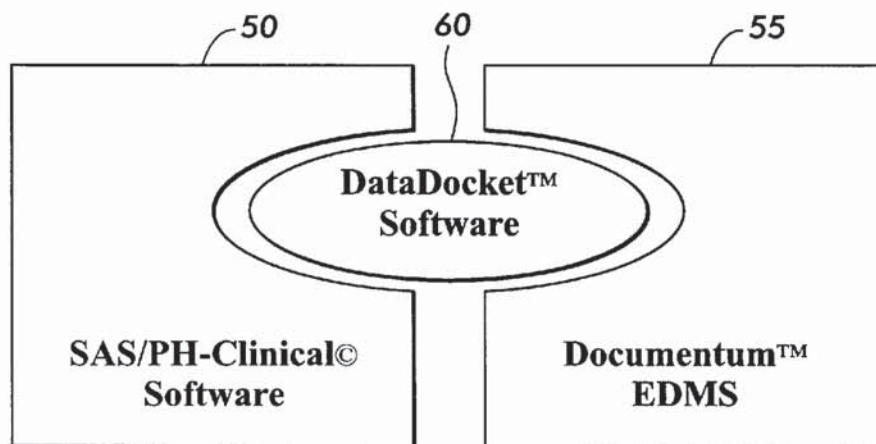




**FIG. 1**



**FIG. 2A**



**FIG. 2B**

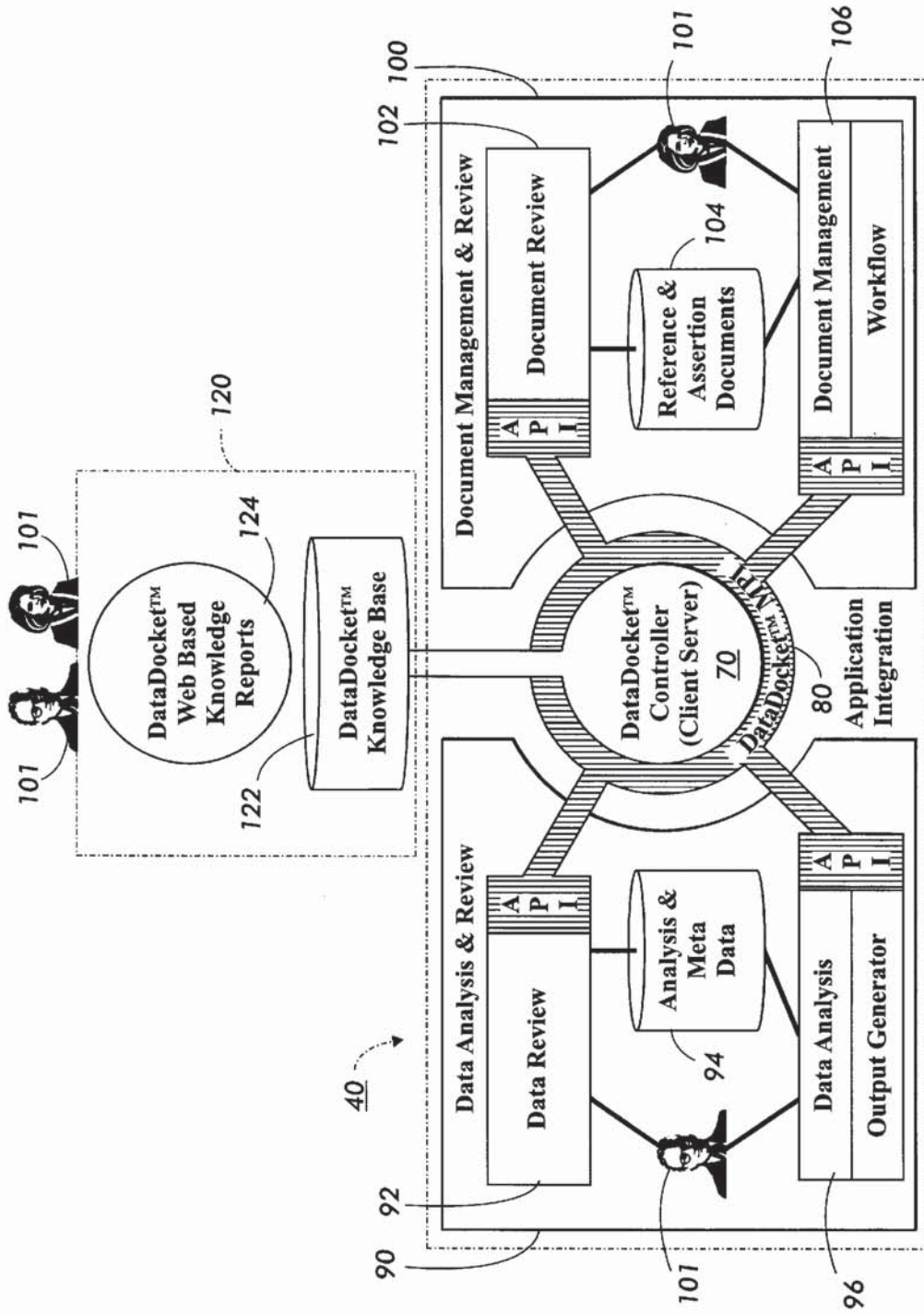


FIG. 3



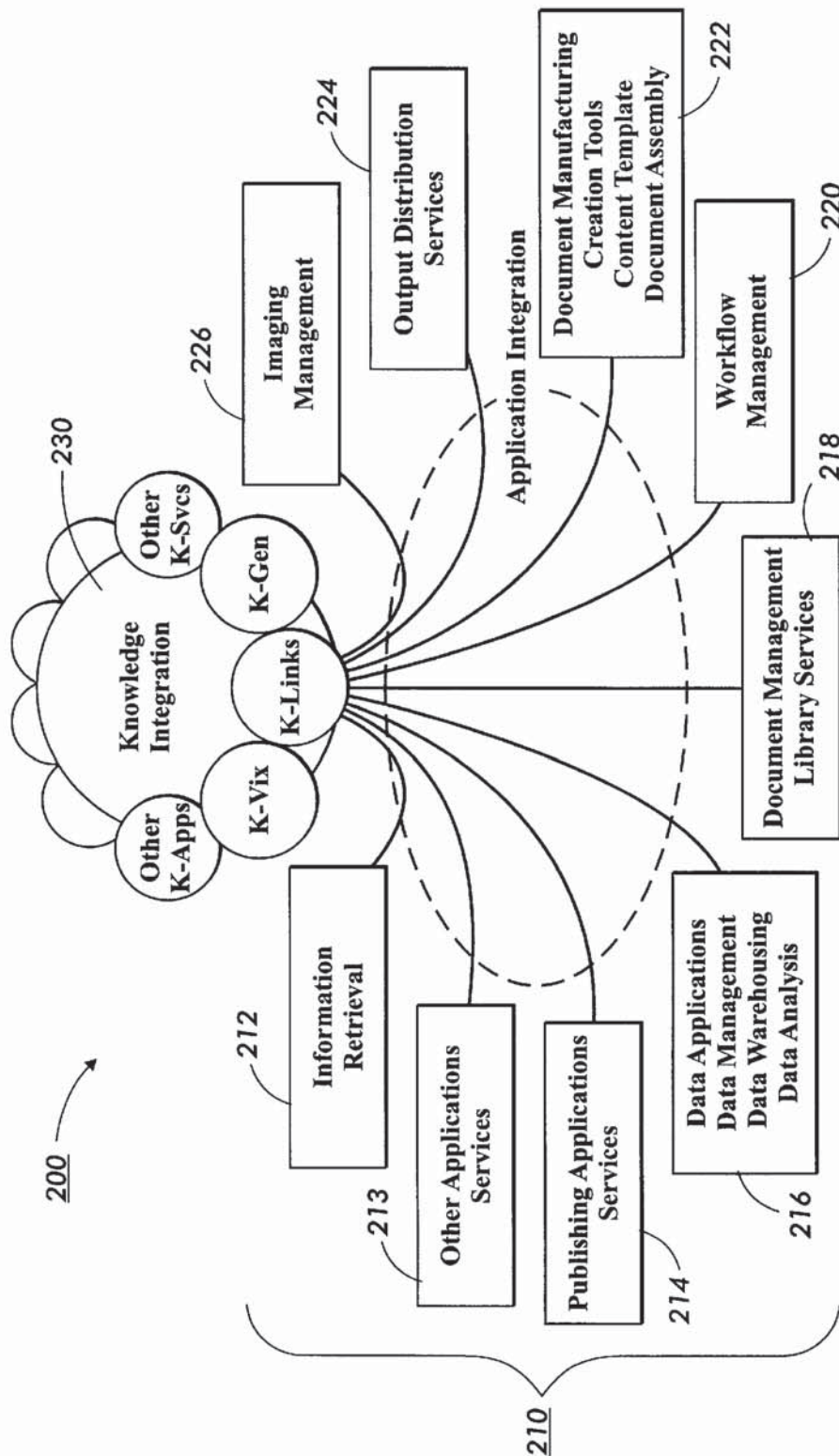


FIG. 4

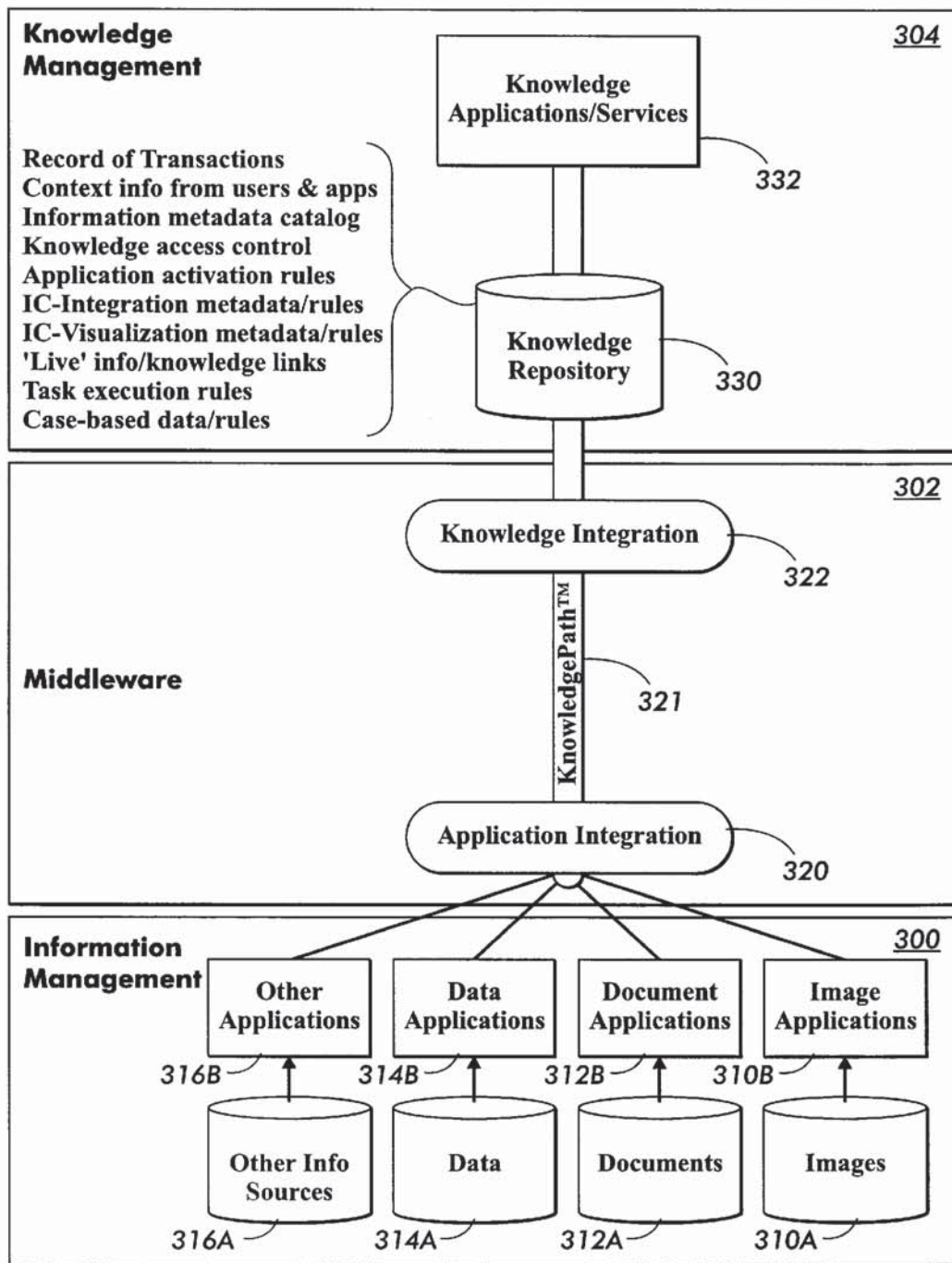


FIG. 5

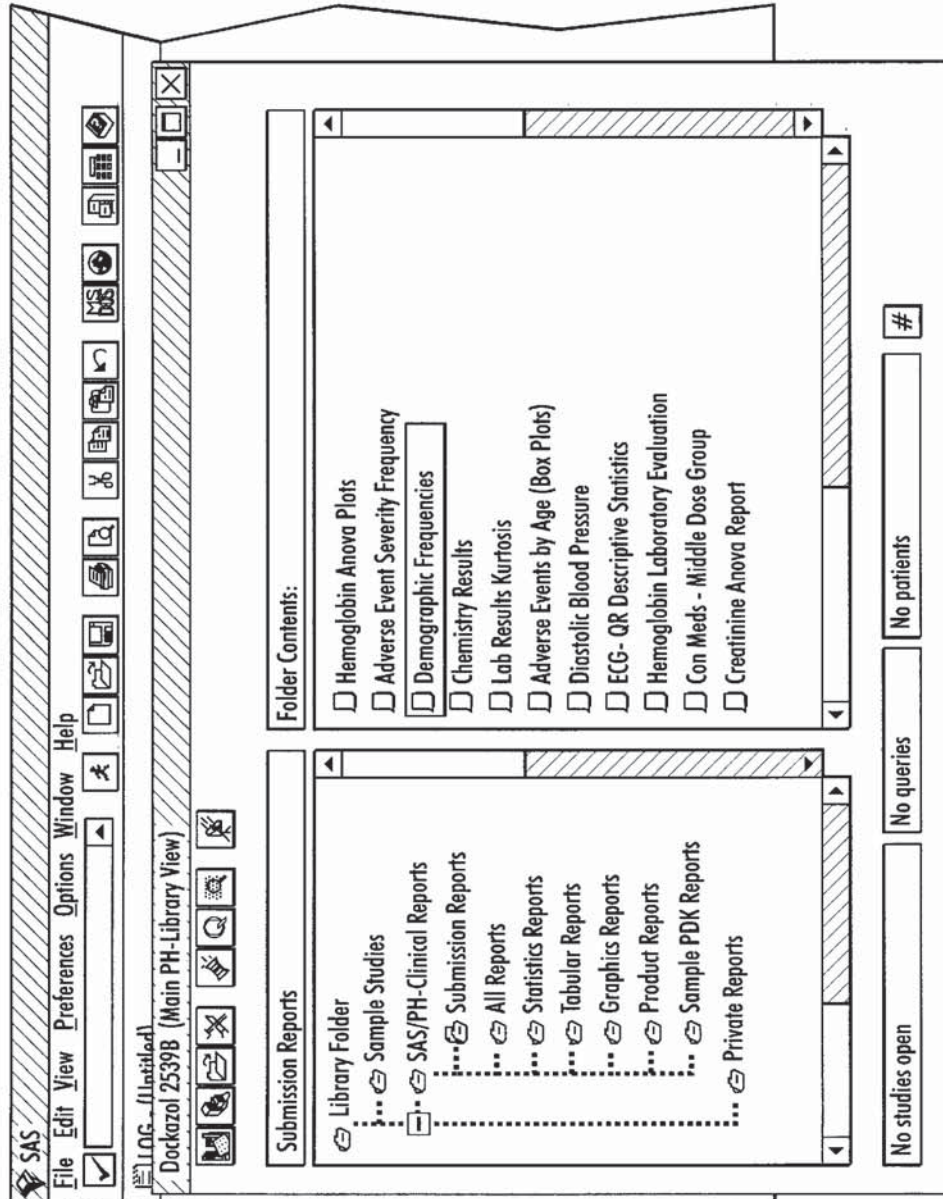


FIG. 6



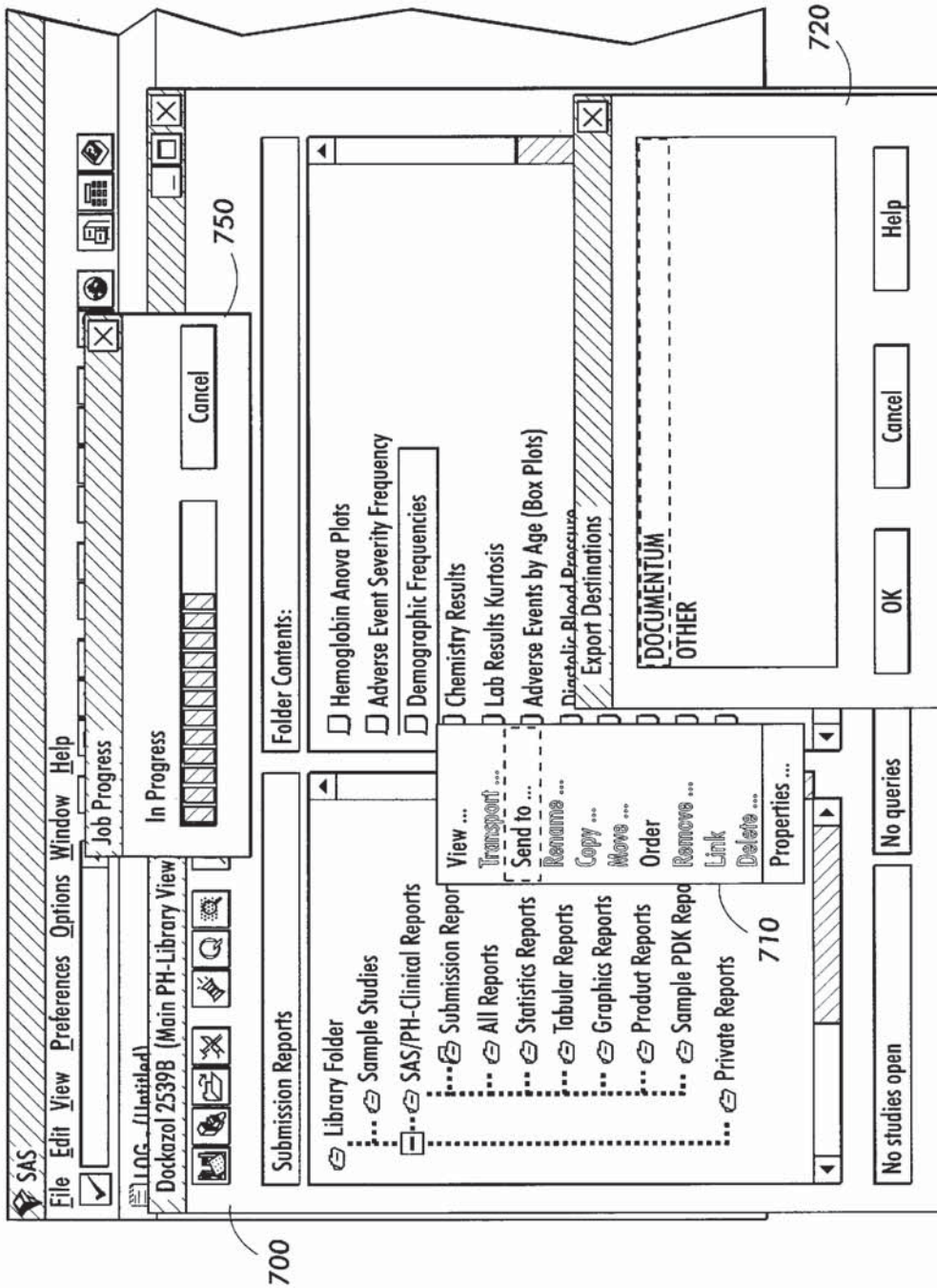


FIG. 7

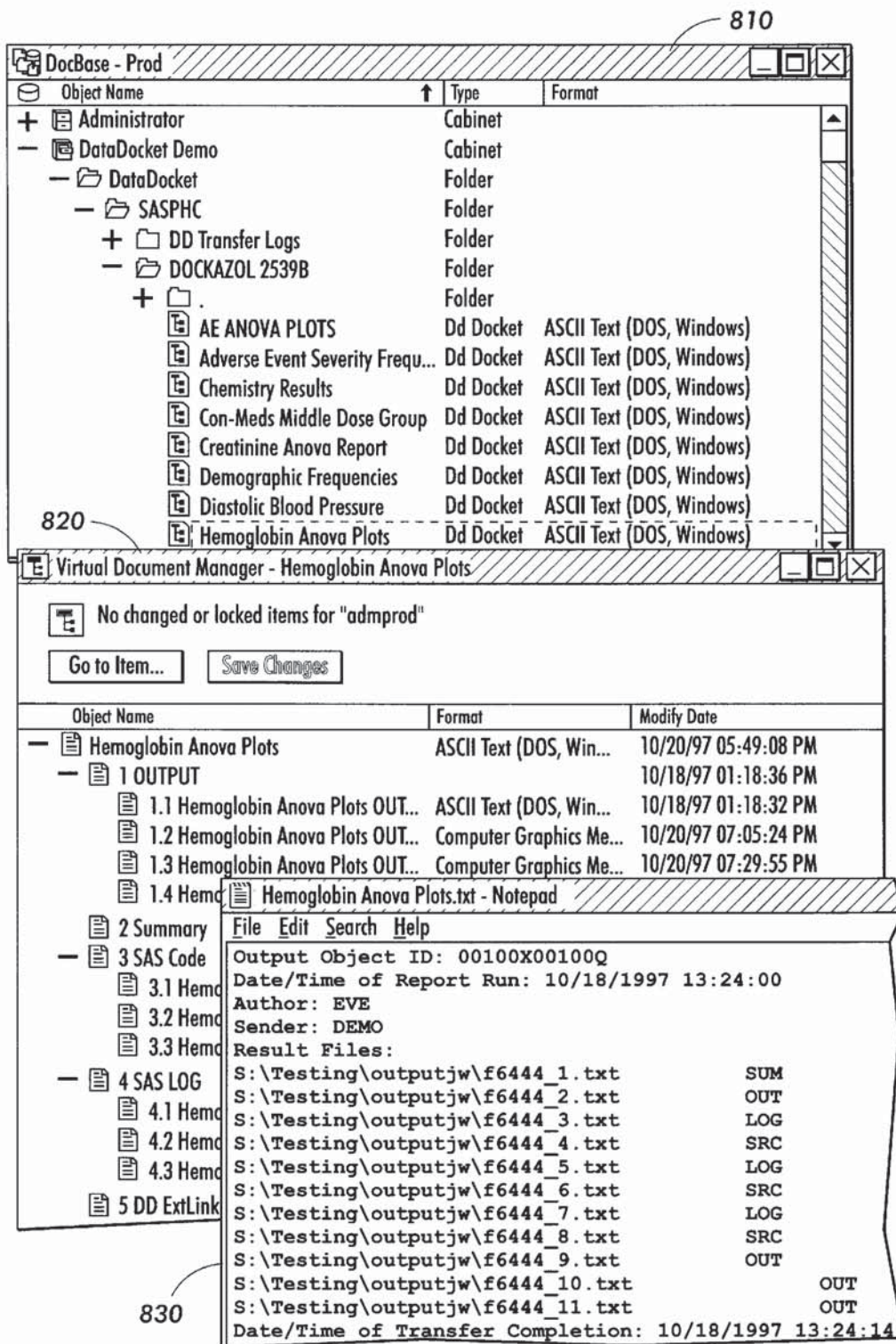


FIG. 8

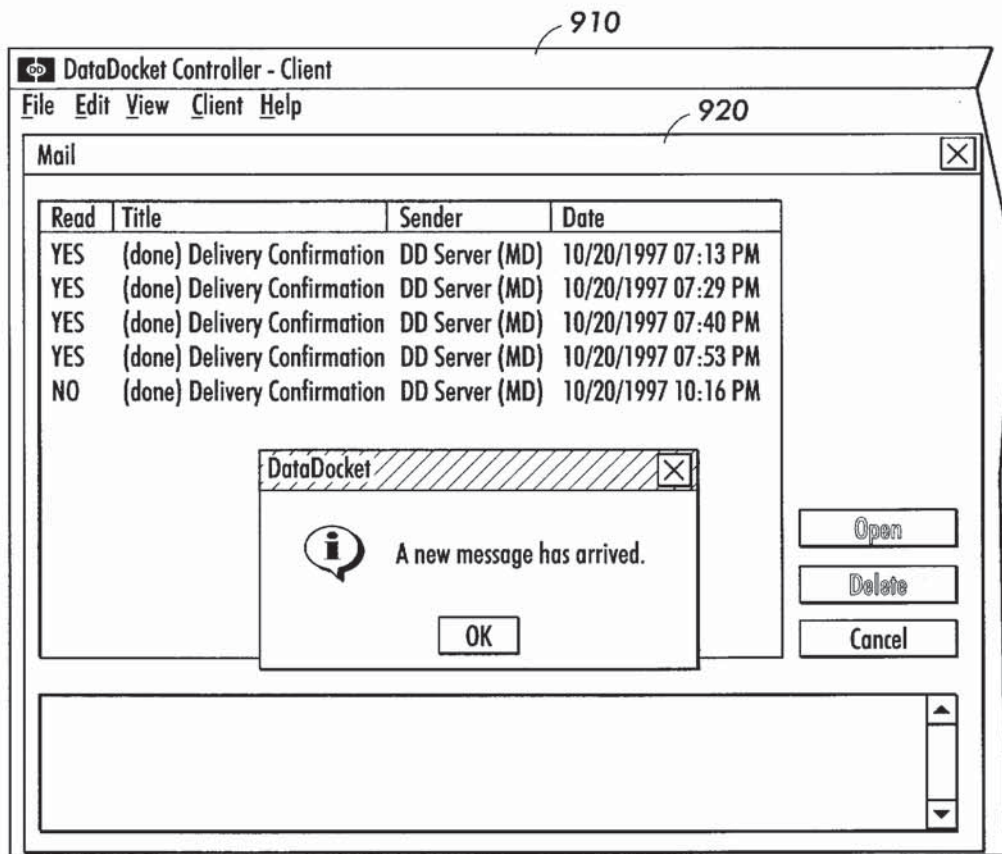


FIG. 9



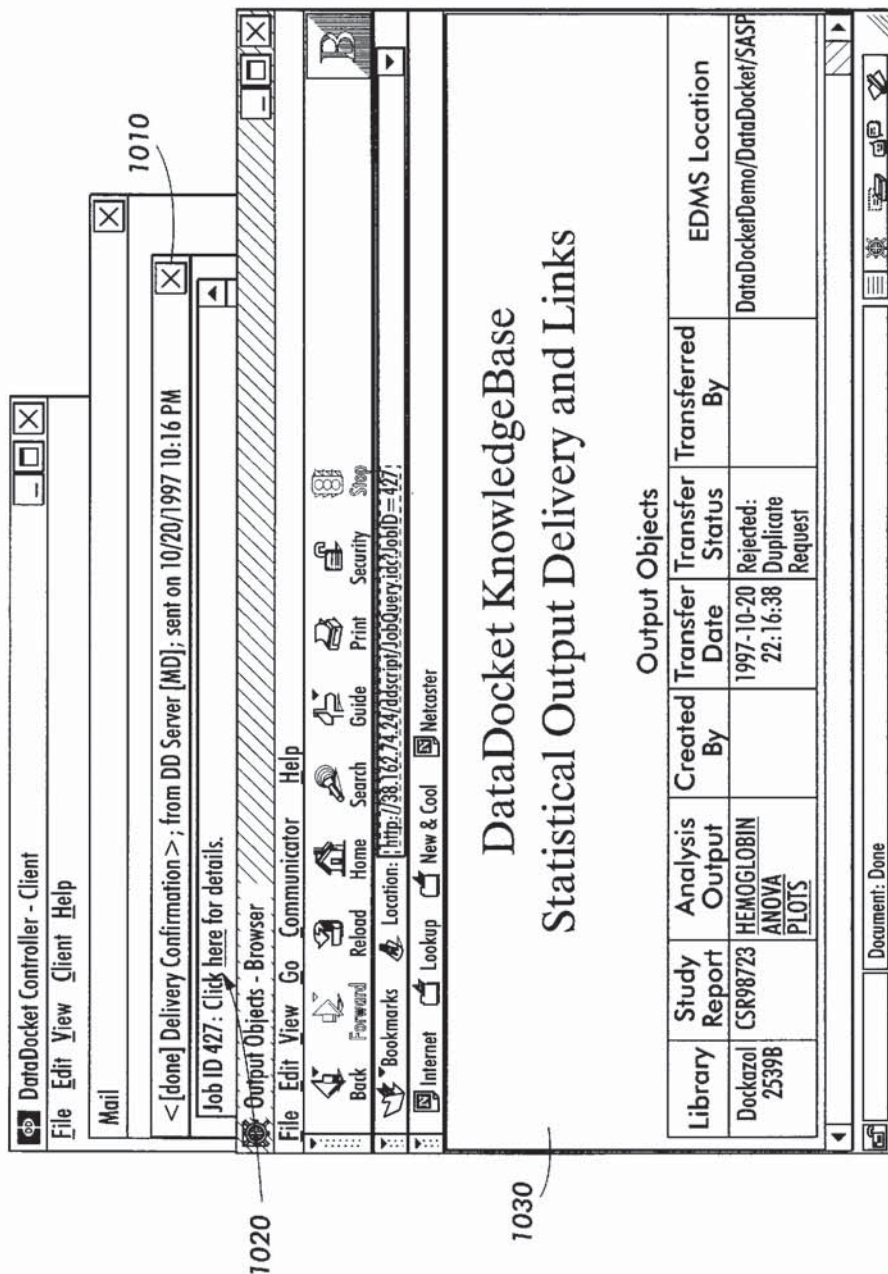


FIG. 10

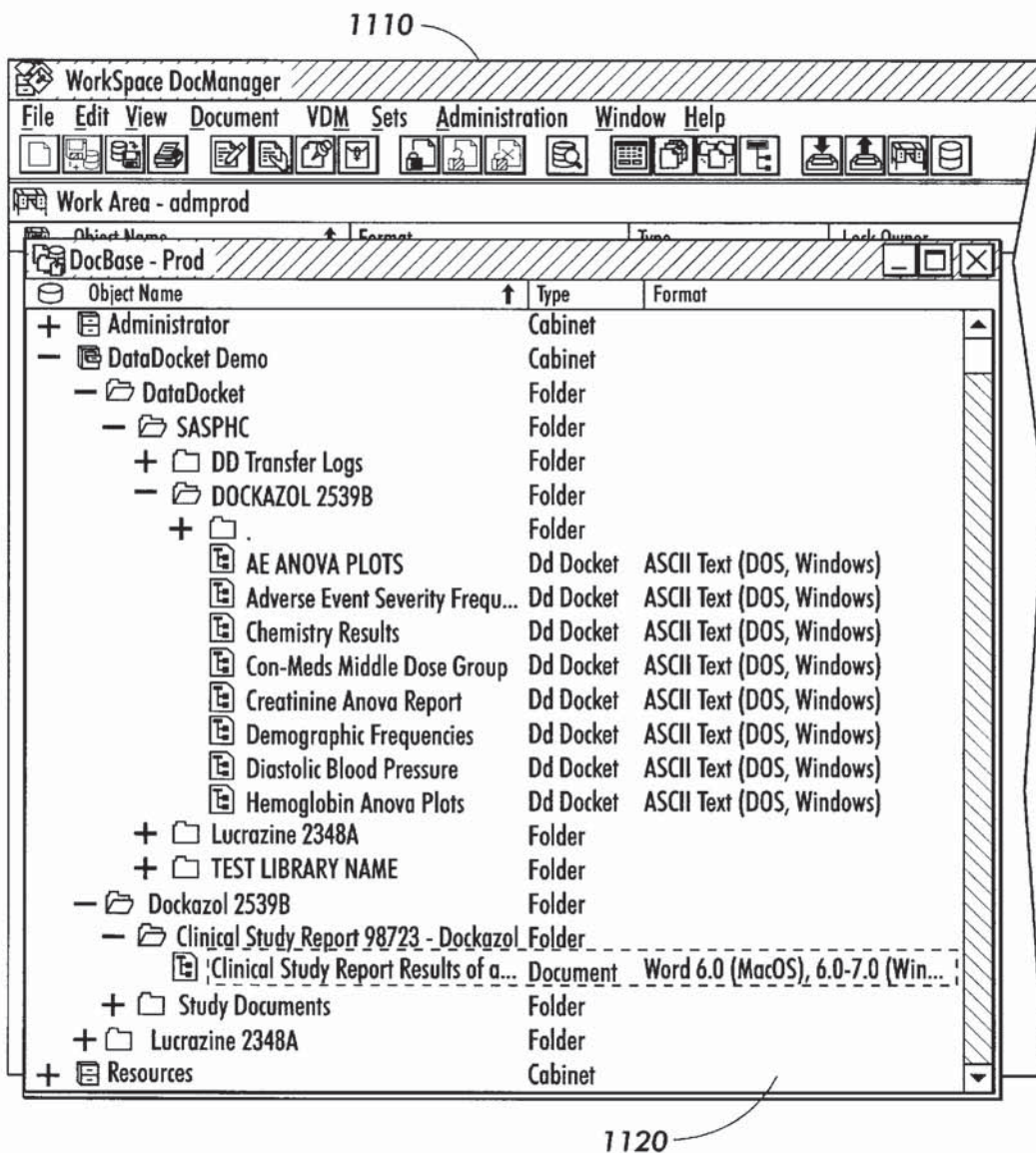


FIG. 11



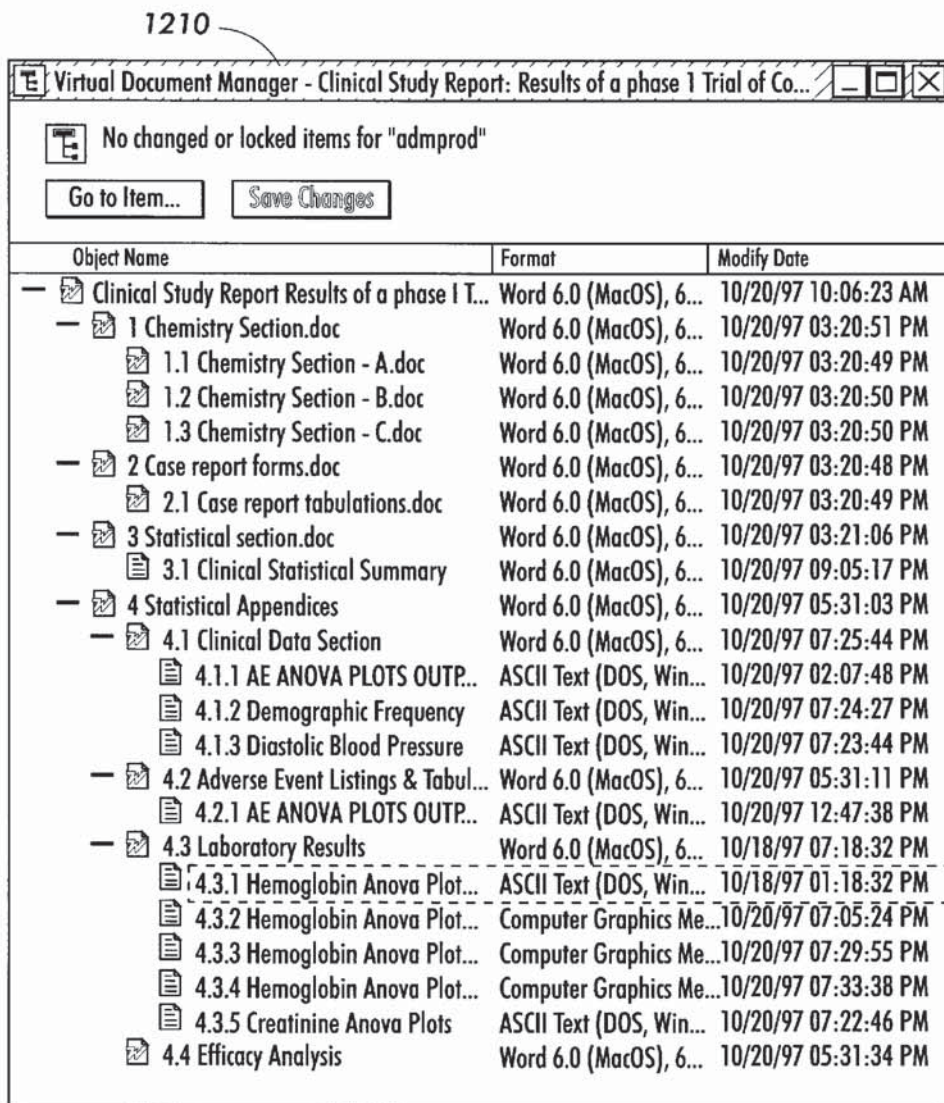


FIG. 12

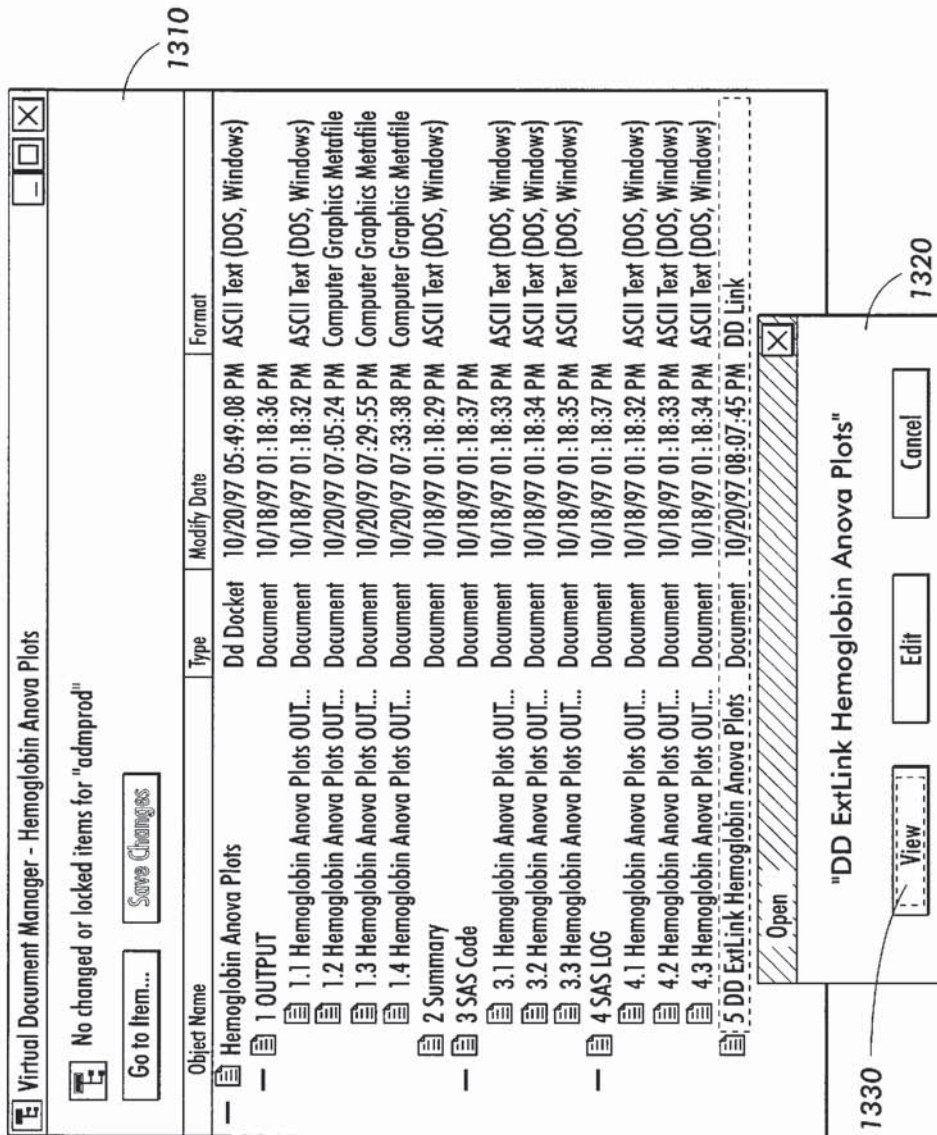


FIG. 13

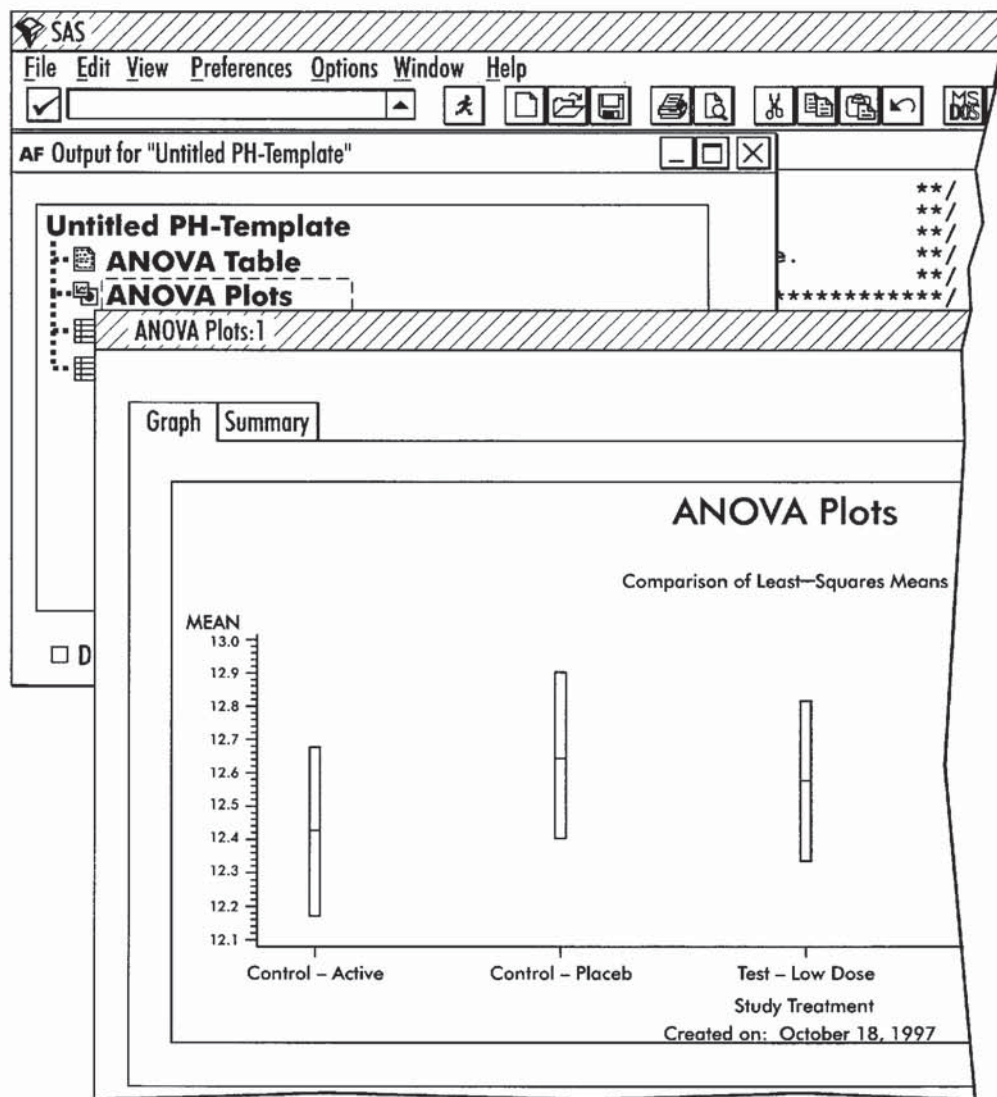


FIG. 14

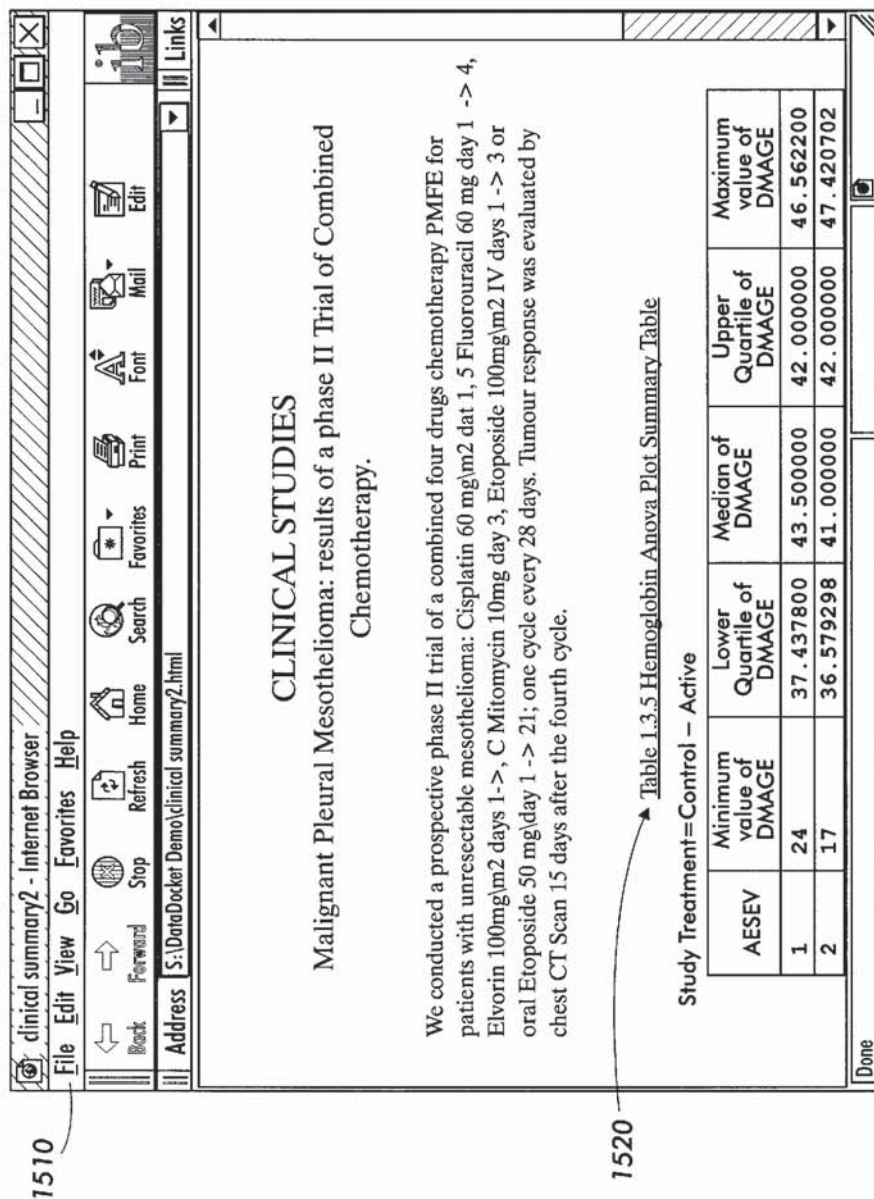


FIG. 15



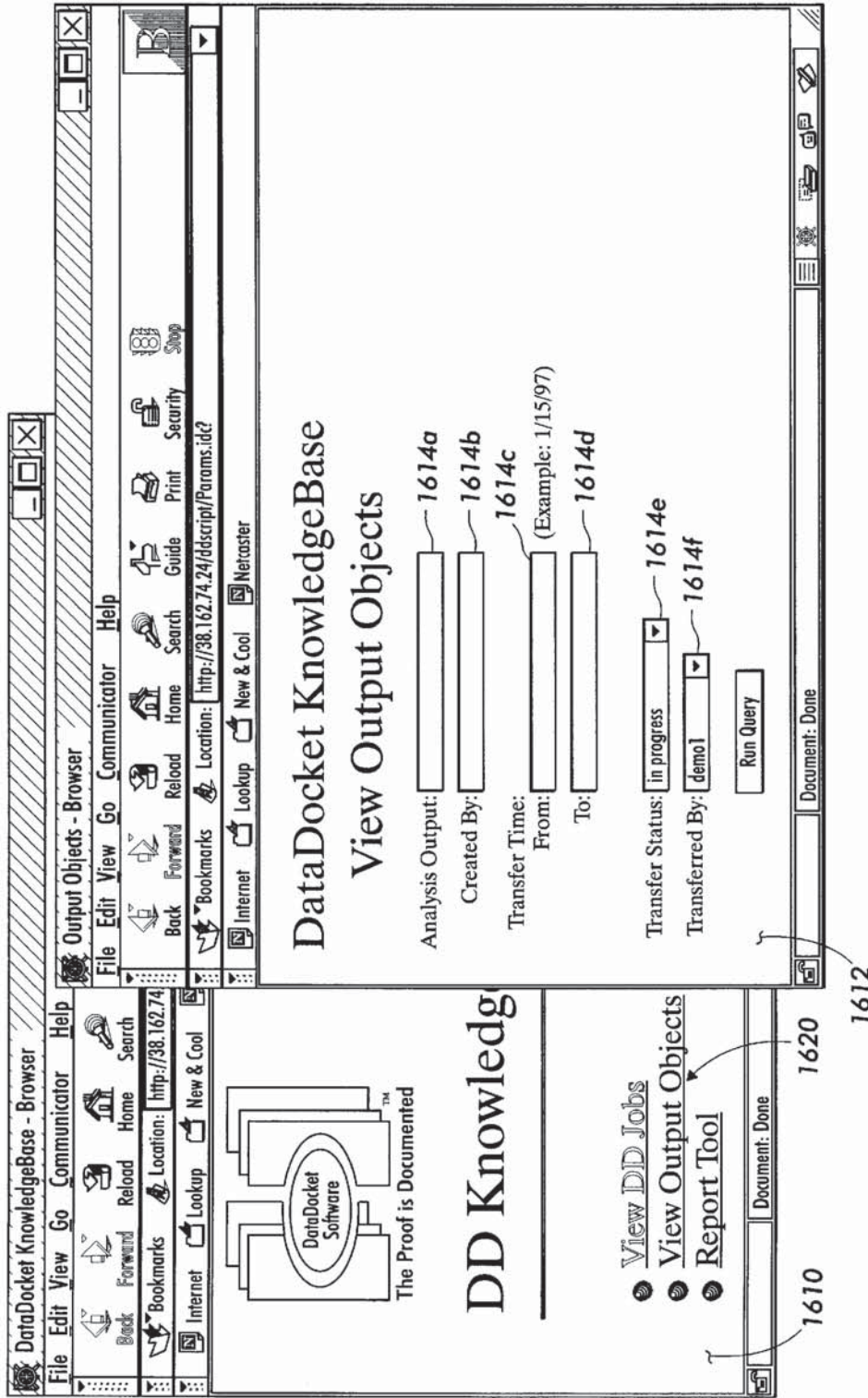


FIG. 16



The screenshot shows a web browser window with the following elements:

- Browser Title Bar:** Output Objects - Browser
- Menu Bar:** File, Edit, View, Go, Communicator, Help
- Address Bar:** Location: p://38.162.74.24/ddscript/DDOutputQuery.IDC?JobName = &Author = &SubmitTime1 = &SubmitTime2 = &Status = &UserNam
- Navigation Buttons:** Back, Forward, Reload, Home, Search, Security, Stop
- Page Content:**

DataDocket KnowledgeBase  
Statistical Output Delivery and Links

Output Objects

Library	Study Report	Analysis Output	Created By	Transfer Date	Transfer Status	Transferred By	EDMS Location
Dockazo 2539B	CSR98723	ae report #1		1997-10-13 12:46:50	complete	demo	DataDocketDemo/DataDocket/SASP
Dockazo 2539B	CSR98723	ae report #1		1997-10-13 14:41:22	in progress	demo	DataDocketDemo/DataDocket/SASP
Dockazo 2539B	CSR98723	ae report #1		1997-10-13 14:50:58	pending	demo	DataDocketDemo/DataDocket/SASP
Dockazo	CSR98723	ae report #1		1997-10-13	pending	demo	DataDocketDemo/DataDocket/SASP

Stop the current transfer

FIG. 17



1

## METHOD AND APPARATUS FOR THE INTEGRATION OF INFORMATION AND KNOWLEDGE

### PRIORITY TO PRIOR PROVISIONAL APPLICATION

Priority is claimed to Provisional Application Ser. No. 60/062,933, filed on Oct. 21, 1997 Pending.

This invention relates generally to an architecture for the integration of data, information and knowledge, and more particularly to a method and apparatus that manages and utilizes a knowledge repository for the purpose of enabling easy access, manipulation and visualization of synchronized data, information and knowledge contained in different types of software systems.

### SOURCE CODE APPENDIX

This patent document contains a source code appendix, including a total of 542 pages.

### COPYRIGHT NOTICE

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure, as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

### BACKGROUND AND SUMMARY OF THE INVENTION

Companies operating in regulated industries (e.g., aerospace, energy, healthcare, manufacturing, pharmaceuticals, telecommunications, utilities) are required to manage and review large amounts of information that is frequently generated over the course of several years. The principal components of this information are the structured numerical data and the unstructured textual documents. The data are collected and run through complex statistical analyses that are then interpreted and reported by industry experts to meet stringent requirements for regulatory review. Separate groups or organizations produce multiple iterations of these data and documents, with potentially thousands of statistical data analysis files linked to thousands of dependent documents. Often such groups have independently evolved specialized and often incompatible procedures and work practices. Correspondingly, separate software systems for data analysis and document management have been adopted as discrete solutions. The dichotomy existing in both the information sources and work groups jeopardizes the common goal. Hence, the challenge is to integrate and synchronize the flow of all information, processes and work practices necessary for making better and faster decisions within an enterprise.

Currently the process of integrating data and data analysis reports with regulatory documents can be characterized as (a) an entirely manual process (i.e., paper is copied and collated into a hard copy compilation), (b) a multi-step electronic process (i.e., files are placed into a central file location by one department and retrieved by another), or (c) an internally developed, custom solution that is used to automate portions of the process. Problems with such processes typically include:

complexity and error prone nature of the systems needed to manage the process(es) (e.g., manual updates to related documents and data, demands for maintaining a "mental" mapping of these objects to each other (i.e., a

2

meta information catalogue) and enforcing the integrity of the defined object "linkages" throughout the business process);

difficulty in locating and working with interrelated documents and data throughout the information generation lifecycle (a lack of integrated textual and numerical information severely constrains enterprise information workflow and decision making);

a lack of an efficient mechanism, in the current document management and data analysis systems, for locating and working with the many different types of information maintained in separate systems;

a failure to recognize, appreciate and enable the dependencies between data and documents throughout the information generation lifecycle—a complex information workspace topology exists that is known only intrinsically by the users who must maintain the referential integrity of these related information objects; and inflexibility of a process, during the information generation lifecycle, to handle situations where data changes force a series of document changes, which may in turn require modifications of other documents.

On the other hand, the present invention will alleviate such problems using an architecture that includes a knowledge repository for the purpose of enabling easy access, manipulation and visualization of complete and synchronized information contained on a plurality of software platforms.

Heretofore, a limited number of patents and publications have disclosed certain aspects of knowledge management systems, the relevant portions of which may be briefly summarized as follows:

U.S. Pat. No. 5,644,686 to Hekmatpour, issued Jul. 1, 1997, discloses a domain independent expert system and method employing inferential processing within a hierarchically-structured knowledge base. Knowledge engineering is characterized as accommodating various sources of expertise to guide and influence an expert toward considering all aspects of the environment beyond the individual's usual activities and concerns. This task, often complicated by the expert's lack of analysis of their thought content, is accomplished in one or more approaches, including interview, interaction (supervised) and induction (unsupervised). The expert diagnostic system described by Hekmatpour combines behavioral knowledge presentation with structural knowledge presentation to identify a recommended action.

U.S. Pat. No. 5,745,895 to Bingham et al, issued Apr. 28, 1998, discloses a method for associating heterogeneous information by creating, capturing, and retrieving ideas, concepts, data, and multi-media. It has an architecture and an open-ended-set of functional elements that combine to support knowledge processing. Knowledge is created by uniquely identifying and interrelating heterogeneous datasets located locally on a user's workstation or dispersed across computer networks. By uniquely identifying and storing the created interrelationships, the datasets themselves need not be locally stored. Datasets may be located, interrelated and accessed across computer networks. Relationships can be created and stored as knowledge to be selectively filtered and collected by an end user.

FileNet's "Foundation for Enterprise Document Management Strategy White Paper", Sep. 1997, suggests a major industry trend that is being generated by users: the convergence of workflow, document-imaging, electronic document management, and computer output to laser disk into a family of products that work in a common desktop PC environment. FileNet's foundation is a base upon which companies can easily build an enterprise-wide environment to access and



manage all documents and the business processes which utilize them. FileNet's architectural model is based on the client/server computing paradigm. Four types of generic client applications are described, the four main elements include:

Searching—the ability to initiate and retrieve information that “indexes” documents across the enterprise by accessing industry standard databases and presenting the results in an easy to use and read format.

Viewing—the ability to view all document types and work with them in the most appropriate way, including viewing, playing (video or voice), modifying/editing, annotating, zooming, panning, scrolling, highlighting, etc.

Development tools—industry-standard based development tool sets (e.g. Active X, PowerBuilder) that allow customers or their selected application development or integration partners to create specific applications that interface with other applications already existing in the organization.

Administrative applications—applications that deliver management and administrative information to users, developers, or system administrators that allow them to optimize tasks, complete business processes or receive data on document properties and functions.

SAS Institute's Peter Villiers has described, in a paper entitled “New Architecture for Linkages of SAS/PH-Clinical® Software with Electronic Document Management Systems” (June 1997, SAS Institute), an interface between SAS Institute's pharmaceutical technology products and document management systems (e.g., Documentum™ Enterprise Document Management System). In the described implementation, a point-to-point system (PH—Document Linker Interface and Documentum to SAS/PH-Clinical link back) is established to enable two-way transfer of information between the statistical database and the document repository.

While application integration solutions are being used to link major information management components, such as imaging, document management and workflow, none of the available integration methods manage the information overload or the contextual complexities characteristic of the regulatory application process. In order to make informed decisions, all information sources that are part of this process must be coalesced as part of a knowledge management architecture.

In the example of a regulated industry (e.g., pharmaceuticals), the primary problem is generally viewed as how to synthesize all the information to prove a regulatory application case as quickly as possible while not losing the context. Automating and synchronizing the flow of all information helps expedite the review process. But the bigger challenge is to preserve the context necessary for applying knowledge. A system is needed that enables users to put their knowledge to work; to answer such questions as: Are the documents consistent with the data? Were iterations of the data and documents synchronized? What was done to preserve the integrity of the data? Who performed the work and what were their qualifications? Appropriate answers to these questions will influence reviewer/regulator confidence in the data and assertions; yet in current systems, the information gets buried, lost or is never recorded. The present invention is directed to a system, architecture and associated processes that may be used to identify, confirm, integrate and enable others to follow the “path” that was used in meeting the regulatory approval requirements.

In accordance with the present invention, there is provided a knowledge integration system for providing application interoperability and synchronization between heterogeneous document and data sources, comprising: a first database memory; a data source suitable for independently performing data analysis operations using data stored within the first database to generate data and analysis results; a

document source, including a document database memory, for capturing knowledge and storing the knowledge in the form of documents, validating the accuracy of the knowledge, and making the captured knowledge available across a network; and a knowledge integration application, running on a client/server system having access to the data source and the document source, for managing the flow of information between the data source and the document source, thereby enabling the integration of data and analysis results with the documents and provide links to automatically update the documents upon a change in the data or analysis results.

The present invention represents an architecture, embodied for example in a software product suite, that manages and utilizes a knowledge repository, via knowledge integration middleware (KIMW), for the purpose of enabling easy access, manipulation and visualization of complete and synchronized information contained in different software systems. Aspects of the present invention include:

- the use of knowledge integration middleware in conjunction with traditional application integration middleware to build and manage an integration knowledge repository;
- providing a generic mechanism for bridging structured and unstructured data with uniform access to information;
- the specification of four integrated knowledge-based software applications (described below) that collectively enable information integration with knowledge linkage, visualization and utilization of structured, unstructured and work practice data and metadata produced by knowledge workers in an enterprise;
- use of a knowledge repository containing record of integration transactions, context information from users and applications, information metadata catalog, knowledge access control, application activation rules, metadata and rules for knowledge integration, knowledge generation, knowledge visualization, “live” knowledge links, task execution, and case-based data for regulatory review;
- use of a three dimensional (3D) interface in conjunction with a user-specific conceptual schema providing access to enterprise information wherever it is stored and managed; and
- implementation of a rule-based paradigm for filing marketing applications to regulatory agencies that uses hypothesis/proof/assertion structures.

The present invention will provide application interoperability and synchronization between heterogeneous document and data sources such as those currently managed by disparate enterprise document management and data analysis systems. Initially, the invention will allow users to establish and utilize “live” links between an enterprise document management system and a statistical database. Alternative or improved embodiments of the invention will enable users to define and execute multiple tasks to be performed by one or more applications from anywhere within a document.

Users of knowledge management systems desire an integrated and flexible process for providing Integrated Document Management, Image Management, Workflow Management and Information Retrieval. Aspects of the present invention focus on the added insight that a majority of the same customers want their data integrated in this document lifecycle platform as well as where the flow of textual and numerical analysis information are systematically synchronized. Such a system will enable decision makers to have complete information.