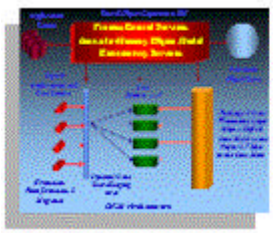

DKMS Briefs

Joseph M. Firestone

DKMS Brief No. One: The Corporate Information Factory or the Corporate Knowledge Factory?

The Corporate Information Factory

W. H. Inmon's vision of the IT future is an information ecosystem

> "with different components, each serving a community directly while working in concert with other components to produce a cohesive, balanced information environment. Like nature's ecosystem, an information ecosystem must be adaptable, changing as the inhabitants and participants within its aegis change. Over time, the balance between different components and their relationship to each other changes as well, as the environment changes. Sometimes the effect will appear on seemingly unrelated parts (sometimes disastrously!). Adaptability, change, and balance, are the hallmarks of the components of a healthy information ecosystem." [1]

Further, "the corporate information factory (CIF) is the physical embodiment of the notion of an information ecosystem." [2] In other words, the CIF "is an architecture for the information ecosystem, consisting of the following architectural components:

- An applications environment
- An integration and transformation layer (I & T layer)
- A data warehouse with current and historical detailed data
- A data mart(s)
- An operational data store (ODS)
- An Internet and Intranet
- A metadata repository" [3]

Both the information ecosystem and the CIF are, according to Inmon, evolving to meet the pressures for business efficiency in three key areas: business operations, business intelligence, and business management. While IT has successfully delivered capabilities in the area of business operations without the CIF, it is because of the increasing need for greater business intelligence and greater business management capability that corporations are developing the information ecosystem and the CIF. [4]

In short the goals are better business operations, improved business intelligence, and enhanced business management. The means are the information ecosystem and the CIF. But left unanswered is the question of why the information ecosystem and the CIF are the appropriate means. The means can be questioned in two ways. First, we can disagree about the make-up of the CIF. That is, it is easy to postulate additional components for the CIF such as: Component Transaction Servers [5], Active Information Managers [6], Business Process Engines [7], and other Application Servers [8]. But second, beyond this and more important, why should we assume that an information ecosystem, and an associated CIF are appropriate objectives for fulfilling our goals? To see whether this is true, we need to examine what information is, and how it is related to other outcomes of our attempts to understand the business environment, namely: data, knowledge, and wisdom.

What are the Differences Among Data, Information, Knowledge, and Wisdom?

To begin with, organizational data, information, knowledge, and wisdom, 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 inter-subjective constructs and to provide metrics for them.

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? Yes, information is provided by a datum, or by data, but only 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 that is a member of that class, some ideas about object operations or behavior, and relationships to other objects and classes.

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. More specifically, the information we call knowledge is information that has been subjected to, and passed tests of validation. 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.

Organizational knowledge, in terms of this framework, is information validated by the rules and tests of the organization seeking knowledge. The quality of its knowledge then, will be largely dependent on the tendency of its validation rules and tests to produce knowledge that improves organizational performance or in Inmon's terms business operations, business intelligence, and business management (the organization’s version of objective knowledge).

Wisdom, lastly, has a more active component than data, information, or knowledge. It is the application of knowledge expressed in principles to arrive at prudent, sagacious decisions about conflictful situations. [9]

From the viewpoint of the definition given of organizational knowledge, what is an organization doing when it validates information to produce knowledge? It seems reasonable to propose that the validation process is an essential aspect of the broader organizational learning process, and that validation is a form of learning. So, though knowledge is a product and not a process derived from learning, knowledge validation (validation of information to admit it into the knowledge base) is certainly closely tied to learning, and depending on the definition of organizational learning, may be viewed as derived from it.

The Corporate Distributed Knowledge Management System and The Corporate Knowledge Factory

It follows directly from the distinctions among data, information, knowledge and wisdom, that if our goals are to improve business operations, business intelligence, and business management, we must do this not through an information ecosystem and a corporate information factory, but with analogous constructs focused on knowledge, rather than just information. Why? Because:

- We need correct information to improve business operations, intelligence, and management; incorrect information will lead to worse performance, not improvement
- Information that has been validated as correct, is more likely to be correct than unvalidated information or validated information that has not survived testing
- Therefore we need a special kind of corporate information ecosystem called a corporate knowledge ecosystem or a **Corporate Distributed Knowledge Management System (CDKMS)** to improve performance
- And we also need a special kind of information architecture called a **Corporate Knowledge Factory (CKF)**, to express the IT architecture supporting this system

Let's get to some definitions.

- The **Knowledge Management System (KMS)** [10] is the on-going, persistent interaction among agents within a system that produces, maintains, and enhances the system's knowledge base. This definition is meant to apply to any intelligent, adaptive system composed of interacting agents, including a corporation. So, the **Corporate Knowledge Management System (CKMS)** is just the KMS of a corporation. An agent is a purposive, self-directed object. Knowledge base will be defined in the next section.

  In saying that a CKMS produces knowledge I am saying that it (a) gathers information and (b) compares conceptual formulations describing and evaluating its experience, with its goals, objectives, expectations or past formulations of descriptions, or evaluations. Further, this comparison is conducted with reference to ***validation criteria***. Through use of such criteria, CKMSs and other intelligent systems distinguish competing descriptions and

evaluations in terms of closeness to the truth, closeness to the legitimate, and closeness to the beautiful. [11]

In saying that a CKMS maintains knowledge I am also saying that it continues to evaluate its knowledge base against new information by subjecting the knowledge base to continuous testing against its validation criteria. We are also saying that to maintain its knowledge, a more complex system must ensure both the continued dissemination of its currently validated knowledge base, and continued socialization of intelligent agents in the use and content of its knowledge base.

Finally, in saying that a CKMS enhances its knowledge base, I am saying that it adds new propositions and new models to its knowledge base, and also simplifies and increases the explanatory and predictive power of its older propositions and models. That is, one of the functions of the CKMS is to provide for the growth of knowledge.

- A CKMS's **knowledge base** is: the set of remembered data; validated propositions and models (along with metadata related to their testing); refuted propositions and models (along with metadata related to their refutation); metamodels; and software used for manipulating these, pertaining to the CKMS and produced by it. [12]

  A corporate knowledge management system, in this view, requires an at least partly automated knowledge base to begin operation. But it enhances its own knowledge base with the passage of time because it is a self-correcting system, and subjects its knowledge base to testing against experience.

- The **Corporate Knowledge Management Process (CKMP)** is an on-going persistent interaction among human-based agents who aim at integrating all of the various agents, components, and activities of the CKMS into a planned, directed process producing, maintaining and enhancing its knowledge base. [13]
- **Corporate Knowledge Management (CKM)** is the human activity within the CKMP aimed at creating and maintaining this integration, and its associated planned, directed process.
- The **Corporate Distributed Knowledge Management System (CDKMS)** [14] manages the integration of distributed IT components into a functioning whole supporting the activities of producing, maintaining, and enhancing its knowledge base. A CDKMS, in this view, is an information systems application. It is a special kind of information ecosystem, in Inmon's sense. It is distinct from the CKMS, which transcends information systems, and covers all types of corporate activity involving the knowledge base. The CDKMS requires a knowledge base to begin operation. But it enhances its own knowledge base with the passage of time because it is a self-correcting system, subject to testing against experience.

  The CDKMS must not only manage data, but all of the information, components, objects, object models, process models, use case models, object interaction models, and dynamic models, used to process data and to interpret it to produce a business knowledge base. It is because of its role in managing and processing distributed data, objects, and models to produce a knowledge base that the term Corporate Distributed Knowledge Management System is so appropriate.

- The **Corporate Distributed Knowledge Management Architecture (CDKMA),** which can also be called the **Corporate Knowledge Factory (CKF)** [15]**,** is the IT architecture needed to implement the CDKMS. It is different from the CIF in essential ways I'll cover shortly.

How are the CDKMS and the CKF different from the information ecosystem and the CIF? The difference is that the CDKMS and the CKF are sub-types of the information ecosystem and the CIF, focused on the objectives of producing, maintaining, and enhancing knowledge, and further on the necessity of validating information as the essential aspect of knowledge production, maintenance and enhancement. I cannot over-emphasize the importance of this difference between the respective information management system/architecture and knowledge system management/architecture constructs. There is nothing in Inmon's conception that orients IT toward Knowledge Management (KM), only toward information management and data management.

Without an orientation toward KM however, there is no tight coupling between the IT remedies -- the information ecosystem, and the CIF -- and the goals of improving business operations, intelligence, and management. An information ecosystem, however balanced and adaptable, is oriented toward producing, maintaining, and enhancing information, and information, again, is just data filtered and formatted in such a way as to give it structure. It has no necessary relationship to reality or the ability to influence it in accordance with corporate goals. But knowledge has such a relationship, and insofar as CKM with the assistance of the CDKMS and the CKF is successful in improving relevant knowledge production, maintenance, and enhancement, it must also provide the capability to improve business operations intelligence, and management.

The CIF and The CKF

As well as being different in their goal orientation the CIF and the CKF are different in their components. Figure One displays the basic CKF or CDKM architecture [16]. The differences from the CIF are substantial. They are:

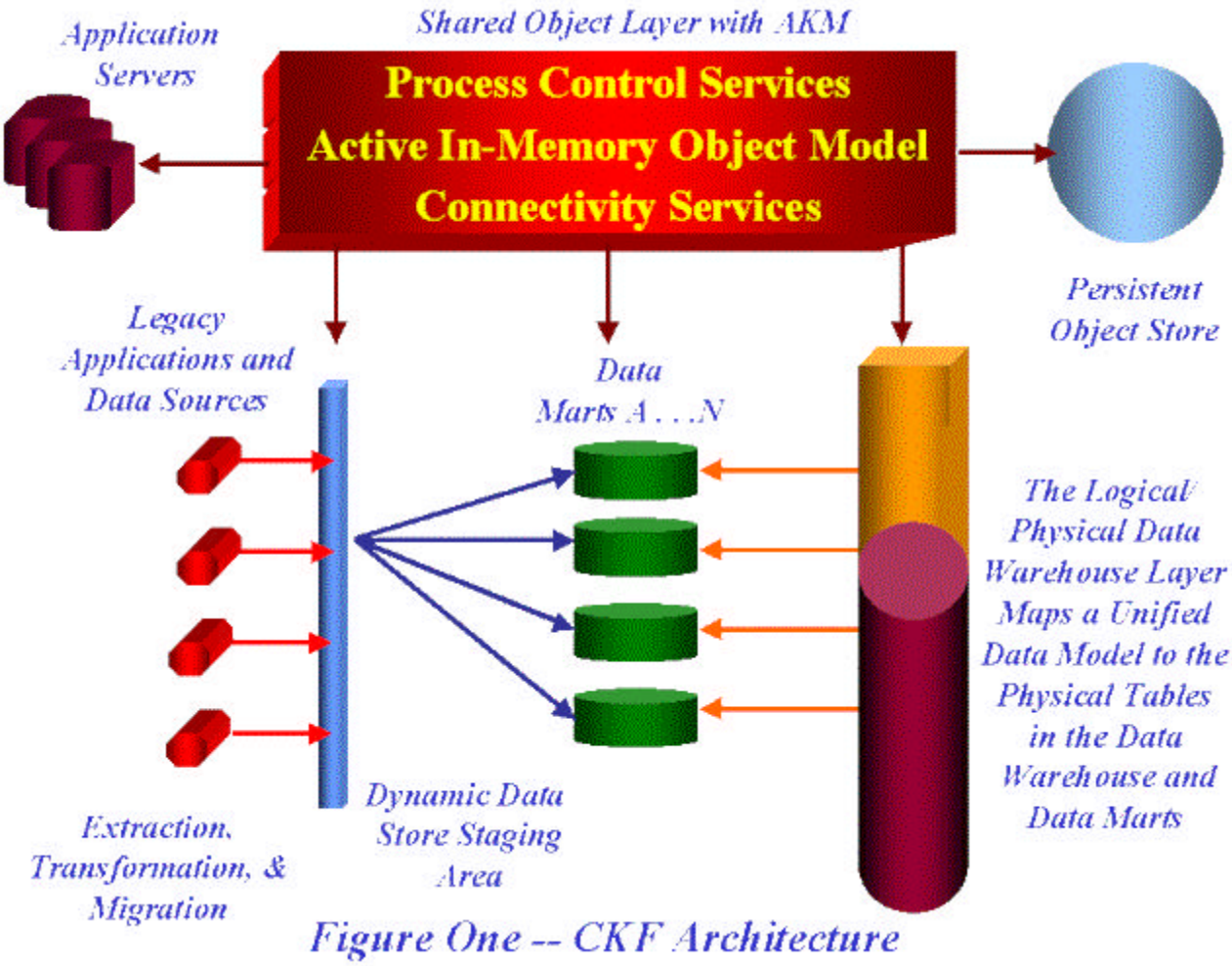

Figure One -- CKF Architecture

- The CKF emphasizes most of the components of the CIF, but not the ODS which is optional in the CKF
- In addition, central to the CKF is the Active Knowledge Manager [17], a server providing: process control services, an active in-memory object model, and connectivity services that may include an Object request Broker (ORB).
- In addition, persistent storage for the AKM is provided by an OODBMS
- In addition, the CKF emphasizes a number of application servers (not distinguished in the figure) as essential to the architecture. These include:
- A Data Mining Server(s) devoted to Knowledge Discovery in Databases (KDD), including pre-processing data for data mining, data mining, and validation [18]
- A Component Transaction Server (CTS) devoted to monitoring and managing transactions among system components
- A Business Process Engine(s) [19] devoted to supporting scalable performance in the CDKMS by breaking the chain of serial requests for data using in-memory business state management, business state synchronization, transactional multi-threading, proactive operations including intelligent software agents, and component management [20].

Most or all of these additions could be added to the CIF to make it equivalent to the CKF, and certainly there is nothing in the CIF concept that would make such additions inconsistent with its conceptual foundation. But this is to beg the question of comparing the orientations provided by the CIF with that of the CKF, and determining whether the CIF or the CKF should be preferred as an architectural concept.

The CIF does not naturally orient the architect to KDD, because it does not emphasize the centrality of knowledge discovery. It also does not naturally orient the architect to object-orientation and therefore to the other added components

either, because "information management" has long been more associated with data processing and data management than it has been with abstract modeling, general systems approaches to knowledge development, and advanced data analysis. On the other hand, when we begin to think in terms of the CKF about KDD-related manipulation and validation; we naturally think about analytical model repositories, tree structures, graph theory, neural networks, genetic algorithms, complex adaptive system simulations, and other matters that are handled well by O-O based components added to the architecture, and not as well by more traditional database components.

In short, for all of the reasons stated above, if you want to improve business operations, intelligence and management, the proper orienting concepts are not the information ecosystem and the CIF, but instead the CDKMS and the CKF. These are more closely related to the business goals that Inmon wanted to support. In addition, they are suggestive of a wider and increasingly necessary set of O-O -based components for IT solutions than are the information ecosystem and the CIF. So let's replace the CIF with the CKF and get on with constructing corporate distributed knowledge management solutions.

References

[1] See W. H. Inmon, Claudia Imhoff, and Ryan Sousa, Corporate Information Factory (New York, NY: John Wiley & Sons, 1998), Pp. 2-3

[2] Ibid. P. 8

[3] Ibid. P. 13

[4] Pp. 5-11

[5] Some examples are Sybase's Jaguar CTS, and Microsoft's MTS.

[6] Later on, I call this component the Active Knowledge Manager (AKM) and define it.

[7] Defined later.

[8] For example, data mining servers, commodity trading servers, etc.

[9] I read Gene Bellinger's views on data, information, knowledge, and wisdom at http://www.radix.net/~crbnblu/musings/kmgmt/kmgmt.htm, before writing my own differing account of these four concepts. His views are certainly worth keeping in mind when considering mine

[10] For more detail see my "Basic Concepts of Knowledge Management," White Paper at http://www.dkms.com/KMBASIC.html.

[11] I'm referring to the view that validation criteria can be applied in arriving at ethical and aesthetic knowledge, as in arriving at factual knowledge. See Nicholas Rescher's Objectivity: The Obligations of Impersonal Reason (Notre Dame, IN: University of Notre Dame Press, 1997), Chs. 9-11, and E. W. Hall, Our Knowledge of Fact and Value (Chapel Hill, NC: University of North Carolina Press, 1961).

[12] For more detail see my "Basic Concepts . . ." op. cit.

[13] For more detail, see Ibid.

[14] I introduced the DKMS concept in two previous White Papers "Object-Oriented Data Warehouse," and "Distributed Knowledge Management Systems: The Next Wave in DSS." Both are available at http://www.dkms.com/White_Papers.htm.

[15] I've used the more general term Distributed Knowledge Management (DKM) Architecture in an earlier context. But the CKF is just the DKM in corporate contexts. See my "Architectural Evolution in Data Warehousing," White Paper No. Eleven, July 1, 1998, at http://www.dkms.com/ White_Papers.htm

[16] Ibid.

[17] The ideas for the AKM owe much to the following White Papers. Template Software, "Integration Solutions for the Real-Time Enterprise: EIT - Enterprise Integration Template," Dulles, VA, White Paper May 8, 1998. See also

http://www.template.com. Persistence Software, "The PowerTier Server: A Technical Overview" at http://www.persistence.com/products/tech_overview.html, and John Rymer, "Business Process Engines, A New Category of Server Software, Will Burst the Barriers in Distributed Application Performance Engines," Emeryville, CA, Upstream Consulting White Paper, April 7, 1998 at http://www.persistence.com/products/wp_rymer.html. Two other products that could be used to develop the AKM component are DAMAN's InfoManager (inquire at http://www.damanconsulting.com), and Ibex's DAWN workflow product along with its ITASCA active database (at http://www.ibex.ch/)

[18] A recent white paper of mine "Knowledge Management Metrics Development: A Technical Approach," treats KDD as a use case in the DKMS. See it at http://www.dkms.com/White_Papers.htm.

[19] See John Rymer, "Business Process Engines . . ." op. cit.

[20] This is a very close paraphrase of John Rymer's statement in Ibid. P. 2. It is not quoted because I omitted bullet points and added the clarification on the use of agents.

Biography

Joseph M. Firestone is an independent Information Technology consultant working in the areas of Decision Support (especially Data Marts and Data Mining), Business Process Reengineering and Database Marketing. He formulated and is developing the idea of Market Systems Reengineering (MSR). In addition, he is developing an integrated data mining approach incorporating a fair comparison methodology for evaluating data mining results. Finally, he is formulating the concept of Distributed Knowledge Management Systems (DKMS) as an organizing framework for the next business "killer app." You can e-mail Joe at eisai@home.com.

[Home] [Up]
[Data Warehouses and Data Marts: New Definitions and New Conceptions]
[Is Data Staging Relational: A Comment]
[DKMA and The Data Warehouse Bus Architecture]
[The Corporate Information Factory or the Corporate Knowledge Factory]
[Architectural Evolution in Data Warehousing]
[Dimensional Modeling and E-R Modeling in the Data Warehouse]
[Dimensional Object Modeling] [Evaluating OLAP Alternatives]
[Data Mining and KDD: A Shifting Mosaic]
[Data Warehouses and Data Marts: A Dynamic View]
[A Systems Approach to Dimensional Modeling in Data Marts]
[Object-Oriented Data Warehousing]