6.0 Information Access technology

Introduction

Information Access Technology (IAT) provides the architectural foundation for Information Systems, that is computer-based systems which gather data and process data to provide information for decision-makers. Information systems also invoke tools to develop and assess alternate courses-of-action. These tools will operate over the global communication networks now emerging, and use a diversity of computer hardware and software to achieve their results.

Many information-processing tasks can be defined independent of specific source databases and of specific recipient applications. In the IAT architecture we assign such sharable services to an active middleware layer, and define domain-specific *mediator* modules to populate this layer. Mediating services must be of value to the customers, so that their applications will access mediators rather than sources directly. Several types of value can be considered: improvement in access and coverage, improvement of content, and delegation of maintenance. Criteria for mediating modules are: ownership by party who assumes responsibility for the results of the services, domain-specificity to delimit the scope of such a responsibility, and conformance with interface standards that enable interoperation.

Applications that benefit from mediation include planning and other decision-making tasks that require information from diverse resources (e.g., databases, reference systems, data obtained from sensors, and analyses projecting trends into the future). The sources are often autonomous, some of them are public. They will be heterogeneous. The heterogeneities include: representation, scope, level of abstraction, and context.

Mediated results are intended to be composable by higher level applications, which have to solve problems involving multiple subtasks. Hence mediators need a machine-friendly interface to support those applications. This interface must provide good communication, while encapsulating the domain-specific tasks, so that the complexity of the composed system is not much greater than that of the individual subtasks. However, questions of effectiveness and efficiency do arise, and must be dealt with by exploiting the processing and storage capabilities of modern hardware.

The corresponding architecture is a generalization of a server-client model. The partitioning enhances maintainability: the applications software staff can concentrate on functional improvements, the data resource managers on operational issues, and the technical maintenance are concentrated in the mediator layer. The concept is network oriented, and it is hoped that mediating services will be provided over the network by domain specialists. On networks, mediating services can be performed by independent entrepreneurs.

6.1 Information Access Technology

Information Access Technology (IAT) provides the framework and foundation services for accessing and shipping data of potential relevance to an Air Force decision maker. Central in this task are the conversion of data from the many heterogeneous data resources available for logistics, plans, local status, and intelligence into an integrated information base that provides situation awareness. In addition to transmitting data IAT needs to support ongoing access to software tools, which carry out the tasks. Some of the tools fall within the IAT domain, such as rule-based interpreters which describe the needed data reduction for a particular application. Others are from their own support domains, such as image processing software for fusing intelligence data, which is invoked when needed for an IAT task.

As information systems are increasing in scope, they depend on many diverse, heterogeneous resources. These resources are typically developed and maintained separate from the majority of the applications that use their results. The applications that motivate the establishing of a database tend to be operational: inventory control, payroll, production control and the like, These data soon become important to support high-level applications: planning and decision-making. Decision support applications are typically designed subsequently and independently. Planning support is synchronized with objectives that change rapidly, and has to rely on existing sources, since it is rare that sufficient time is available to build planning systems and their data collection from scratch. Other sources of information are data systems such as digital libraries, geographic information systems, and simulations.

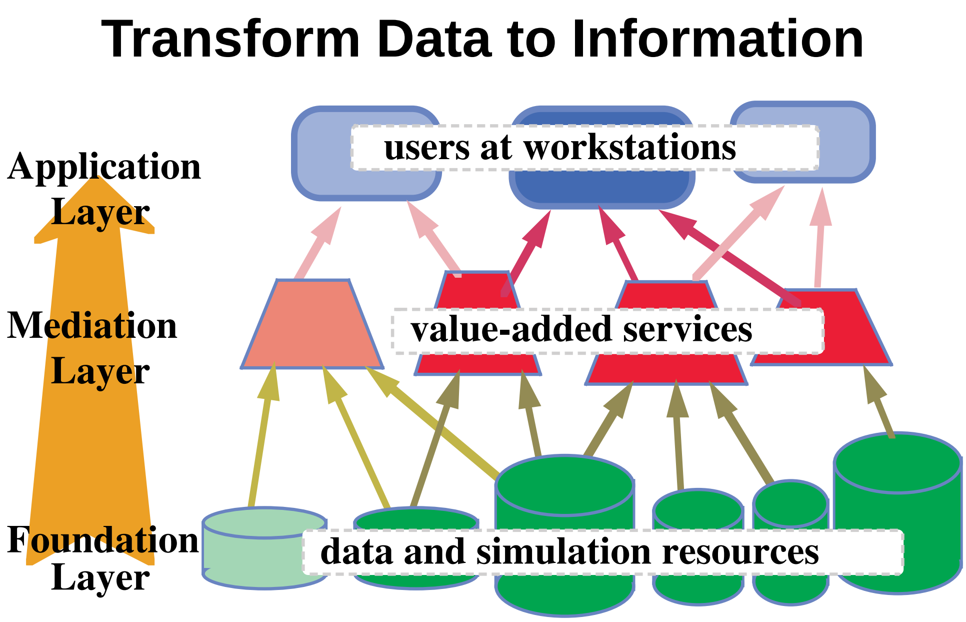

Figure 6 Transforming Data to Information

Dealing with many, diverse, and heterogeneous sources overwhelm high-level applications with excessive emphasis on application-irrelevant but crucial details.

Since integrating many streams of data further increases the volume of data that can be delivered, a critical and implied task of IAT services is the reduction of this excessive volume to a manageable amount of essential information.

It is crucial to avoid information overload at the decision-maker by presenting only relevant information at the appropriate level of abstraction. That level will differ for an airman on the maintenance line, a pilot on a mission, or the commander of a joint task force. Information overload also has to be avoided when lower bandwidth links are to be used, often in the last 100 miles to the warfighter.

6.2 Mediators

Mediators are the modules which provide intermediary services in an IAT architecture, linking data resources and application programs. Their function is to provide integrated information, without the need to integrate the data resources.

Specifically, the tasks required to carry out these functions are comprised of:

1. Accessing and retrieving relevant data from multiple heterogeneous resources
2. Abstracting and transforming retrieved data so that they can be integrated
3. Integrating the homogenized data according to matching keys
4. Reducing the integrated data by abstraction to increase the information density in the result to be transmitted

These are shown in Figure 6.

Advanced IAT services must also access simulation software, so that the decision maker can also develop and compare multiple courses-of-action. Overall, this processing adds value by converting data to information.

Definition: A mediator is a software module that exploits encoded knowledge about certain sets or subsets of data to create information for a higher layer of applications. It should be small and simple, so that it can be maintained by one expert, or at most, a small and coherent group of experts.

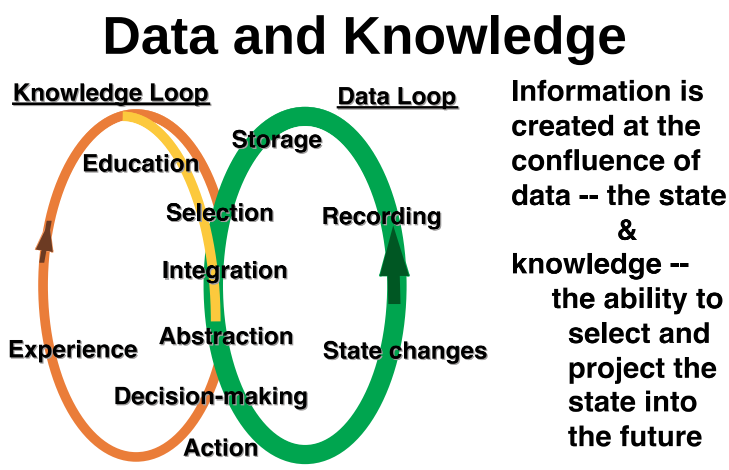

Figure 7 Interaction of Data and Knowledge

Access is a necessary prerequisite to deal with distributed information. Well-known examples of approaches to improve access are knowbots, facilitators, a variety of webcrawlers, and multi-database systems.

6.3 Module Types

We recognize two types of IAT modules: *mediators and facilitators*; IAT systems can be composed out of one or both of these. A mediator provides data processing functions, and is managed by a domain specialist who determines the functions and supports the needs of the recipient. The program in a mediator is typically an expert system, with knowledge about: the data sources; their significance; data transformations appropriate to the domain; and the information structure preferred by the recipient. A facilitator is a module which searches the network for sources of data, and delivers newly discovered data to a mediator and its maintainer, or directly to an end-user recipient.

In mediation more effort is devoted to processing the results of the retrievals (task types 2-4) than to access. Because of the processing effort, and the expectation that mediation services can support multiple applications, mediation is done in distinct modules in the communication networks, as sketched in Figure 8 (a logical progression of technology).

Mediated IAT technology depends on having adequate and secure communication networks, and builds on the technological infrastructure currently being established. Mediation technology augments the communication network with software modules, that typically will execute at intermediate nodes on the network, placed somewhere between the data sources and the computers used by the recipients of the information.

6.4 Status of This Technology

Mediation is an architecture which extends the client-server model and provides scalability. The sketches in Figure 8 show the development from:

(a) Client-server applications in a setting that assumed compatibility (dealing in a single domain, often dominated by local data) to

(b) The addition of wrappers to deal with legacy data and an integrator to combine resources, to

(c) The sharing and associated generalization of the integrators, to

(d) Domain-specific services performed by independent mediators on the network that are not bound to the client or to the server.

(e) Finally, applications that require information from multiple domains will use multiple mediators.

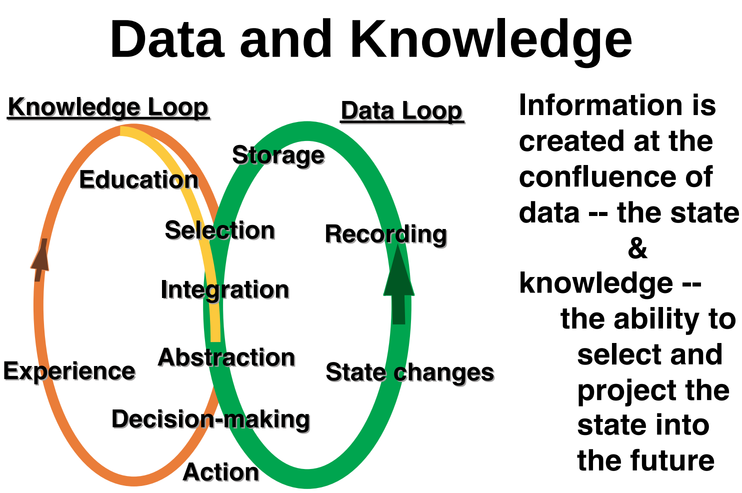

Figure 8 Network services from Client-Server to Mediation Layer

Mediators are being built now by doing careful domain knowledge acquisition and handcrafting the required code. Building mediators by hand is essential to validate the concept and establish the needed standards for the interfaces. Since the mediator approach is conceptually comprised of three layers, as shown in Figure 9, there will be two major interfaces:

a: Base resources to mediation

b: Mediators to applications

In practice there will also be intermediate interfaces, since it is likely that within the layer a number of sublayers exist as well. For these the technologies of type b. are also appropriate.

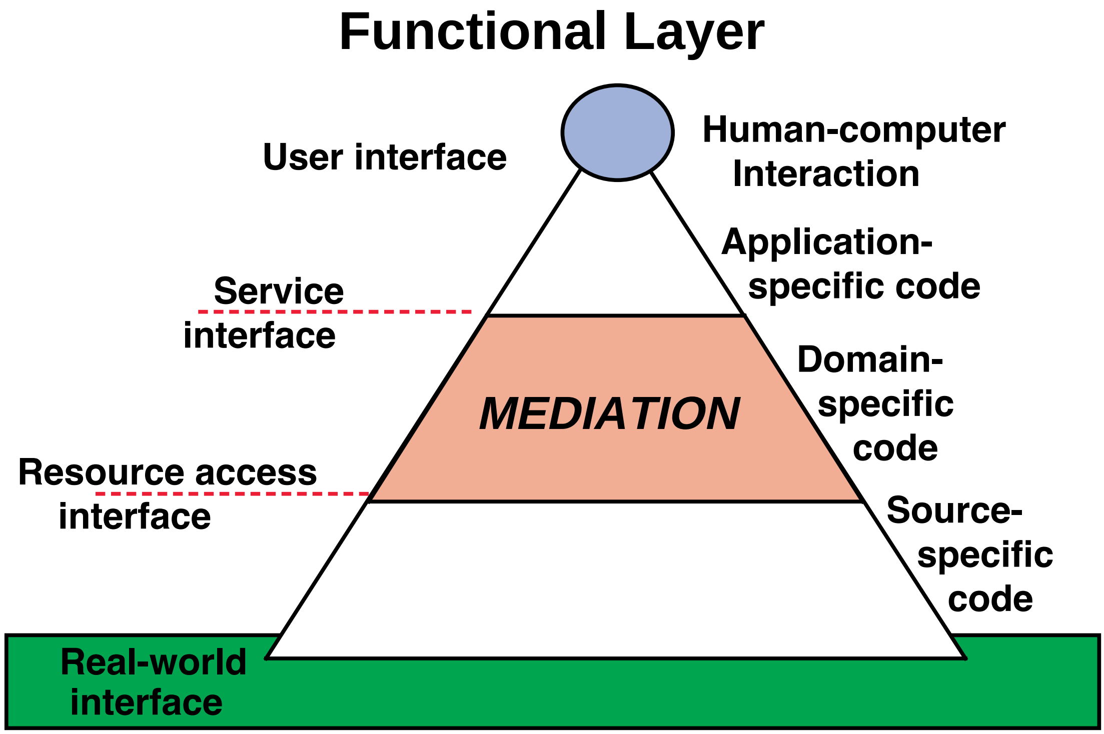

Figure 9 The Mediation Layer

6.4.1 Interfaces

For the base interface the many tools that are becoming available to serve the two-layer server-client model are appropriate. These include:

a. Distributed and augmented SQL

b. The interface tools for object-oriented access, such as the object fetching protocols of COBRA

At the application layer the interfaces need greater capabilities. Here agent languages are appropriate (See Intelligent Agents, Chapter 5). The sender and the receiver must agree on the chosen representation and on the vocabulary and its structure, i.e., the so-called ontology.

6.4.2 The User Interface

Mediation is simplified by delegating the complexities of the customer interface to the application program. Code to deal with the variety of graphical user interfaces (GUI), interface devices (screens, pop-up, roll-down, scroll, cut-and-paste, drag-and-drop, speech, etc.) often occupies more than 70% of application programs. Mediators and the invoking applications only need a machine-friendly interface, as represented by a KQML application program interface (API) block.

Trading internal code for an API recapitulates an earlier paradigm shift. We recall that in the 1960s the programming of file operations took a major amount of effort and competence. The acceptance of the database paradigm removed that code from applications and delegated it to database management systems (DBMS). The DBMS provides all the really difficult and specialized code associated with managing files, as backup, recovery, integrity, and internal consistency management.

Mediating modules can either provide information according to a pre-planned schedule, according to triggers that indicate that a significant change has occurred, or in response to a request. Principle functions that can be incorporated in a mediator include:

a. Adaptation to the location and format of a data resources

b. Selection of relevant data, aggregation of data into symbolic form to provide an effective abstraction for intelligent analysis

c. Fusion of symbolic information according to knowledge-based rules

d. A transformation to present the information to the recipients' computers for final integration of information from multiple domains and display of actionable items.

6.5 Services Provided by Intelligent Mediators

IAT systems provide services to information consumers. The nodes in the network that provide such services are the mediators. Value-added services in a mediator include combinations of:

1. Determination of likely resources using information extracted earlier
2. Invocation of wrappers that deal with legacy sources
3. Selection of relevant source material
4. Optimization of access strategies to provide small response times or low cost
5. Imposition of security filters to guard private data
6. Resolution of domain terminology and ontology differences
7. Resolution of scope mismatches
8. Interpolation or extrapolation to match differences in temporal data
9. Reduction of historical data to limited snapshots
10. Abstraction to bring material to matching levels of granularity for integration

11. Integration of material from diverse source domains based on join keys
12. Omission of replicated information
13. Assessment of quality of material from diverse sources
14. Pruning of data ranked low in quality or relevance
15. Omission of information already known according to the user model
16. Statistical summarization into higher level categories as defined in the user model
17. Reporting exceptions from expected values or trends
18. Triggering of actions due to exceptions from expected values or trends
19. Resolving temporal asynchrony in source data by referring to past, cached data
20. Completing current state information from past data by invoking simulation programs
21. Projecting status data into the future using simulations for Course-of-Action planning
22. Transformation of material to make presentation effective for the customer
23. Adaptation to the bandwidth and media capabilities of the customer
24. Transmission of resulting information and meta-information to the customer application

The conceptual underpinning for automating the mediation process is based on knowledge acquisition to provide the added value. Tools from the field of artificial intelligence that are a major contributor in mediation.

6.6 Where does Mediator Technology Stand Today?

Projects that are using mediators today have included manufacturing systems, such as the design of a new fighter aircraft at Lockheed; and gimbals for antenna positioning on spacecraft at Lockheed Space Systems. Early applications have been in integrating information for flexible military command systems. The technology for these projects is supplied by a variety of vendors, ISX corporation serves as a contact point for these projects.

6.7 Future Mediators for Effective IAT Systems

Technologies must be developed for automation of the building of mediators. Since the technology is modular there is promise in that direction, but the current short commercial funding cycles do not allow the development of the technology, since individual mediators, as now employed in demos and pilot situations can be more rapidly hand-crafted, one-at-a-time.

In order to build a mediator, resources have to be identified. This task is a common task carried out on the networks to enhance browsing, but not with the intent of building persistent paths and reusable software. Knowbots and facilitators provide the required base concepts. Once suitable data resources are located on the net, they have to be placed into a user-sensible hierarchy, as sketched in Figure 10. To achieve the required functionality, tools have to be attached to the nodes that provide for fusion and summarization. Some current ARPA funding is focused on that goal, but it has to be sufficiently long-term to provide the vendor community with credible science.

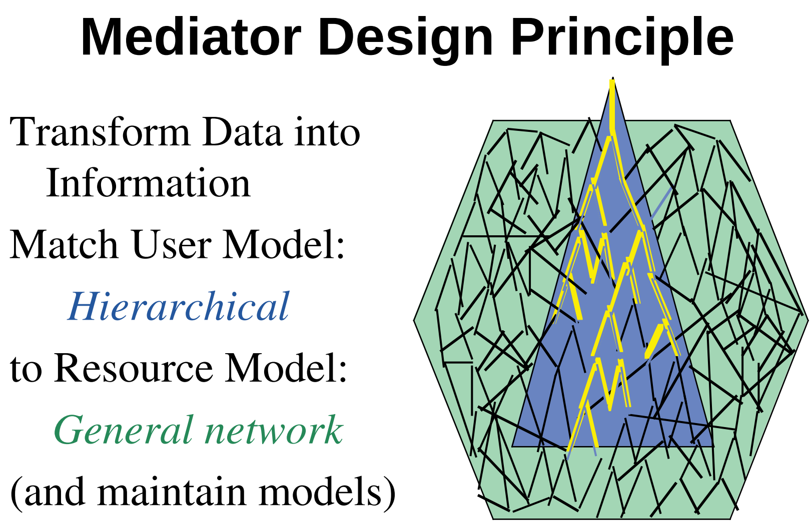

Figure 10 Mediator Design Principle

Beyond the initial creation of mediating modules is the need to update them. Where mediators are built with rule-based declarative technologies an excellent opportunity exists for automated adaptation, using machine learning technology developed in the artificial intelligence domain (See Artificial Intelligence, Chapter 7).

6.7 A Partitionable Architecture for IAT

The amount of knowledge to perform all possible aspects of mediation is enormous. Modularization or partitioning is the divide-and-conquer tool used in science to deal with excess size and complexity. Here size and complexity go together, because as the number of concepts grows, their linkages, their distinctions based on context, and their subsumption relationships also grow, and that growth is more than linear.

Since the number of domains is large we cannot foresee a central solution to that problem, and hence will have to deal with subsets of the domains, and bring those subsets together, as and when needed.

6.6.1 Domain-Specific Partitioning

Just as we specialize people in organizations, so we partition knowledge by domains. We define a domain to be a subset of knowledge in the world that can be managed by a single person, or at least by a small coherent group, as indicated in Figure 11. Within a domain no compromises should be necessary to define terms, and no committee effort should be required. It is best if there is an organization willing to take responsibility for a domain (e.g., the Defense

Logistics Agency for services supplying materiel; the Automotive Engineering Society for problems with cars).

The knowledge in a mediator module has to be maintained, since resources and user needs change over time. It is important to associate a mediator with a domain expert, responsible for the maintenance of its content.

Having the knowledge modularized in a domain-specific mediator allows convenient replications, since the mediators are programs that can be replicated onto multiple nodes on the network. The location of a mediator can be determined to optimize bandwidth, enhance security and be effective in terms of assignments of domain experts. The functionality of a mediator should not be affected by its location.

Specialization makes maintenance feasible. An expert owner can handle a specialized domain, without having to consider the different constraints imposed by handling unrelated domains, say, religion versus carpentry when discussing miters. Differing domains may also be best served by different programming paradigms, say finance versus engineering. Other subsets of domains may use similar technology, but differ in the concepts and structure of their knowledge representations, say electronic versus civil engineering.

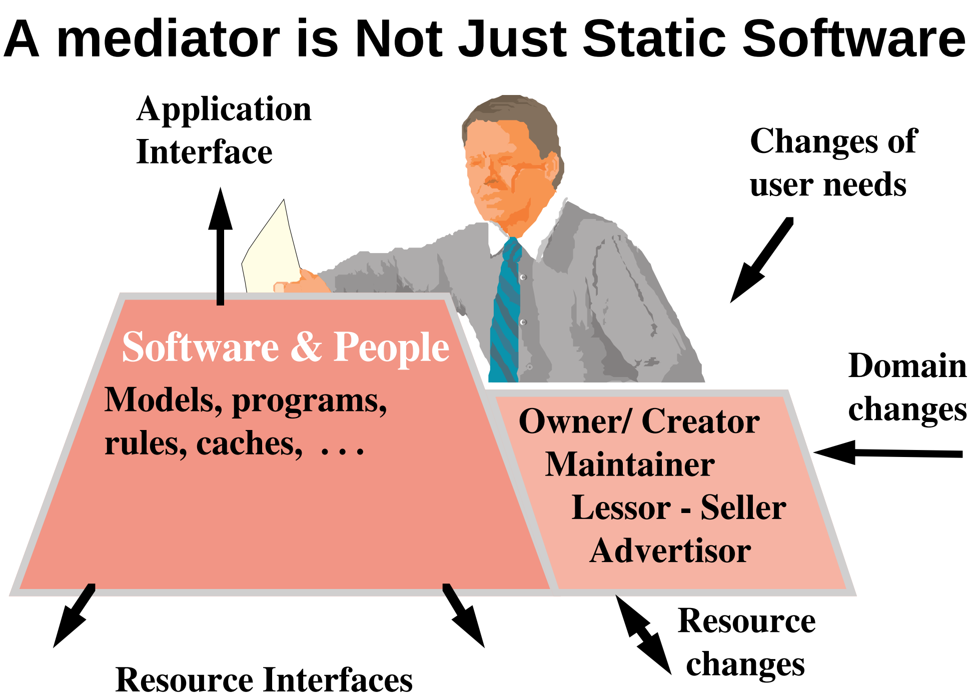

Figure 11 A Mediator has an Owner

Information from multiple domains still has to be integrated at the workstation of the the decision-maker, as indicated in Figure 12. The mediators provide functions that can be delegated to subordinate services, such as logistics and intelligence, but do not carry out the vital decision tasks of the commander in the field.

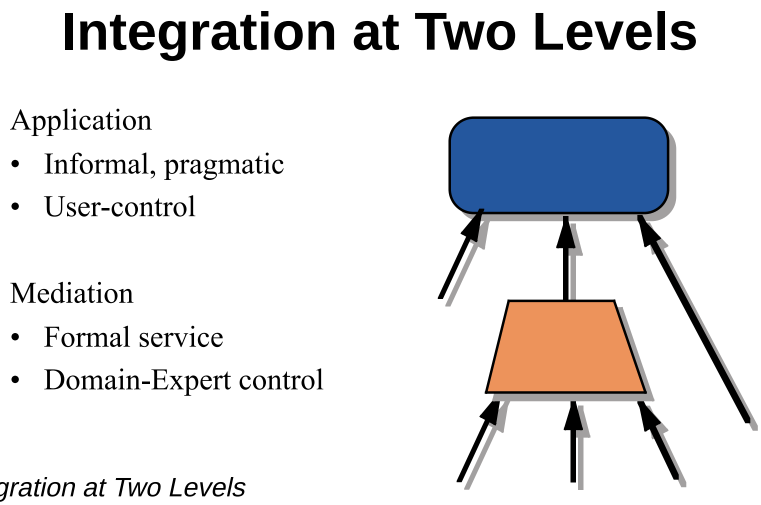

Figure 12 Integration at Two Levels

6.6.2 Distribution

Mediation is achieved by software. A mediator transforms data available on the network, to make it more suitable and relevant to the consumer. This software function can be carried out on the computer where the mediation was developed, or can be allocated to other computers on the network. Since software is easy to copy over the digital networks, mediators can be rapidly moved and replicated. Software is typically much smaller than the data it processes, so that it is easy to install at any location where compatible computers are available, as shown in Figure 13. Reallocation provides optimization for the IAT system, as well as an opportunity to tailor dataflow to the available bandwidth, a crucial concern once we leave the IAT backbones and proceed onto the difficult "last 100 miles" to the warfighter.

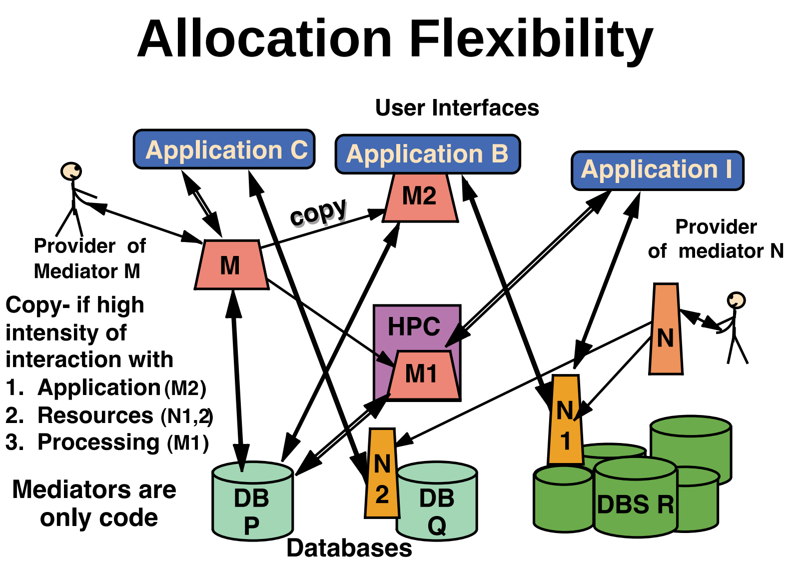

Figure 13 Allocation Flexibility

It is crucial that experience be gained with IAT concepts involving modern technology, so that rapid assembly of IAT systems will be possible when needed to assemble mission nets for new situations and rapidly assembled task forces.

Integration of simulation results will be defense-specific and must be integrated so that results of developing stand-alone simulations can become a part of ongoing tactical support systems.

If the processing algorithms used in mediation are costly, the software can be moved to a high performance processor on the network, again improving response time. Some of the features associated with a mediator architecture are sketched in Figure 14.

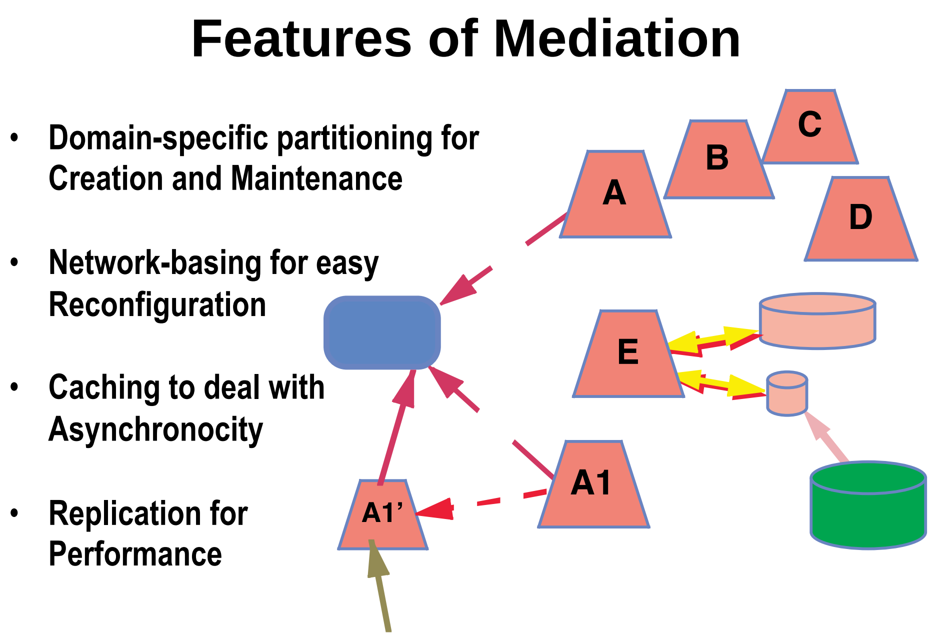

Figure 14 Some Features of Mediators

If demand for a particular type of mediator is strong, replicates can be distributed to additional sites in the IAT network. Since now the communication links will shorten, response time is enhanced as well. It is also possible to store acquired data or intermediate results to enhance the performance of mediators. For instance, detailed terrain maps are best kept on a node close to the customer, so that only changes have to be transmitted when they occur.

6.6.3 Standards

Since we assume that most mediation services are performed in autonomous computing nodes, the requirement for interoperation is the ability to communicate according to some standard conventions. Heterogeneity of computing platforms, operating systems, and message passing infrastructures is being overcome by concerted efforts in many communities. For mediation we must also consider terminology and representation conventions.

A number of projects are now using the mediator concept, and some standards (KQML is an example) are being promulgated for the interfaces that are required. A series of prototype efforts supports the US Air Force concept of an Integrated Weapons Systems Database. A phase is sketched in Figure 15. These prototypes allow for the gathering of experience that can validate the proposed standards. Investment will be needed to provide commonality for that effort, since there is a tendency for independent suppliers to compete in terms of leadership in standards as well, disabling the objective of vendor independence for the purchasers.

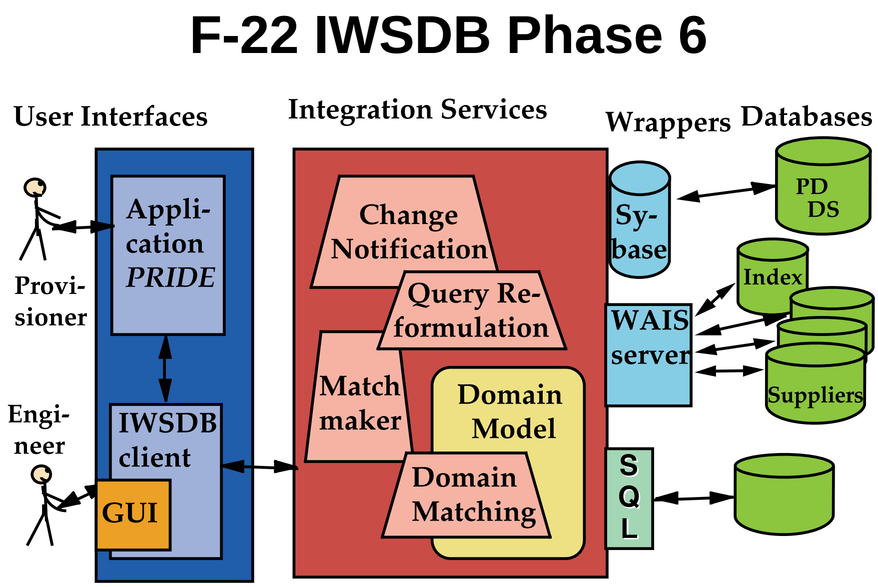

Figure 15 F-22 Integrated Weapons Systems Database

6.6.4 Maintenance

Mediation adds value to the data by applying the expert knowledge of the expert who has created the mediator. Mediators should also be maintained by those experts, so that the mediators remain effective in a constantly changing world. As soon as an improved mediator is developed it can be advertised over the network, both to existing subscribers as well as to potential new clients. A poorly maintained mediator will lose value over time, and be a candidate for replacement by a competitor.

6.7 Conclusion

Information technology is serving us well in specific domains, although we have remained dependent on specialist model designers and programmers for our implementations. Object technology has lessened our dependence on specialists by being able to use an infrastructure that aggregates detail into meaningful units.

Mediation is a technology that is intended to scale systems so that sources from many domains can contribute services and information to the end-user applications.