

The Irresistible Move towards Interoperable Database Systems

Philip E. Cannata

Microelectronics and Computer Technology Corporation
Object Oriented and Distributed Systems Laboratory
3500 West Balcones Center Drive
Austin, TX 78759-6509
philpec@mcc.com

The successful use of interoperable database technology in major applications has not been widespread to date. Most industrial applications rely on highly centralized data resources, and data used by one application is rarely shared by other applications. This situation is rapidly changing to one where applications commonly share data resources and isolated application/data pairs may soon be the exception. The move to these applications with interoperable database systems is irresistible for a variety of reasons:

(1) Corporate data, which has been accumulated across global corporate operations on various incompatible systems, is now recognized as a strategic asset. Because of this, many large corporations are developing interoperability architectures and implementing new systems in accordance with these architectures.

An example of such is Bellcore's OSCATM architecture [1]. OSCA is a strategic architecture to be used by Bellcore and all of the Regional Bell Operating Companies in the United States to provide software product interoperability. It is being developed to allow telecommunication operation systems that are distributed over a variety of computing environments to interoperate. It is intended to accommodate a wide variety of computing environments; to allow a wide choice of database management systems and application architectures; and to be compatible with a variety of communications architectures and data architectures. In OSCA terms, interoperability is the ability to interconnect software products irrespective of their supplier and vintage, to provide access to corporate data by any authorized user, to deny access to any unauthorized user, and to maintain that interconnection and access over changes in supplier and vintage. The OSCA architecture is designed to allow heterogeneous software products to operate as a system-of-

systems in a loosely coupled, distributed configuration so as to achieve end-to-end automation, corporate data access, and rapid deployment of advanced technologies.

It is important to note that these architectures do not mandate an instantaneous jump to a resystemization of an entire business. They are designed to facilitate a transition from today's closed systems, where information is held captive, to tomorrow's open systems, where information is a shared resource. Ann Fuelberg, Executive Director of Information Resources for the State of Texas, recently said at MCC,

"In the State Strategic Plan of Information Resources the goals for the state are to move towards open systems, towards connectivity when appropriate, towards standards, and basically, towards a change of mind set—the mind set being that information is important to the state as a whole. It's an important asset. A lot of money is spent in gathering it, entering it, manipulating it, and reporting it. The uses of information can be valuable to state decision makers as well as agency leadership in accomplishing their missions and goals. So, in the state strategic plan, we are trying to get people to work towards a broader perspective of technology and information... I would love to see a (research) project that had an open systems, open standards base by which we could demonstrate that these things can occur and they do happen and they do work, and they work now. It would be very important to demonstrate and to show to agencies, because we all have been through conversions, we all know the painful process of change, and it is always beneficial to show somebody it can be done and not be so painful."

(2) There is a growing willingness by competing information system vendors to cooperate. An

¹OSCA is a trademark of Bellcore.

example of this is the SQL Access Group (SAG), which consists of a large number of leading systems and software vendors. The Group has developed a set of specifications that will enable an SQL-based application developed by one vendor to access data on an SQL server supported by another vendor. This will make it possible for DBMSs and application tools from different vendors to work together. The Group is planning a demonstration of this interoperability in the middle of 1991. The specifications and demonstration will be based on the ANSI SQL standard and ISO's Remote Data Access (RDA) work. The demonstration is intended to show that two major technical barriers to interoperability (mutually incomprehensible SQL dialects and communication protocol incompatibilities) are being torn down.

(3) The emergence of OSI as a useful tool may finally mean that communication aspects of computers can now be viewed as a commodity. With so many vendors having the ability to produce similar systems, standardized, rather than proprietary, functionality will be the distinguishing characteristic of successful corporations that gain market share. In a recent interview at MCC, George Emmanuel, Manager of Manufacturing and Database Systems Consulting Services for Hughes Aircraft Co. and a member of the board of directors for Computer Associates, said, "There is absolutely no question in my mind that OSI standards are winning out and will continue to win out. I believe and I predict within the next couple of years that vendors will be singing the song of OSI, as if they found OSI, as if that is their differentiating advantage over other vendors. And I predict that within a year or two we'll see advertising campaigns claiming who is more pure OSI." To succeed, information system providers will have to be proactive, not reactive. They will have to aggressively use standards as they are adopted. Otherwise, the standards will simply become order forms for their competitors.

(4) Dramatic changes in the world-wide telecommunications fabric will be seen in the 1990's. This decade will be remembered for a massive overhaul of regulations governing the telecommunications industry in the United States and the introduction of remarkable new digital networking technology. The slow, unreliable, expensive, long-haul networks linking geographically dispersed data resources will disappear. They will be replaced by a high-capacity, transparent, ubiquitous telecommunications network that will serve as a global information infrastructure. If you can call *anyone* in the world from *anywhere* in the world, shouldn't you be able to locate and manipulate any relevant data in the world from anywhere in the world?

The global digital network will reach not only

businesses, government, and schools, but homes as well. It will carry not only numbers, text, and speech, but high definition pictures and video. It will support an information infrastructure of products, services, and capabilities that will forever change the nature of newspapers, periodicals, and book publishing; of television, broadcasting, and movies; of education; of advertising; of shopping; of work; of travel and tourism; of business enterprise integration and computerized commerce, including manufacturing; of financial institutions; and of government.

I think the database interoperability research community has a responsibility to recognize and understand these driving forces and formulate research plans that focus on problems faced by industry as it tries to move to this new paradigm for application development and help position the world for the explosion of global digital networking.

There is a great deal of research needed to help develop efficient implementations of corporate interoperability architectures. When asked about OSCA at MCC, Bob Martin, Vice President of Software Technology and Systems at Bellcore, and who has the responsibility of implementing OSCA compliant software systems, said, "The main idea in OSCA is that there are two things that are important to share: data and user interface. And since those are two important things, OSCA decomposes the system into its user interface on one end, and data on the other end, and puts all the rest of the junk in the center. Associated with this notion are two very complex technological issues: 1) What is meant by a corporate database, how do you control its access, what are its semantics, and how does one manage redundancy? 2) Associated with user interfaces, how do you create amazingly simple human interfaces that hide the complexity of all this junk underneath."

At MCC, we have a project called Carnot that is doing the research necessary to develop tools that will allow development of open applications that can be tightly integrated with information stored on existing, closed systems. These tools will allow our participants to achieve efficient implementations of architectures such as Bellcore's OSCA. We are working directly with the SQL Access Group, and we are building everything on top of existing standards, paying particular attention to OSI. The purpose of the Carnot system is to provide for the integration of a variety of resources within an enterprise and, ultimately, across global enterprises. We are committed to recognizing the thirty year investment that all major corporations have in information technology. We intend to build an elegant bridge between the closed centralized systems of yesterday and the open distributed systems of

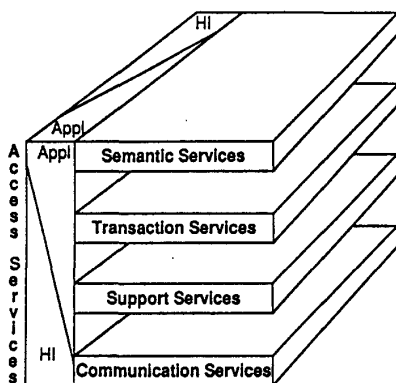


Figure 1: The Carnot architecture

tomorrow.

The Carnot system's architecture consists of five groups of services: communication services, support services, transaction services, semantic services, and access services. These are shown in Figure 1.

The communication services provide the user with a uniform method of interconnecting heterogeneous equipment and resources. While these services provide the interconnection, it is the groups of services above them that allow for their effective use. The communication services are primarily built upon the OSI standards using the ISODE [2] implementation. TCP/IP is also supported for some services and other protocols, such as MCC's ESP, OSF's DCE, and UI's ONC, may someday be supported.

The support services provide a collection of basic applications that make use of the interconnected resources supported by the communication services, and can be thought of as network-wide utilities available to applications and other higher level services. The support services consist of implementations of existing and emerging standards, such as the Sequel Access Group's Remote Data Access, and CCITT's X.400 message handling and X.500 directory services. Also included in this group of services are Electronic Data Interchange services, such as those found in the Department of Defense's CALS standards, various authentication services, and Information Repository Dictionary System ser-

vices.

The transaction services provide the basic facilities needed to coordinate the execution of work across a variety of heterogeneous resources, including databases, files, and other applications used in running a business. It is with these services that the Carnot System begins to add unique value to the standards efforts.

The main element in the transaction services is a distributed shell environment called Rosette [5, 6], which coordinates the control flow and data flow among components of the various Carnot services. Rosette is based on the MIT actor model [3]. The actor model provides a clean, well understood semantics for the asynchronous execution of distributed, concurrent services. Rosette is particularly nice because it allows for the rapid prototyping of new capabilities within either the OSI or TCP/IP environments. It also allows for remote evaluation, fan-in/fan-out data flows, virtual synchrony, and reliable broadcast. We believe that these overall mechanisms are likely to be of great benefit in providing support for fluent, nonbrittle transaction semantics.

Conventional two-phase commit-based distributed transaction processing is available as a transaction service through products from our participating vendors, such as NCR's TOPEND. The capabilities offered by these products are also coordinated by Rosette and therefore are an important piece of our relaxed transaction environment.

Distributed transaction/query generators build Rosette scripts that reflect current business realities and accumulated corporate folklore. These generators build the scripts through interactions with directory services, repository managers, and deductive knowledge in Carnot's declarative resource constraint base. The declarative resource constraint base is essentially a collection of predicates that express business rules, inter-resource dependencies, consistency requirements, and contingency strategies throughout the enterprise. The declarative knowledge base may be changed to adapt to a changing environment without modifying the application programs that generate the transactions; new applications can easily be added to the environment.

The semantic services provide a global or enterprise-wide view of all the resources integrated within a Carnot system, rather than the view that occurs at the transaction services layer, which essentially treats each resource as accessible but separate. With the semantic services, we want to have a view of the distributed environment in which we can use essentially the same language or languages for communicating among resources. We are not going to attempt to make everything appear as a

single centralized system, but rather have a common semantic framework.

We are employing MCC's CYC [4] knowledge base as a part of this framework. The idea is that the CYC ontology and reasoning mechanisms, as they exist, can be used as a backdrop for representing sets of concepts that may be found in various databases and other resources. As another piece of the framework, we are using MCC's Reasoning Architecture software to provide mechanisms for independent, communicating agents. These agents encapsulate chunks of expertise as separate active entities within the Carnot system. The framework also includes schema integration and database design tools for updating, reconceptualizing, and generating information schema.

An interesting capability that may arise with the addition of reasoning agents into the semantic services layer is an inversion of the notion of the way standards work. In the typical approach, you have standards that define a particular set of operations and ways of using those operations to accomplish tasks. With the malleability and adaptability offered by the reasoning agents, we may be able to provide an environment in which the Carnot system can adapt to the needs of the client.

Finally, the access services provide mechanisms for manipulating the other four Carnot services. They allow developers to use a mix of human interface software and application software to build enterprise-wide systems. In some situations there is mostly application code and no human interface component. This would be typical of background tasks that are being performed in a user system. In other situations there is a mix of human interface and application code. Finally, there are situations in which nothing but human interface code provides direct access to functionalities of one or more of the four layers of the architecture.

The human interface software, at present, is a set of tools that provide for graphical interaction with the Carnot environment. It supports a drag and drop type of interaction, where graphical objects can be selected ("picked up") and then dropped onto active regions of the screen, causing various actions to take place. The tools come with application libraries that can be modified and extended to build interfaces for new Carnot domains. Navigation and building appropriate graphical metaphors of the information space are the foci of existing tools. Support is also provided for structuring dialogues involving data entry, where incomplete or erroneous information may be discovered and appropriately handled. Future work includes human interface support for multiple cooperating users accessing and updating common information stores, and declarative languages for associating the se-

mantics of application data to grammars for parsing and generating graphical displays.

The timeline for all of this work began in late 1990. We plan prototype deliveries of ISODE-based tools in the second quarter of 1991. The Carnot tool kit will continue to grow through application development collaborations with end user participants in 1992 and 1993, culminating in final deliverables in late 1994. Our vendor participants will productize and support these tools as we work with them and the various standards bodies to ensure compliance with and incorporation into appropriate industry standards.

References

- [1] "The Bellcore OSCATM Architecture," Issue 2, Technical Report No. TA-ST-000915, Bellcore, Piscataway, NJ, July 1990.
- [2] Marshall Rose, *The Open Book*, Prentice Hall, Englewood Cliffs, NJ, 1990.
- [3] G. Agha, "Concurrent Object-Oriented Programming," *Communications of the ACM*, Vol. 33, No. 9, September 1990, pp. 125-141.
- [4] Douglas B. Lenat, Ramanathan Guha, Karen Pittman, Dexter Pratt, and Mary Shepherd, "CYC: Toward Programs with Common Sense," *Communications of the ACM*, Vol. 33, No. 8, August 1990, pp. 30-49.
- [5] Chris Tomlinson, Gul Agha, Won Kim, Yehoshua Sagiv, Mark Scheevel, Vineet Singh, and Rebecca Will, "Report on the Rosette Language," MCC Technical Report ACT-ODS-449-89(Q), Microelectronics and Computer Technology Corporation, Austin, TX, December 1989.
- [6] Chris Tomlinson and Vineet Singh, "Inheritance and Synchronization with Enabled-Sets," *Proceedings of OOPSLA '89*, SIGPLAN Notices, Vol. 24, No. 10, October 1989, pp. 103-112.