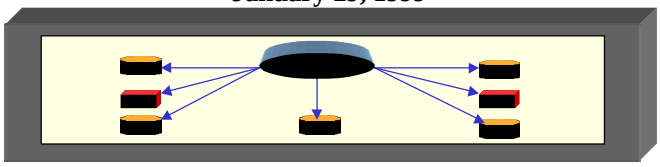
Executive Information Systems, Inc.

Project Five: Standard Recommended Practice On Artificial Knowledge Base Management Systems

By

Joseph M. Firestone, Ph.D.
eisai@moon.jic.com

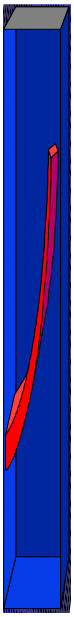
Prepared for First KMC/AIIM
KM ANSI/ISO Standards Committee Meeting
January 29, 1999



EIS Web Site: http://www.dkms.com

15 seconds

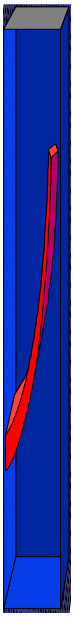
Objective and Approach



- We need to develop definitions of the concepts of Knowledge Base Management System (KBMS) and the Knowledge Warehouse (KW), and a standard recommended practice on Artificial KBMSs.
- KBMS and KW are clear analogues of DBMS and DW.
- I will begin with some basic distinctions, then discuss DBMSs, the DW, DW evolution, and Data Warehousing as a process, and end by developing the analogous concepts in the knowledge and KM spheres.

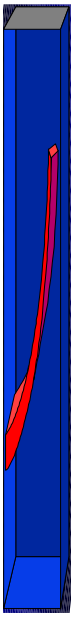
- 30 seconds, SUBT = 45 seconds

Data, Information, and Knowledge



- What is the difference between data, information, and knowledge in human organizations?
- To begin with, organizational data, information, and knowledge, all emerge from the social process of an organization, and are not private.
- We are trying to specify inter-subjective constructs and to provide metrics for them, not specify personal data, information, or knowledge

- 45 seconds SubT = 90 seconds
- What is the difference between data, information, and knowledge in human organizations?
- To begin with, organizational data, information, and knowledge, all emerge from the social process of an organization, and are not private.
- In defining them, we are not trying to formulate definitions that will elucidate the nature of personal data, information, knowledge, or wisdom.
- Instead, to use a word that used to be more popular in discourse than it is at present, we are trying to specify intersubjective constructs and to provide metrics for them.

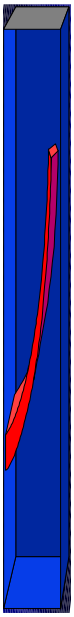


Data, Information, and Knowledge (Page Two)

- A datum is the value of an observable, measurable or calculable attribute. Data is more than one such attribute value.
- Is a datum (or is data) information?
- Not in itself; but information is provided by a datum, or by data, because data is always specified in some conceptual context.

- 30 seconds SubT = 120 seconds
- A datum is the value of an observable, measurable or calculable attribute. Data is more than one such attribute value.
- Is a datum (or is data) information?
- Not in itself; but information is provided by a datum, or by data, because data is always specified in some conceptual context.
- At a minimum, the context must include the class to which the attribute belongs, the object which is a member of that class, some ideas about object operations or behavior, and relationships to other objects and classes.

Data, Information, and Knowledge (Page Three)

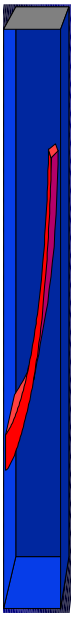


- Data alone and in the abstract therefore, does not provide information.
- Rather, information, in general terms, is data plus conceptual commitments and interpretations.
- Information is data extracted, filtered or formatted in some way.
- Knowledge is a subset of information. But it is a subset that has been extracted, filtered, or formatted in a very special way.
- It is information that has been subjected to, and passed tests of validation.

- 45 seconds, SUBT = 165 seconds, 2 minutes, 45 seconds

- Data alone and in the abstract therefore, does not provide information.
- Rather, information, in general terms, is data plus conceptual commitments and interpretations.
- Information is data extracted, filtered or formatted in some way (but keep in mind that data is always extracted filtered, or formatted in some way).
- Knowledge is a subset of information. But it is a subset that has been extracted, filtered, or formatted in a very special way.
- It is information that has been subjected to, and passed tests of validation.

Data, Information, and Knowledge (Page Four)



- Common sense knowledge is information that has been validated by common sense experience.
- Scientific knowledge is information (hypotheses and theories) validated by the rules and tests applied to it by some scientific community.
- More formally, the hierarchical network of the organization's validated rules is the knowledge base of the organization or enterprise.

- 60 seconds, SUBT = 225 seconds, 3 minutes, 45 seconds

■Common sense knowledge is information that has been validated by common sense experience.

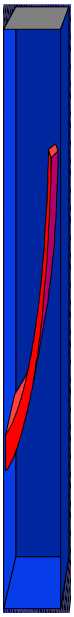
■Scientific knowledge is information (hypotheses and theories) validated by the rules and tests applied to it by some scientific community.

■More formally, the hierarchical network of the organization's validated rules is the knowledge base of the organization or enterprise.

■Each rule in the network relates antecedent attribute values to consequent attribute values, concepts, or rule sequences. The attributes involved belong to a number of concepts that represent the components of the model. Declarative Rule networks are those whose rules fire in parallel to determine an outcome. Procedural Rule networks are those whose rules fire in sequence.

■The knowledge base is composed of both declarative and procedural rule networks.

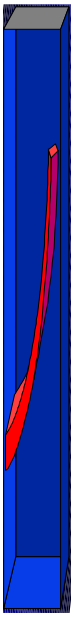
The Organizational Knowledge Base



- The knowledge base rule network of the organization contains:
- its set of remembered data;
- its validated propositions and models (along with metadata related to their testing);
- its refuted propositions and models (along with metadata related to their refutation);
- its metamodels; and (if the system produces such an artifact)
- the software it uses for manipulating these.

- 30 seconds, 255 seconds, 4 minutes, 15 secs.
- The organization's knowledge base enables it to explain, anticipate, and predict events and interaction patterns in the organization and in its environment.
- The knowledge base rule network of the organization contains: its set of remembered data; its validated propositions and models (along with metadata related to their testing); its refuted propositions and models (along with metadata related to their refutation); its metamodels; and (if the system produces such an artifact) the software it uses for manipulating these.

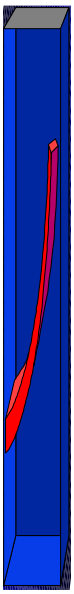
DBMS and Related Definitions



- A database is a self-descriptive, permanent, repository storing a collection of records kept for a common purpose.
- A DBMS is a computer program for managing this repository.
- A specific DBMS programming application, is produced by using a DBMS-template to create, maintain, and enhance it.
- Sometimes the template software is called a DBMS in common usage.

- 45 seconds, 300 seconds, 5 minutes
- A database is a self-descriptive, permanent, repository storing a collection of records kept for a common purpose.
- A DBMS is a computer program for managing this repository.
- A specific DBMS programming application, is produced by using a DBMS-template to create, maintain, and enhance it.
- Sometimes the template software is called a DBMS in common usage (e.g., Oracle, DB2, Sybase).
- But we should not lose sight of the fact that the program that manages a database in any specific situation is the concrete product of using a particular template or tool for producing an actual database management application.

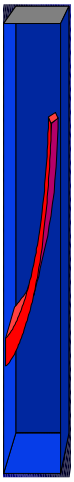
The Data Warehouse and Data Warehousing



- In the beginning, there were only “islands of information”
  - operational data stores,
  - legacy systems needing enterprise-wide integration, and
  - mission-specific DSSs.
- Then “along came Bill” (Inmon) and his concept of the Data Warehouse (DW) (seen as the solution to the problems of information integration and redundancy) -- the embodiment of enterprise-wide DSS for the ‘90s.

30 seconds, 330 secs.

The DW Definition



- Inmon defined the DW as “a
  - subject-oriented
  - integrated,
  - time-variant
  - non-volatile
  - collection of data in support of management’s decision making process.”

- 30 secs. 360 secs., 6 minutes
- This is the classic definition of the Data Warehouse. According to it, the DW is a type of database managed by a DBMS.
- Indeed, in its present form the DW is a database that uses a relational DBMS. Inmon's definition is now undergoing change as the DW field evolves. Figure One depicts where DW began.

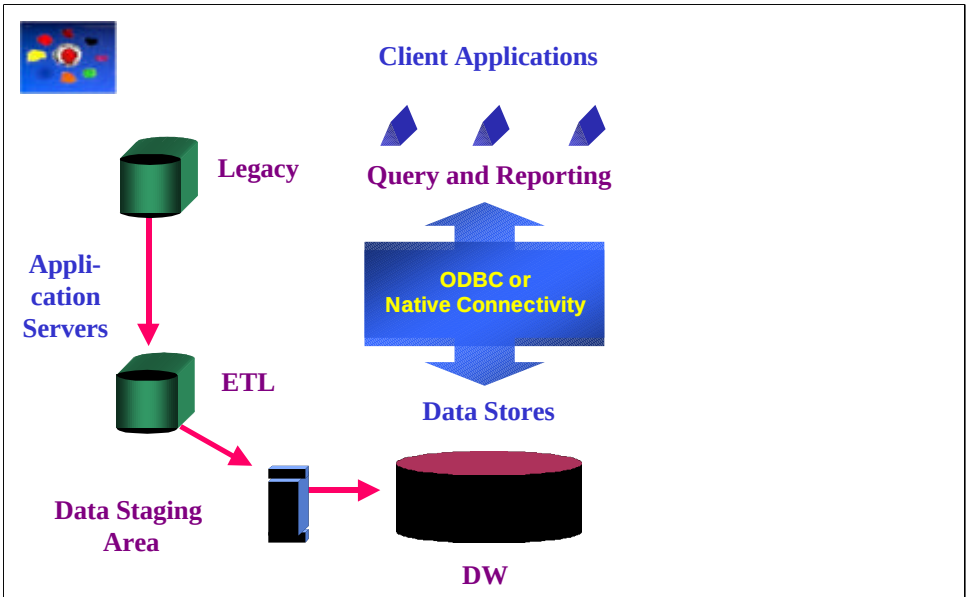
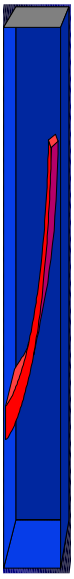


Figure One -- Data Warehousing: The Beginning

- 45 secs., 405 secs., 6 minutes, 45 secs.
- DW has focused a lot on ETL and Data Staging
- The DW Data Store is Relational
- A Big Issue is how the DW should be modeled
  - Star Vs. 3NF
- A Very active Area has been that of query and reporting tools which are claimed to implement business intelligence

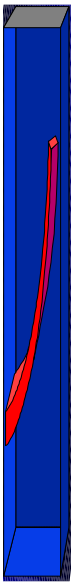
Data Warehousing Reality



- Data Marts and Data Mining were not part of this vision. There was only the DW
- But the vision was too sweeping. DW's were too
  - costly,
  - impolitic,
  - took too long to implement, and
  - the architecture of DW systems was too simple to handle customer requirements

30 secs., 435 secs., 7 minutes, 15 secs.

Evolution in Data Warehousing



- So evolution in data warehousing systems began with:
  - Data Marts
  - Dynamic Data Staging areas
  - Operational Data Stores
  - Web and OLAP Clients
  - A Variety of application Servers were added to the ETL, Legacy, and Database Servers in DW systems to fill a variety of user needs

- **So evolution in data warehousing systems began with:**
  - **Data Marts 120 secs., 555 secs.**
  - **Dynamic Data Staging areas**
  - **Operational Data Stores**
  - **Web and OLAP Clients**
  - **A Variety of application Servers were added to the ETL, Legacy, and Database Servers in DW systems to fill a variety of user needs**
- Intelligent Agent Technology is being integrated into DW systems, though we do not yet see Agency application servers and general use of agents.
- We also see the introduction of "second generation" Metadata exchange architectures based on a commitment to COM and Object Technology for improving metadata managers and integration of ETL, other application servers and DSS data stores
- Figure Two depicts where DW is now.

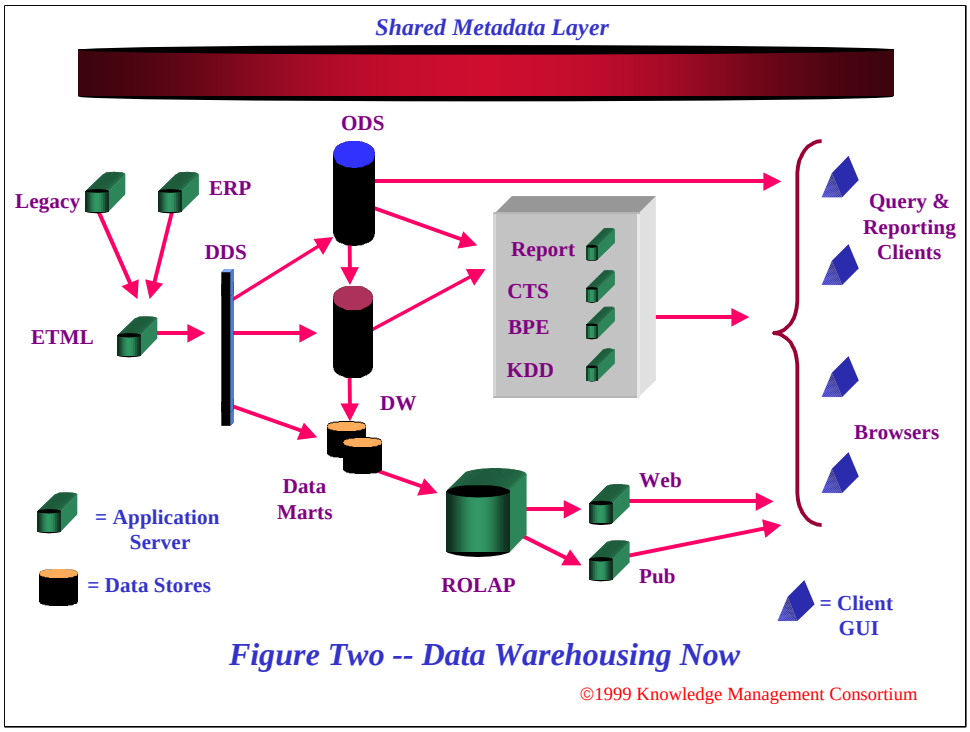
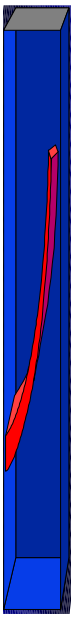


Figure Two -- Data Warehousing Now

- **180 seconds, SUBT = 735 seconds, 12 minutes, 15 seconds**
- **Web** = Web Information Server **Pub** = Publication & Delivery Server
- **KDD** = Knowledge Discovery in Databases/Data Mining Servers
- **ETML** = Extraction, Transformation, Migration and Loading
- **DDS** = Dynamic Data Staging Area **DW** = Data Warehouse
- **ODS** = Operational Data Store **ERP** = Enterprise Resource Planning
- **Query** = Query and Reporting Server **CTS** = Component Transaction Server
- **BPE** = Business Process Engine
- **ROLAP** = Relational Online Analytical Processing
- Note the great increase in functionality and complexity in the above system, and the correspondingly greater need for integrative mechanisms
- In particular, the greater and increasing role of application servers in general, and BPEs in particular is manifest in data warehousing. ROLAP and KDD servers are also BPE types.
- Metadata is now heavily emphasized as an integrative mechanism

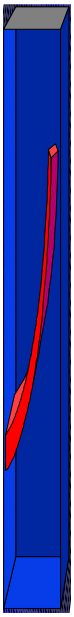
The Evolution of Data Warehousing (Page Two)



- Given the changes in DW system complexity, Data Warehousing is now, increasingly, a problem of integrating a variety of distributed warehouse data stores with various specialized application servers and front end access devices that need warehouse data.
- The Data Warehousing System, which began as a low volatility system, is now a system that may integrate DSS, batch and OLTP processing, and that therefore may incorporate considerable volatility.

- 45 seconds, 780 seconds, 13 minutes
- So, Data Warehousing used to focus on gathering data from legacy sources of various kinds, putting it through the ETL process, loading it into the data warehouse, and providing reporting tools and report templates to access it conveniently.
- Given the changes in DW system complexity, Data Warehousing is now, increasingly, a problem of integrating a variety of distributed warehouse data stores with various specialized application servers and front end access devices that need warehouse data.
- The Data Warehousing System, which began as a low volatility system, is now a system that may integrate DSS, batch and OLTP processing, and that therefore may incorporate considerable volatility.

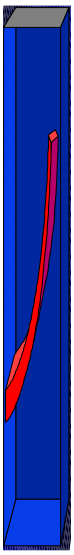
Challenges in Data Warehousing



- How can increasingly complex data warehousing systems:
  - achieve dynamic integration?
  - comprehensively integrate and support Knowledge Production?
  - Store knowledge for high capability decision support?
  - efficiently deliver tactical decision support using volatile data stores?
  - integrate ERP systems?
  - Integrate increasingly varied business process engines?

- 45 seconds, SUBT = 825 seconds, 13 minutes, 45 seconds
- The current state of Data Warehousing raises the following issues. How can increasingly complex data warehousing systems:
  - achieve dynamic integration?
  - comprehensively integrate and support Knowledge Production?
  - Store knowledge for high capability decision support?
  - efficiently deliver tactical decision support using volatile data stores?
  - integrate ERP systems?
  - Integrate increasingly varied business process engines?
  - Business Process Engines are application servers that maintain state in memory rather than in persistent storage.

The Evolution of Data Warehousing and The AKM



- To successfully resolve these issues data warehousing systems need an integrative component with the capabilities of the Artificial Knowledge Manager, so that future Data Warehousing Systems will look like Figure Three.

- 15 seconds SubT = 840 seconds, 14 minutes
- To successfully resolve these issues data warehousing systems need an integrative component with the capabilities of the Artificial Knowledge Manager, so that future Data Warehousing Systems will look like Figure Three.

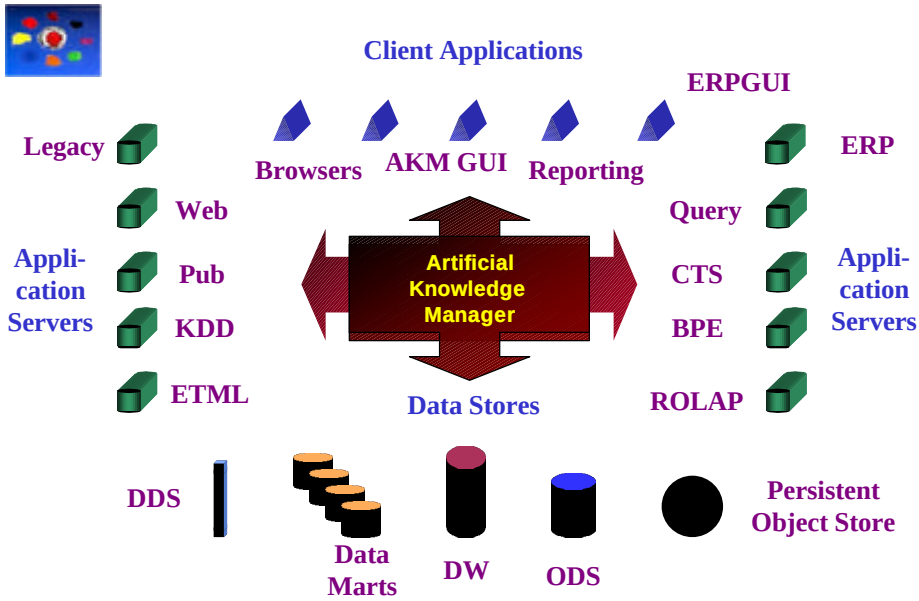
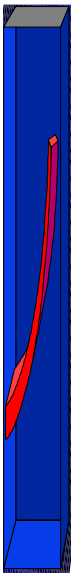


Figure Three -- A Data Warehousing System with the AKM

90 seconds Subt = 930 seconds, 15 minutes, 30 secs.

- **Web** = Web Information Server **Pub** = Publication & Delivery Server
- **KDD** = Knowledge Discovery in Databases/Data Mining Servers
- **ETML** = Extraction, Transformation, Migration and Loading
- **DDS** = Dynamic Data Staging Area **DW** = Data Warehouse
- **ODS** = Operational Data Store **ERP** = Enterprise Resource Planning
- **Query** = Query and Reporting Server **CTS** = Component Transaction Server
- **BPE** = Business Process Engine
- **ROLAP** = Relational Online Analytical Processing
- **Image Management ? Document Management? Multimedia (video? Audio?)** Combination of Application Servers, BPEs, and Data Stores
- **AI?** Fits in across the board**.** Imaging, KDD, BPEs, the AKM, Pub, Agents.
- **Agents?** Also fits in across the board**. Use your own paper**
- **Computer Simulation?** BPEs
- **Balanced Scorecard?** A BPE

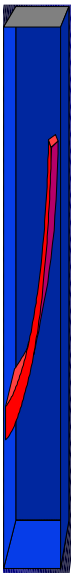
The Knowledge Base Again



- The organization's knowledge base is an abstract phenomenon.
- It is one that emerges from the interaction of the various agents comprising the organization.
- Measures of an organization's knowledge base may be found in its cultural artifacts, including its linguistic products, its electronic artifacts, and its artistic expressions, if any.
- One type of cultural artifact of an organization is its Artificial Knowledge Base (AKB).

- 60 seconds, 990 seconds, 16 minutes 30 seconds.
- With this account of Data Warehousing as background let's discuss AKBs, KBMSs, and KWs. Previously we stated that the knowledge base of an organization contains: its set of remembered data; its validated propositions and models (along with metadata related to their testing); its refuted propositions and models (along with metadata related to their refutation); its metamodels; and (if the system produces such an artifact) the software it uses for manipulating these.
- The organization's knowledge base is an abstract phenomenon. And it is one that emerges from the interaction of the various agents comprising the organization. Measures of an organization's knowledge base may be found in its cultural artifacts [12], including its linguistic products, its electronic artifacts, and its artistic expressions, if any.
- One type of cultural artifact of an organization is its Artificial Knowledge Base (AKB).

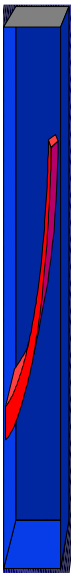
The Artificial Knowledge Base (AKB)



- An AKB is the portion of an organization's knowledge base expressed in the persistent storage and non-persistent memory of its computers.
- The AKB, like a database, is self-descriptive, is ultimately composed of bits and bytes, is permanent in the sense that it is an on-going system, is located both in specific in-memory locations and in specific persistent storage location, and is kept to fulfill an organization's purposes.

- 30 seconds, 1020 secs., 17 minutes
- An AKB is the portion of an organization's knowledge base expressed in the persistent storage and non-persistent memory of its computers.
- The AKB, like a database, is self-descriptive, is ultimately composed of bits and bytes, is permanent in the sense that it is an on-going system, is located both in specific in-memory locations and in specific persistent storage location, and is kept to fulfill an organization's purposes.

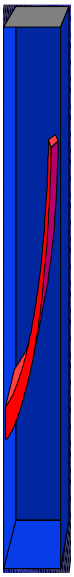
The Artificial Knowledge Base (AKB) (Page Two)



- Unlike a database which stores records, however, an AKB stores a network of objects and components, and these encapsulate data and methods (validated and unvalidated procedural or declarative rules that use validated and unvalidated data).
- So the AKB stores data and information as well as knowledge.

- 30 seconds, 1050 secs., 17 minutes 30 secs
- Unlike a database which stores records, however, an AKB stores a network of objects and components, and these encapsulate data and methods (validated and unvalidated procedural or declarative rules that use validated and unvalidated data).
- So the AKB stores data and information as well as knowledge.

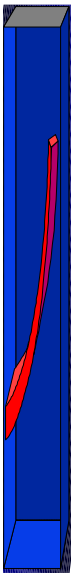
The Knowledge Base Management System (KBMS)



- A Knowledge Base Management System (KBMS) is a computer application for managing (creating, enhancing, and maintaining) the AKB, just as a DBMS is a computer application for managing a database.
- But what does such a computer program do? To answer this question, look again at Figure Three.

- 30 seconds, 1080 secs, 18 mins.
- A Knowledge Base Management System (KBMS) is a computer application for managing (creating, enhancing, and maintaining) the AKB, just as a DBMS is a computer application for managing a database.
- But what does such a computer program do? To answer this question, look again at Figure Three.

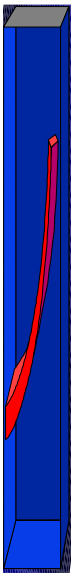
The Data Warehouse, The KBMS, and the AKMS



- Figure Three is not simply a Data Warehousing System. It is an Enterprise AKMS.
- It is also a KBMS, because it can manage (through the AKM, its database management, application server, and communications and connectivity software) not only data and information, but also the network of objects and components constituting an AKB.

- 30 secs., 1110 secs., 18 mins., 30 secs.
- Figure Three is not simply a Data Warehousing System. It is an Enterprise AKMS.
- It is also a KBMS, because it can manage (through the AKM, its database management, application server, and communications and connectivity software) not only data and information, but also the network of objects and components constituting an AKB.

The AKMS and The Knowledge Warehouse

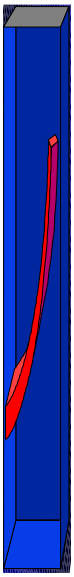


- In fact, the persistent data stores in Figure Three are not simply data stores, but taken together, including their OODBMS component, they are knowledge stores.
- They can store objects, and methods, and rules, and validation information, as well as data.
- And that makes Figure Three a Knowledge Warehousing System, and not just a Data Warehousing System.

- 30 secs. 1140 secs., 19 minutes

- In fact, the persistent data stores in Figure Three are not simply data stores, but taken together, including their OODBMS component, they are knowledge stores.
- They can store objects, and methods, and rules, and validation information, as well as data.
- And that makes Figure Three a Knowledge Warehousing System, and not just a Data Warehousing System.

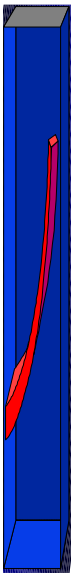
Convergence



- In a nutshell, the changes summarized above, indicate that data warehousing systems are about to evolve into AKMSs, or equivalently, Knowledge Base Management Systems, or Knowledge Warehousing Systems.
- Convergence between data warehousing, DSS, and KM is about to occur.
- There is no separate Knowledge Base Management System.
- The KBMS is both the AKMS and the Knowledge Warehousing System.

- 30 secs., 1170 secs. 19 mins. 30 secs.
- In a nutshell, the changes summarized above, indicate that data warehousing systems are about to evolve into AKMSs, or equivalently, Knowledge Base Management Systems, or Knowledge Warehousing Systems.
- Convergence between data warehousing, DSS, and KM is about to occur.
- There is no separate Knowledge Base Management System. The KBMS is both the AKMS and the Knowledge Warehousing System.
- Take your pick on the name.

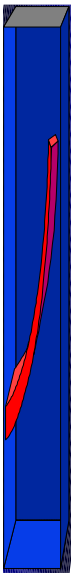
What of The Knowledge Warehouse?



- Like the DW, it may be viewed as subject-oriented, integrated, time-variant, and supportive of management's decision making processes.
- But unlike the DW, it is a combination of volatile and non-volatile objects and components, and, of course, it stores not only data, but also information and knowledge.
- The KW is not co-extensive with the AKMS. It is also not a physical subsystem of the AKMS, as the data warehouse is of the DSS is supports, to which one can easily point.

- 45 secs., 1215 secs., 20 minutes, 15 seconds
- What of the Knowledge Warehouse?
- Like the DW, it may be viewed as subject-oriented, integrated, time-variant, and supportive of management's decision making processes.
- But unlike the DW, it is a combination of volatile and non-volatile objects and components, and, of course, it stores not only data, but also information and knowledge.
- The KW is not co-extensive with the AKMS. It is also not a physical subsystem of the AKMS, as the data warehouse is of the DSS is supports, to which one can easily point.

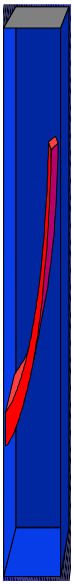
The Knowledge Warehouse and the AKB



- The KW is physically resident both in-memory and in persistent distributed data stores.
- Abstractly, however, the KW is the AKB itself.
- There is no distinction between the AKB and the KW as there is between an enterprise wide federated database, and its data warehouse component.

- 30 secs., 1245 secs., 20 minutes, 45 secs.
- The KW is physically resident both in-memory and in persistent distributed data stores.
- Abstractly, however, the KW is the AKB itself. There is no distinction between the AKB and the KW as there is between an enterprise wide federated database, and its data warehouse component.

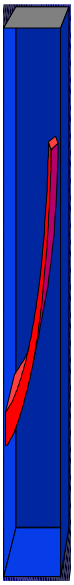
The Consequences of Convergence on the AKMS



- The AKMS is an On-line Complex Processing (OLCP) System.
- The AKMS, given present technology, is a distributed processing system.
- Since the KBMS is the AKMS, it follows that the standard the AKMSC is developing for the AKMS is also the KBMS standard, and any software tools developed on the basis of the standard will be KBMS tools as well as AKMS tools.

- 45 secs., 1290 secs., 21 minutes, 30 seconds
- The AKMS is an On-line Complex Processing (OLCP) System. Not merely a DSS system like today's data warehousing system. Nor an OLTP system, like today's ERP systems.
- The AKMS, given present technology, is a distributed processing system, or as I have called it elsewhere, a Distributed Knowledge Management System (DKMS).
- Since the KBMS is the AKMS, it follows that the standard the AKMSC is developing for the AKMS is also the KBMS standard, and any software tools developed on the basis of the standard will be KBMS tools as well as AKMS tools.

The KBMS Standard Recommended Practice



- If the KBMS and the AKMS are one and the same, and the KW and the AKB are also equivalent, it follows that the standard recommended practice for the KBMS is the same as that for the AKMS.
- To develop such a recommendation, we first need to define the AKMS standard in much greater detail.
- To do this we need to implement the AKMSC "straw man" program outlined in Working Paper No. One. Here again is a list of tasks in the program.

- 30 seconds, 1320 secs., 22 minutes
- If the KBMS and the AKMS are one and the same, and the KW and the AKB are also equivalent, it follows that the standard recommended practice for the KBMS is the same as the standard recommended practice for the AKMS.
- To develop such a recommendation, we first need to define the AKMS standard in much greater detail.
- To do this we need to implement the AKMSC "straw man" program outlined in Working Paper No. One.
- Here again is a list of tasks in the program.

AKMSC "Strawman" Program

- Specify AKMS Use Case Model and Relate to EKM Processes and Activities
- Specify the Artificial Knowledge Manager (AKM) Logical Component
- Specify Types of Client Application Components.
- Specify Types of Application Servers
- Specify Communication Buses including Object Request Brokers (ORBs)

90 seconds SubT = 1410 seconds = 23 minutes 30 seconds

- Specify AKMS Use Case Model and Relate to EKM Processes and Activities: In the Unified Modeling Language (UML) a use case is defined as "a set of sequences of actions a system performs that yield an observable result of value to a particular actor." A use case model requires description of the all the use cases and associated diagrams
- Specify the Artificial Knowledge Manager (AKM) Logical Component: We'll talk about this component in a minute.
- Specify Types of Client Application Components: These refer to Interface Components of Applications
- Specify Types of Application Servers: Application servers provide services to other components in a distributed processing system by executing business logic and data logic on data accessed from database servers.
- Specify Communication Buses including Object Request Brokers (ORBs): Which ORBS are essential for an open standard?

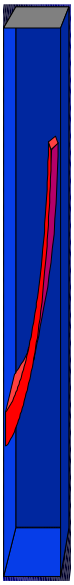
AKMSC "Strawman" Program

- Specify Types of Data Stores
- Specify AKMS Architectural Model
- Specify AKMS Model
- Specify Artificial Knowledge Manager Standard
- Specify Knowledge Warehouse Standard

90 seconds SubT = 1500 seconds = 25 minutes

- Specify Types of Data Stores: RDBMS, OODBMS, Flat File, etc.
- Specify AKMS Architectural Model: An abstract of the AKMSs technical structure
- Specify AKMS Model: An abstract of component structure, relationships and dynamics in the AKMS
- Specify Artificial Knowledge Manager Standard: Define requirements for the AKM logical component
- Specify Knowledge Warehouse Standard: Define the knowledge warehouse in the context of the AKMS and the AKM.

The KBMS Standard Recommended Practice



- Once the AKMS standard is developed, we can proceed to develop the standard recommended practice for implementing an AKMS.
- In the mean time, fields likely to contribute to the standard can be studied. The two main ones are Object-Oriented Software Engineering (OOSE) and Data Warehousing.
- Both fields are in ferment right now, and practitioners and vendors alike are offering methodologies for their Communities of Practice (CP).
- The standard recommended practice for the AKMS may perhaps be developed as a synthesis of these two developing CPs.

- 90 seconds, 1590 seconds, 26 minutes 30 seconds.
- Once the AKMS standard is developed, we can proceed to develop the standard recommended practice for implementing it.
- In the mean time, fields likely to contribute to the standard can be studied. The two main ones are Object-Oriented Software Engineering (OOSE) and Data Warehousing.
- Both fields are in ferment right now, and practitioners and vendors alike are offering methodologies for their Communities of Practice (CP).
- In OOSE, methodologies utilizing the Unified Modeling Language (UML), aimed at rapid application development of distributed object applications are now beginning to appear. In data warehousing, the simplistic methodologies of the early days of two tier data warehousing are giving way to incremental, iterative methodologies for developing distributed data warehouses over time.
- The standard recommended practice for the AKMS may perhaps be developed as a synthesis of these two developing CPs.