Presented at COOPIS, Vienna, 1995

The Basis for Mediation

Gio Wiederhold and Michael Genesereth
Stanford University
February 1995, minor updates September 1997
A later version of this work is appearing as
The Conceptual Basis for Mediation Services in *IEEE Expert*, November 1997

Abstract
Mediator modules provide intermediary services in information systems, linking data resources and application programs. Programs that led to the concept of mediation were either constructed to support specific applications, or extended services from databases. Mediators are being built now by careful domain knowledge acquisition and hand crafting the required code.

In this paper we present the conceptual underpinning for automating the mediation process. Our goal is not fully automatic code generation, since programming knowledge is necessary to provide the added value that offsets the costs of having the extra architectural layer due to formal mediation. The concept is based on the extraction of a hierarchical domain model out of the general network representing the available resources. Associated with the method are domain ontologies. Ontologies list the terms used by the models, and document their relationships. These terms provide the semantic foundation needed to perform the match operation.

This paper is prescriptive, rather than a research report. The objective of publishing it now is to gain conceptual coherence among the projects that have adopted a multi-layer information architecture, while the implementations can differ greatly.

1. Introduction

As information systems are increasing in scope, they depend on many diverse, heterogeneous resources. These resources are typically developed and maintained autonomously of most of the applications that use their results. While immediate applications tend to be inventory control, payroll, production control and the like, eventually the data become important to support high-level application, as planning and decision-making. Decision support applications are typically designed subsequently and independently. Planning support is synchronized with management objectives, 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 systems as digital libraries, geographic information systems, and simulations.

Dealing with many, diverse, and heterogeneous sources overwhelms high-level applications with excessive emphasis on irrelevant, but crucial details. Platform and operatings systems differnces are commonly recognized, and are getting easier to manage. Differences in interface protocols, data description, level of abstraction, and the precise meaning of terms being used do not yet yield well to automation. Mediators provide intermediary services, linking data resources and application programs [W:92C]. 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 Processing the integrated data to increase the information density in the result.

as shown in Figure 1. Overall, this processing adds value by converting data to information. More detail is provided in Table 1.

One should note that in mediation more effort is devoted to processing the results of the retrievals than to access, i.e., locating and gathering the data. Those access tasks are the current emphasis of much work in information systems, and are a necessary prerequisite to deal with distributed information. Well known examples of approaches to improve access are systems as KNOBOTS and the variety of WEBCRAWLERS, as well as multi-database systems. [Dunstan:89, Elmagarmid:90, BCGP:92, HBK:93].

Earlier programs that exhibited such functions and led to the concept of mediation were either constructed to support specific applications, or extended services from databases. Mediators are being built now by careful domain knowledge acquisition and hand crafting the required code. Building mediators by hand is essential to validate the concept and establish the needed standards for the interfaces. Since the mediator architecture is conceptually comprised of three layers, as shown in Figure 2, there will be two major interfaces:

a Mediators to Applications

b Base resources to Mediatiators

In practice there will also be intermediate interfaces, since it is likely that within the mediation layer a number of sublayers exist as well. For these the technologies that support the Mediator to Application interface are also appropriate.

Slide 2 goes here

Figure 1. Transforming Data to Information

1.1 Architecture and its Support

Mediation is primarily an architectural concept. The precise implementation of a module is less important than its ability to perform its function in a larger systems context. In buildings we can have big doors, small doors, wooden doors, doors with locks, etc. But all doors have to fit into the openings that are provided in the building. Using non-standard door openings increases the cost, reduces the flexibility to chose among types of doors, and may hinder moving equipment in and out of rooms. Much of the effort in moving to sharable mediator architecture involves the recognition of interface standards, so that configurations, i.e., instances of the architecture, can be rapidly assembled. There is a strong linkage here to concepts as virtual enterprises [Rheinhold:93]; mediation provides the required openness in the architecture.

For the base interface the many tools that are becoming available to serve the two-layer server-client model are appropriate, as distributed and augmented SQL [DD:93] and the interface tools for object-oriented access, as CORBA [OMG:91]. At the application layer the interfaces need greater capabilities. Here agent languages, as KIF for content and KQML for managing the kinds of multi-user, multi-representation, and multi-function operations are appropriate [GK:94, PFPMFGR:92]. While KIF is defined to exchange knowledge in a first-order logic representation of rules, KQML also can handle other representations, as engineering objects using STEP (The ISO 10303 standard for Product Data Exchange) [STEP:92], mathematical equations [FCGB:91], and even tuples. The requirement is, of course, that the sender and the receiver agree on the chosen representation, as specified in the preamble of a KQML statement [NFFGPSS:93]. The sender and the receiver must also agree on the vocabulary and its structure, i.e., the *ontology*, specified in the preamble [Gruber:93]. This aspect will be the focus of Section 8 of this paper.

Mediation is simplified by delegating the complexities of the customer interface to the application program. Code to deal with the variety of graphical user interface (GUI) devices, as windows, 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 between them, as represented by a KQML application program interface (API). The mediator will need a third interface to allow to interact with the owner of the mediator for its maintenance. That interface can be a friendly GUI.

Trading internal code for an API recapitulates an earlier paradigm shift. We recall that before database management systems existed, 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 DBMSs now provides all the really difficult and specialized code associated with managing files, as backup, recovery, integrity, and internal consistency management. The DBMS-based resource and its application share a model, as represented in the schema, and this concept will be a basis for the mediation paradigm presented here.

2. Services

In this paper we present the conceptual underpinning for automating the mediation process. Automation does not mean fully automatic code generation, since knowledge is necessary to provide the added value that offsets the costs of having the additional architectural layer implied by formal mediation. In an earlier paper we presented some of the tools from the field of artificial intelligence that are of help in mediation [W:92I]. At that time the more general vision presented in this paper had not yet been formulated.

The services listed in Table 1, and related services specialized for domains as finance, scheduling, etc., are now available in some or other existing applications, but are rarely identified as composable and reusable tasks. Many more citations than the ones listed are possible. Placing them into mediators within this architecture enables their reuse, and partitions information processing tasks by functionality and domain.

Slide 5 goes here

Figure 2. The Mediation Layer

Value-added services in a mediator include combinations of:

1 Determination of likely resources using information extracted earlier[Gravano:94]
2 Invocation of wrappers deal with legacy sources [Franchetti:94]
3 Selection of relevant source material [Salton:90]
4 Optimization of access strategies to provide small response times or low cost
5 Imposition of security filters to guard private data [Ware:93]
6 Resolution of domain terminology and ontology differences [W:94]
7 Resolution of scope mismatches [W:92I]
8 Interpolation or extrapolation to match differences in temporal data [W:93]
9 Reduction of historical data to limited snapshots [Snodgrass in TCGJSS:93]
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 from the customer model
16 Statistical summarization into higher level objects, as defined in the customer model
17 Relaxation of search terms to satisfy query expectations [Chaudhuri:90], Chu:92]
18 Reporting deviations from expected values or trends
19 Triggering of actions due to exceptions from expected values or trends

20	Transformation of material to make presentation effective for the customer
21	Adaptation to the bandwidth and media capabilities of the customer
22	Sending the information and meta-information to the customer application

Table 1: Services to be provided in mediators.

2.1 Value

To warrant implementation of a service as a distinct module there must be enough added value to overcome the cost of adding a layer and its interfaces in the information processing flow. But the costs and benefits to be considered are only partially related to performance. Having identifiable and maintainable service modules provides significant long-term management benefits. Some of these services will be best provided by independent enterprises over the networks, or their programs can be leased to provide these services at customer sites, as detailed for digital libraries in [W:95A]. More specifics about these services will be found in [W:95D].

2.2 Assigning responsibilities

When these tasks are performed in a two-layer, client-server, architecture they are either built into customer applications or database services. When built into applications the applications have difficulties keeping up with the variety and change in resources, and where resources are used for multiple flavors of results, much redundant and potentially inconsistent efforts are expended. For instance, when data reduction is bound to the databases, the abstractions are often not those that are most relevant to the customer, and integration from multiple sources is unlikely.

Models

The mediation paradigm depends on models; models of the resources and models of the customer needs. Since models can have a wide variety of flavors, this section focuses on making the model requirements for mediation more specific. Three well-accepted concepts combine to create the resource model:

- Schema integration
- Connecting schema entities with relationships
- Reducing the volume of requested data by using views.

Each of these concepts needs some formalization, so that mediation and the generation of mediating programs can become reliable and predictable.

3. Resource Models

Each of these resources can be described by its own model. The modeling technologies differ greatly, and their integration is a challenge when building mediators.

3.1 Relations and their structure

In a relational database the schema describes the base classes of the model. During the database design phase an Entity-Relationship model may have been used. We use our particular form of that model, the *structural model*, which uses relations and connections among them. The connections also have algebraic properties, which can support the transformations that enable integration [ElMasri:79]. Figure 3 presents a simple model of a health care setting. In practice, even in a single hospital the entities depicted may reside in multiple databases. Certainly, if the model is extended to cover multiple clinics, and internal and external laboratories, there will be a distributed operation, and some of the resources will actually be autonomous and heterogeneous. Connections will, of course, extend over distributed resources as well [WQ:94].

Relational schema integration and evolution is a well developed sub-specialty in the database arena, and we will here benefit by modeling an integrated schema without trying to integrate the actual databases [BLN:86].

Slide 14 goes here

Figure 3. Resources in a Health Care Setting

3.2 Object-structured Data

Object technology enables applications to use an infrastructure which aggregates detail into meaningful units in many important domains. When the underlying resources are modeled as objects, then some of the connections, the ones within the objects, have already been made explicit, although they may be encapsulated so that they are hidden from the programmer. Internally, objects have hierarchical linkages since the class definitions that control them are based on hierarchies. The use of multiple ownership connections, implying multiple inheritance within an object, is rare. For encapsulated objects those connections are not directly accessible or transformable, although it is likely that methods will exist within the objects that allow fetching of all implied data. An interface language, as IDL for OMG CORBA or ISL for Xerox ILU provides access to the descriptive meta-data needed to interpret objects [OMG:91, CJSSW:94].

Since the granularity of objects is limited there is still a considerable scope for modeling referencing connections among objects. Outside of the objects one finds the more arbitrary, i.e., non-hierarchical and multiple-ownership linkages.

Some large-scale hierarchies are being constructed in the object world with great pain and compromise. We don't believe that very large hierarchical objects are viable since conflicts of viewpoints are bound to arise. A reasonable upper limit for an object class seems to be at a 200 element granularity. The assumption that one hierarchical viewpoint is right for all occasions is demonstrably false [W:89]. Forcing unsuitable representations onto processing programs increase cost of finding and executing solutions; it is well known in mathematics that finding the right representation for a problem is 80% of the effort, and the same holds true in computing. However, computers can transform general, well-defined representations into specific ones.

3.3 Text

Textual data is also an important information resource. We find text in messages, reports, papers, and books. Much of its processing has been distinct from database activities in the past. Now the ability of linking through texts using hyper-text and World-wide-web technologies enable shared processing, where linkages, and hence information integration can join structured data to text. Furthermore, such links can point into multi-media objects, as images, video clips and speech segments, and multi-media objects can in turn have references to other objects. Such inter- and intra-document linkages are more ad-hoc than the structured connections seen in databases. Since these linkages cannot be expressed by a general model, some of the operations described later in this section will have to be deferred to execution time, and performed on the instances, rather than performed once only on the model. Since at this time we envisage that model matching may well be interpretive, the gap is not as wide as it would be if traditional data-processing and information systems had to be integrated.

3.4 The result

Taking all of these resources together, we find that the result is network of linkages. Meta-linkages or connections exist in both in the model and actual linkages exist among the instances, where they exist, and in the actual instantiation of the resources. Some simple hierarchies may have contributed to the integrated model, being derived from specific customer or application views.

4. Customer models

In general, the concept of having a user model is daunting. We will be more specific and speak of a customer model, since our information systems have many flavors of users [W:95A]. Such a customer model focuses on a task set and a domain of interest. The same user may be assume a different customer role at other times, and is then represented by a distinct model. Furthermore, we limit ourselves to concepts that can be machine-processable, since the customer's actions are always transmitted via the customer's workstation to the mediating service. We do not look into the dark recesses of the human mind.

We now have a crucial hypothesis:

Customer models are hierarchical.

Unfortunately, it is impossible to prove this hypothesis formally. All instances of user models cannot be enumerated, and if we define customer models formally, it would create a tautology in our proof. Resorting to informal reasoning we consider that people, when faced with complex tasks, categorize the processes and objects to be dealt with, so that they can apply a divide-and-conquer paradigm. Good categorizations are taxonomies with two attributes: disjointness, i.e., no object belongs to more than one category, and completeness, i.e., all objects can be classified.

The acceptance by customers of object models, which are also hierarchical in nature argues for the hypotheses that customer models can be hierarchical, and hence manageable within this paradigm.

While we can envisage systems of objects with multiple inheritance, say a person being both in the profession category of accountant and in the sports category of skydiver, these categorizations belong to distinct models. We support this distinction below by introducing domains. Having a hierarchical customer model driving a mediation process does not inhibit customer applications from integrating results from multiple mediators, and hence accommodating dissimilar domains. Such high-level integrations tend to be pragmatic, since it is difficult to apply formal comparative metrics in dissimilar domains. For instance, while employee competence and cost may be analyzed in financial measures, that basis for comparison is not appropriate for amateur sports. Figure 7 denotes the two levels of integration, here we focus on the methods within the mediators.

The need for conceptual clarity also supports the use of hierarchies. In mediators simplicity is essential. This rule holds, of course, for any software you want to work reliably. A mediating module carries out tasks to serve a customer, and a desirable aspect of such a computer-based servant is that it has a model to which permits the customer to understand its capabilities. Such a customer is initially the programmer building the application. If the application is designed for flexibility, it will also make the model available to the customer, so that it can be adapted, within the rage of its inherent capabilities, by the customer.

In database technology, a view relation, defined by a single view expression, is also essentially a hierarchy. Although the result will be represented as a table, each join in the view expression has defined a relationship. The attribute named by `WHERE` clause of a join along a relationship defines the superior element. In general, a view relation is not in normal form. Any subsequent normalization will expose its hierarchical structure. Such view relations have been deemed adequate for applications, further bolstering our hypothesis.

Slide 10 goes here

Figure 4. A Customer Hierarchy in a Resource Network

5. Model Matching

The central task in mediation is the identification of relevant resources for the customer model and the subsequent retrieval and reduction of the relevant data for a customer inquiry. We focus here on the initial process, matching the customer needs to the resources.

Given the resource network and the customer-model hierarchy, the matching process is based on mapping of the customer's model hierarchy onto the integrated resource network model. Since the mapping is part of the design process, the steps can be performed manually, by the database or view designer and the eventual owner of the mediator.

As the design information, i.e., the models and the ontologies that define the terms and their structures is being committed to machine-processable representations some automation can ensue. In Section 5.3 we present work on facilitators, which automate resource identification and adaptation. Major benefits of automation can accrue in long-term maintenance. The creation of new functionality in mediation is a creative effort, and is effectively performed by programmers or knowledge-engineers with high-level tools. Figure 4 presents the concept symbolically.

5.1 Mapping the hierarchy

Matching of a customer model to the resource model requires that the terminology used to describe objects and their attributes matches. Limiting mediation to closed-world domains is a precondition. Interoperation among domains is addressed in Section 8. The following steps extract the customer model out of the integrated model of available resources:

1 Locating the pivotal linkage: the root node of the customer's hierarchical model has to be found in the general resource network. If we assume that the root node identifies an object class, the search may be restricted to entities of the resource model. This assumption requires that the domain ontologies match, otherwise there is a chance that an object in one model has been characterized as an attribute in another model. For instance, a six-cylinder engine can be an attribute of car in a detailed automotive catalog resource, but may be an object in the garage model. When matching attributes are located, a transformation of the resource representation has to be initiated. These transforms have been considered when integrating databases, but not, as far as we know, applied dynamically. If any match of the root fails, objects lower in the customer hierarchy may be searched for, and the root becomes an object for aggregation only.

2 Now relationships in the resource must be matched to those emanating from the root of the customer object. There are again potential mismatches in semantics. Relationship representations may be value- or reference- based. The customer's model may be simpler than the resource model, requiring the match process to combine intermediate linkages. For instance, the tire of the customer's car can be linked to the wheel of the car, since there is a dependency if various wheels are being offered. But the customer's model may not have considered this choice.

* Now further entities and relationships can be matched until the leaf nodes of the customer's model are reached.

+ Entities and relationships beyond the customer's model can be collected, until the leaf nodes of the resource model are reached. These nodes may not be reported in the result, but contain data for aggregation into the customer's model.

Unmatched links and objects of the customer's model will be reported to the builder of the mediator, who must decide how to deal with the problem. If more resources are legitimately needed to satisfy the customer, then they must be obtained and the resource model correspondingly be augmented. In practice, resource cost or accessibility makes the matching of all of an ideal customer model pragmatically infeasible. At other times the customer's model is out of scope for the domain. In both cases the application has to be informed by the mediator of the mismatch of expectation to available resources.

5.2 Improving the Mapping

The process of matching, like any design process, may require some iterations. A customer or domain specialist may have ideas on substitute resources. When dealing with structured databases the breadth of the resources is typically well defined, since their schemas establish a closed world [Reiter:78]. To determine their scope, i.e., their coverage of object instances requires again analysis or expertise. For instance, nowhere was it documented that a certain Navy database about ships did not include fishing boats. In many contexts that may not matter, but when trying to stem smuggling it does.

When dealing with unstructured information sources, as browsing through bibliographies in the world-wide-web, no schema exists. No clear definition of scope is likely and resources must be selected based on content analysis [Gravano:94]. The most likely sources are selected, but now no guarantee can be given that relevant instances will be found at query time.

The augmented customer's model may also become too large. Potential resources may be identified which are actually redundant. The designer of the mediator may omit resources, or add conditions on their use, say, only when actual retrieval from a primary source is incomplete, will the secondary source be consulted [Agarwal:95]. Some level of detail may not be needed. Some summaries may exist in the resources and will not have to be recomputed. Exception search may be selected instead of summarization, as discussed in Section 6. It is here where much of the added value occurs.

5.3 Facilitation

Automation of mediating tasks is obviously desirable, even though lack of automation does not invalidate the concepts. Facilitation provides automation, primarily in resource identification and data format conversion [G:92, GSS:94].

The vision underlying facilitators is one in which any system (software or hardware) can interoperate with any other system, without the intervention of human cutomers or their programmers. This level of automation depends on having very thorough ontologies to describe the resources, since now no human is interposed between the customer's application and the resources. Human input will still define the ontologies of the resources, but the task of matching now occurs at query processing time.

By definition, facilitators are required to accept runtime submissions of

1 meta-data about information resources
2 logical statements relating disparate concepts (usually definitions)
3 format descriptions for the data items being accessed

Facilitators are expected to use this meta information in converting, translating, or routing of data and information.

For example, when a new resource becomes available, the facilitator will expect a machine-processable description of the new resource, which must be coherent with the existing description used by the facilitator. Then the facilitator can integrate the new resource, or elide an existing one, and the consumer can be immediately served with the new resource.

For the same situation the party responsible for a mediator, its owner, is to be informed of the new resource and its capabilities. The knowledge that defines the value-added function has to be augmented to include the new capability. The augmentation may be done by the owner, perhaps aided by an automated process if the semantic difference of current and new resource capabilities is modest. A new version of the mediator, with augmented capabilities is then made available. Existing applications will be informed, but the owner also has a responsibility to continue serving current applications, with the results that remain compatible.

For added value-services a facilitator may call upon a mediator as a resource. Then the required meta-information is provided by the owner of the mediator, moving the hard task of matching ontologies one layer away from the base data, potentially simplifying adoption of facilitators in the architecture. To achieve this level of interoperation, compatible interface languages must be employed [GK:94, Tourtier:94]. Both mediators and facilitators can use wrappers to access non-conforming resources.

Slide 11 goes here

Figure 5. Systems with Facilitators and Mediators

In summary, a facilitator can be viewed as an instant mediator to the extent that automation allows. A mediator, in turn, can be viewed as a petrified facilitator, a human mason rebuilds its role in a system architectural configuration.

6. Exploiting the Model

A major task for an effective mediation service is the reduction of data volume to be shipped to customer' applications, while maintaining its information content. More information embedded in less data increases the information density. Having a high information density deals with the complaint that customers voice now: that there is *information overload*. Reduction of transmission to the customer's workstation also reduces communication delays and costs. The principal tool for data reduction is abstraction, either by summarization or by exception seeking. Both functions depend on having a simple, hierarchical model of the customer's needs.

6.1 Summarization

Summarization provides aggregated data, classified along the dimension specified by the customer. Summarization is a common task now mainly performed by people who support high-level decision-makers. They will, of course, process the data with the help of computers, but will explicitly program their view of the customer's model. For example, cost data collected in a factory are at a level of detail which records the activity of every worker and every machine with respect to every task. For the payroll domain the worker's efforts are aggregated to daily hours, and then processed with data which determine overtime rates, and then further aggregated to weekly totals, which determine pay checks. At the pay level benefits are added, taxes are computed, and contributions are withheld.

The same source data from the factory floor can be aggregated according to a different customer model to arrive at costs per product, and to this aggregation the allocation of product development costs is added to arrive at the base costs which eventually can determine sales prices and profits.

Other examples are summarization of medical records, of sales trends by product-type, of population dynamics etc. The final presentation is often graphical, but aggregations are performed prior to the graphical presentation [deZeegher:88, MG:85].

6.2 Exception seeking

An alternate, even more effective abstraction is just seeking and reporting exceptions. Here only results that differ significantly from the customer's expectation are presented, for instance, abnormal clinical findings or an unexpected drop in sales for a product line. The need for a customer model is obvious here. A change in a patient's weight by 10% over a short time is typically a cause for concern; putting absolute limits on weights would lead to useless exceptions, even if patients were categorized by age, height, and gender.

Many business decisions are motivated by changes in customer demand. Simple tabulations do not tell the full story. Sale amounts are affected by exchange rates and promotions. Factory sales are buffered by inventories. Many products are affected by the weather and regional preferences. Only when these have been taken into account by specialists is it useful to use the information for production planning and investment decisions.

6.3 Caching

Not all resources maintain historical data. To enable summarization it may be necessary for the mediator to cache past data [Roussopoulos:91]. Understanding what the range and point of validity requires meta-data which is rarely available today. While processing routines for historical data are common, their invocation today requires human interaction [TCGJSS:93]. Dealing automatically with the heterogeneity of temporal representations requires adoption of standards and adequate algebras.

6.4 Guiding the processing flow

As a byproduct of matching the models, sequences of high-level operations can be generated, similar to automatic code generation during proof procedures [CG:91, Smith:93, Kieburtz:94]. The operations identified during model matching are collected for subsequent elaboration. The functional granularity and completeness will differ substantially from the codes created in automatic programming research. Rather than lines of code, model matching in will generate invocations to standard package interfaces. The inputs and results for these invocations are attached to the model.

Processing packages have to be linked to those invocations to perform the intended functions over the data. These packages will be selected by the designer of the mediator to carry out the function that the mediator is to perform, since only the framework of the information flow can be generated during model matching, and not the internal code within the functions. This level of added-value mediation cannot be created automatically until we have programming languages at a much higher level of abstraction. Abstraction for processing paradigms is much more elusive than abstraction over information models [WWC:92].

In any case, in mediation full automation is less crucial than having reliable guidelines for the structuring of the mediating program, since we do not see removing the programmer from the creation process, but rather providing access to the proper data and proper paths for the results.

7. Maintenance

Mediation adds value to the data by applying the 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 become a candidate for replacement by a competitor. Existing customers can continue to use the old mediator version, and not be disturbed until they decide that their application needs the upgrade. This flexibility is crucial, since now upgrades are not constrained by the effect on the existing community of customers. The maintainer will, of course, try to keep the number of versions of a mediator service modest. The charges for old mediators may increase, to encourage applications that depend on old versions to upgrade.

Automation is likely to play an important role in long-term maintenance. When changes are needed in the customer model or the resource model, the ability to re-execute the match process, while invoking the same processing functions that defined the role of a mediator can rapidly regenerate up-to-date services. Today, maintenance of software occupies the dominant fraction of programming resources. Without tools to provide some automation the situation is likely to be worse in large-scale heterogeneous information systems because

1 Information should be novel, , requiring constant source updates
2 Customer's legitimately change as they learn more about the domain they are exploring
3 Autonomous resources cannot afford to stay stable to make life easy for existing customers. They must advance in order to serve new customers with better, perhaps more detailed, data.

A small amount of formalization and automation of the process of programming the value-added services in mediators is likely to have a large leverage in keeping costs tolerable and benefits high. Support tools and modules are needed to build the required wrappers and transformations for mediators. Automatic generation of wrappers has been considered by [Franchetti:94]

Enabling the customer to alter the customer model as needed can keep the information system useful and viable. A prerequisite for customer interaction is a presentation of the current customer model. To indicate what changes can be accommodated requires browsing in the resource model and discovering what alternative attributes and categories are available

7.1 Facilitation versus Mediator Generation

Notice that tools to automatically generate mediators are not facilitators, because they still require human interaction and assumption of responsibility. Any such interaction also disables true dynamic interaction of consumers and resources. Version maintenance is also not an issue in facilitation. Programs capable of generating mediators, routers, translators, or wrappers need formal specifications. In some cases, these generators may work automatically, in some cases interactively with humans.

Slide 3 goes here

Figure 6. Resource, Mediators, and Applications in the Architecture

8. Domain Partitioning

We present mediation as the principal means to resolve problems of semantic interoperation. However, mediation will be needed in many topics, and we cannot expect that single mediator can cover all topics of interest to any application. We can expect even less that a single group of individuals can develop and maintain such a general mediator. We also expect that most applications need to combine more than one topic, and hence need support from multiple mediators, as indicated in Figure 6. Different applications will use different configuration of mediators. For instance, a production planner needs production cost estimates and product demand information. The sales manager needs the demand information, perhaps at a lower level of granularity, and inventory data.

We hence expect to have many mediators, each focusing on their own domain. A domain should be limited by its ability to maintain internal consistency. The consistency requirement means that the owners and maintainers have identical semantics for all terms. This requirement translates to a number of technical constraints. For instance, it is best if the terms used within a domain can formally enumerated, so that we can be specific about the scope of a domain.

The issues of domain specificity and interoperation have been addressed in [W:94 and W:95D]. These papers also propose a knowledge-based algebra to enable formalization and, in due time, optimization of interactions among domains. Domain technology inherits many of its concepts from database view models. Notice that views also help in making large resources understandable to the mediator maintainer and enable optimization of access.

Slide 7 goes here

Figure 7. Integration at Two Levels

9. Status

No system exists today that performs the set of tasks envisaged above. However, we do have partial examples. The PENGUIN system creates hierarchical objects classes from a structural database network model [Barsalou:91]. A prototype commercial implementation includes many of the needed functions [RDCLPS:94]. We cited examples of code generation earlier. If that technology can be applied to mediator generation, a considerable increase in scale and significance of that technology may ensue; several projects are now working on such a scale-up [Fisher:94].

Since we assume that most mediation services are performed in autonomous computing nodes, a requirement for interoperation is the ability to communicate according to some standard conventions [GK:94]. To enable a greater variety of communication actions, participants, and representations a *Knowledge Query and Manipulation Language* (KQML) has been developed [FFMM:94]. Automatic access procedures to heterogeneous resources are created in CARNOT, and the results are automatically joined at their *articulation* points, where labels match, although not reduced [Collet:91]. A hierarchical model of terms was used to automatically control the volume of topic citations to be returned in support of the tasks of an editor searching for referees [Chaudhuri:90].

Slide 13 goes here

Figure 8. Fat versus thin mediators

A number of applications have been developed using mediator technology [ISX:94]. Current applications have focused on manufacturing, where design and prototype production data have been combined [WCC:94]. This environment is being extended with services to handle engineering change management. A specific application has been the selection and validation of gimbals for antenna positioning on spacecraft at Lockheed Space Systems. The gathering and integration of intelligence data is an obvious application. Other areas being developed now are in healthcare management and plant safety and environmental cleanup. An interesting project is in the collection and integration of satellite data for land-use planning.

The implementation of mediators varies greatly. If knowledge-based processing is crucial, we may find mediators programmed in languages as LISP. If optimization is crucial to processing, the mediators may depend on packages written in FORTRAN. Maintenance would be enhanced by using declarative approaches that could be understood and modified by end-users. Many current mediators have been coded in the C and C++ languages.

A number of contractors now have the capability to build the required application interfaces and implement the architecture. The number of platforms and languages varies, and there is some discussion on **fat** versus `thin` mediators. Effective mediators must balance depth of function in terms of cost versus added value and breadth of function in terms of applicability versus maintainability.

To minimize the divergence of the mediator technology and its tools there ARPA has sponsored workshops to develop a shared reference architecture. The shared architecture document for the effort is currently being maintained at the University of Colorado [King:95].

10. Conclusion

Information technology is serving us well in specific domains, although we have remained dependent on specialist model designers and programmers for the implementation. Mediation is an architecture intended promote reuse and scalability, so that sources from many domains can contribute services and information to the end-user applications. The layered structure adopts for information structuring the domain management strategy used by the INTERNET distributed naming conventions [Kahn:87]. Both software and artificial intelligence technologies has been hard to scale when domains grew large or became diverse.

We presented a concept for generation of mediators which is based on the extraction of a hierarchical domain model out the general network. The result provides an object model for the using application, and hence a much simpler world-view then is represented by the underlying, heterogeneous resources. Mediators represent responsible, predictable and stable services. To warrant the use of mediators, there should be significant value-added processing. Data reduction, exception search, dealing with uncertainty among heterogeneous resources, and ranking of results are examples. The owner of the mediator assumes responsibility for the correctness of such processing. Facilitators provide automation and rapid response to changing situation. Automation depends on consistency in the ontologies which describe resource capabilities and customer needs.

A partitioning into domains creates a desirable autonomy to reduce the cost of maintenance. The focus on maintenance distinguishes the mediated approach from many other proposals, which attempt to design optimal systems. In large systems the major costs are in integration and maintenance, rather than in achieving initial functionality and optimality. Integration in mediation can proceed at multiple levels of abstraction, avoiding the centralization that hinders progress in data exploitation of data from diverse sources.

Mediators and facilitators inhabit the central layer of a three-layer architecture. Facilitators and mediators can interact with each other and resources within an information system.

From the primary distinction made in Section 5.3 we can derive the following strengths and weaknesses in the main areas addressed by the I3 program

function	Mediators	Facilitators
Control and Routing	determined by human design	dynamic and responsive
Content Manipulation	cricial to add value	per rule technology
Format Conversion	via wrapper invocation	dynamic, per descriptions
Run-time flexibility	no model changes	high
Efficiency	high with human involvement	requires quantitative meta-information
Mutual usage	facilitation in design phase	mediators for services

Table 2. Distinction of Mediators and Facilitators.

Tools are needed to support the development and maintenance of mediators. To have effective tools a common formal paradigm is needed. We cited some early tools used to support the initial mediation projects. The emerging standards have been made public on the networks, to avoid proprietary dominance. Rapid development is possible by exploiting the same modern networks that make the architecture itself feasible. With network access to standards and the potential consumers, whoever build the best tools or provides the best services will gain the deserved advantage.

Acknowledgment
This research direction would not have been feasible without the support of many colleagues in the ARPA Intelligent Integration of Information (I3) program, and, of course, from ARPA itself. Recent workshops sponsored by Bob Neches and David Gunning of ARPA, and chaired by Bill Mark of Lockheed, Tom Gruber of EITech have contributed valuable ideas. Work in progress on these topics can be located via the World-Wide Web node http://www-ksl.stanford.edu/knowledge-sharing. Our own papers can be found on http://www-db.stanford.edu and http://logic.stanford.edu/...; a dynamic glossary for terms used here is http://www-db.stanford.edu/pub/gio/1994/vocabulary.html.

References
Since much of this work is recent, the list provides handles to the World-Wide-Web, where much of the current information appears, especially when it deals with applying the network's capabilities to today's systems.

[Agarwal:95] Shailesh Agarwal, A.M. Keller, K. Saraswat, and G. Wiederhold: "Flexible Relation: An Approach for Integrating Data from Multiple, Possibly Inconsistent Databases"; Proc. IEEE Data Engineering Conference, Taipei, Taiwan, March 1995, pp.495–504.

[Barsalou:91] T. Barsalou, N. Siambela, A. Keller, and G. Wiederhold: "Updating Relational Databases through Object-Based Views"; *ACM SIGMOD Conf. on the Management of Data 91*, Boulder CO, May 1991, pp.248–257.

[BCGP:92] T. Berners-Lee, R. Cailliau, J. Groff, and B. Pollermann: "World-Wide Web: the Information Universe"; *Electronic Networking: Research, Applications, And Policy*, Vol.2 No.1, Spring 1992, pp.52–58.

[BLN:86] C. Batini, M. Lenzerini, and S.B. Navathe:"A Comparative Analysis of Methodologies for Data Base Schema Integration"; *ACM Computing Surveys*, Vol.18 No.4, Dec.1986, pp. 323-364.

[CG:91] D.E. Cooke and A. Gates: "On the Developmemnt of a Method to Synthesize Programs from Requirements Specifications"; Int. J. on Software Eng. and Knowledge Eng., Vol.1 No.1, March 1991, pp.21–38.

[Chaudhuri:90] Surajit Chaudhuri: "Generalization and a Framework for Query Modification"; *Proc. IEEE CS Intl. Conf. on Data Engineering 6*, Feb. 1990.

[Chu:92] W.W. Chu and Q. Chen, "Neighborhood and Associative Query Answering,"; *Journal of Intelligent Information System*, Vol.1 No.3/4, 1992, pp.355–382.

[Collet:91]C. Collet, M. Huhns, and W-M. Shen: "Resource Integration Using a Large Knowledge Base in CARNOT"; *IEEE Computer*, Vol.24 No.12, Dec.1991.

[Cox:94] Brad Cox: The Coalition On Electronic Markets; George Mason University, Fairfax VA, 1994, http://www.site.gmu.edu/˜bcox/ CEM/00CEM.html.

[CJSSW:94] A. Courtney, W. Janssen, D. Severson, M. Spreitzer and F. Wymore: *Inter-Language Unification* rel.1.5; Xerox PARC, ISTL-CSA-94-01-01, ftp://ftp.parc.xerox.com/pub/ilu/ilu.html, P94-58, May 1994.

[DD:93] C.J. Date and Hugh Darwen: *A Guide to the SQL Standard, 3rd ed*; Addison Wesley, June 1993.

[deZegher:88] Isabelle deZegher-Geets, Andrew G. Freeman, Michael G. Walker, Robert L. Blum, and Gio Wiederhold: "Summarization and Display of On-line Medical Records"; *M.D. Computing*, Vol.5 No.3, March 1988, pp.38–46.

[Dunstan:89] James E. Dunstan (ed.): *Knobots in the Real World*; Report of Workshop on the Protection of Intellectual Property Rights in a Digital Library System, Corp. for Nat. Res. Initiatives, Reston VA., May 1989.

[Elmagarmid:90] Ahmed Elmagarmid, Y.Leu,Y., W.Litwin, and M.Rusinkiewicz: "A Multidatabase Transaction Model for Interbase"; *Proceedings of the Conference on Very Large Databases* 16, Morgan Kaufman pubs. 1990.

[ElMasri:79] Ramez El-Masri and Gio Wiederhold: "Data Model Integration Using the Structural Model"; *Proceedings 1979 ACM SIGMOD Conference*, pp.191–202.

[FCGB:91] R. Fikes, M. Cutkosky, T.R. Gruber, and J.V. Baalen: *Knowledge Sharing Technology Project Overview*; Knowledge Systems Laboratory, KSL-91-71, Nov. 1991.

[FFMM:94] Tim Finin, Richard Fritzson, Don McKay, and Robin McEntire: "KQML as an Agent Communication Language"; in *The Proceedings of the Third International Conference on Information and Knowledge Management* (CIKM'94), http://www.cs.umbc.edu/kqml, ACM Press, Nov.1994. pp. 456-463,

[Franchitti:94] J.C. Franchitti, R. King, and O. Boucelma: "A Toolkit to Support Scalable Persistent Base Infrastructures"; in *Proceedings of the Sixth International Workshop on Persistent Object Systems*, Tarascon, France, September 1994, Springer-Verlag LNCS.

[G:92] Michael Genesereth: "An Agent-Based Framework for Software Interoperability"; in Software Agents, edited by Jeffrey M. Bradshaw, AAAI Press/MIT Press, 1997, pp 317-345.

[GK:94] Michael Genesereth and Steven Ketchpel: "Software Agents"; *Comm. ACM*, Vol.37 No.7, July 1994, pp.48–53,147.

[GSS:94] Michael R. Genesereth, Narinder P. Singh and Mustafa A. Syed: "A Distributed and Anonymous Knowledge Sharing Approach to Software Interoperation"; *Proc. Int.Symp. on Fifth Generation Comp Systems*, ICOT, Tokyo, Japan, Vol.W3, Dec.1994, pp.125–139.

[Gravano:94] Luis Gravano, Hector Garcia-Molina, and Anthony Tomasic: "The Effectiveness of GLOSS for the Text Database Discovery Problem"; *Proc. of the 1994 ACM SIGMOD Conference*, ACM Sigmod Record, Vol.23 No.2, May 1994, pp.126–137.

[Gruber:93] Thomas R. Gruber: "A Translation Approach to Portable Ontology Specifications"; *Knowledge Acquisition*, Vol.5 No.2, pp.199–220, 1993

[Fisher:94] David Fisher: *Adanced Technology Program, Component-Based Software* (94-06); National Institute of Standards and Technolgy, U.S. Dep.of Commerce, 1994.

[HBK:93] A.R. Hurson, M.W. Bright, S.H. Pakzad (ed.s): *Multi-database Systems: An Advanced Solution for Global Information Sharing*; IEEE Press, 1993.

[ISX:94] ISX Corporation: Intelligent Integration of Information (I3); http://isx.com/pub/I3, Westlake Village CA, 1994

[Kahn:87]Robert E. Kahn: "Networks for Advanced Computing"; *Scientific American*, Vol.257 No.5; Oct. 1987, pp.136-143.

[Kieburtz:94] Richard B. Kieburtz: "Generating Software from Specifications"; *Proceedings of the Monterey Workshop on Formal Methods*, Luqi (ed.), U.S. Naval Post Graduate School, Sept. 1994, pp.122–128.

[King:95] Roger King et al: *Reference Architecture, Intelligent Integration of Information Program*; University of Colorado, Sept.1995, http://www.cs.colorado.edu/ dbgroup/i3-ref-arch.html.

[Koster:94] Martijn Koster: *Rules for Web Robots*; Nexor Ltd., Nottingham UK, 1994, http://web.nexor.co.uk/users/mak/doc/robots/robots.html.

[MG:85] Jock Mackinlay and Michael Genesereth: "Intelligent Presentation: The Generation Problem for User Interfaces"; *Data- and Knowledge-Engineering*, Vol.1 No.1, Jun.1985, pp.17–29.

[NFFGPSS:93] R. Neches, R. Fikes, T. Finin, T.R. Gruber, R. Patil, T. Senator, and W.R. Swartout: "Enabling Technology for Knowledge Sharing"; *AI Magazine*, Vol.12 No.3, 1993, pp.37–56.

[OMG:91] Object Management Group: *The Common Object Request Broker: Architecture and Specification*; OMG Document 91.12.1, OMG and X/Open, distributed by QED-Wiley, Wellesley MA, 1991.

[PFPMFGR:92] R.S. Patil, R.E. Fikes, P.F. Patel-Schneider, D. Mckay, T. Finin, T.R. Gruber, and R. Neches: The DARPA Knowledge Sharing Effort: Progress Report; in Nebel, Rich and Swartout, eds., *Principles of Knowledge Representation and Reasoning*; pp.777–788, Morgan Kaufmann, 1992.

[RDCLPS:94] B. Reinwald, S. Dessloch, M. Carey, T. Lehman, H. Pirahesh and V. Srinivasan: "Making Real Data Persistent: Initial Experiences with SMRC"; *Proc. Int'l Workshop on Persistent Object Systems*, Tarascon, France, Sept.1994, pp.194–208.

[Reiter:78] R. Reiter: "On Closed World Data Bases"; in Gallaire and Minker (eds): *Logic and Data Bases*, Plenum Press, 1978.

[Rheingold:93] Howard Rheingold: *The Virtual Community: Homesteading on the Electronic Frontier*; Addison Wesley, 1993.

[Roussopoulos:91] Nick Roussopoulos: "An Incremental Access Method For Viewcache: Concept, Algorithm, and Cost Analysis"; *ACM Transactions on Database Systems*, Vol.16 No.3, Sep. 1991, pp.535–563.

[Salton:90] Gerard Salton: "Full Text Information Processing Using the Smart System"; *IEEE CS Database Engineering Bulletin*, Vol.13 No.1, March 1990.

[Smith:93] David R. Smith: "Constructing Specification Morphisms", *Journal of Symbolic Computation*, Special Issue on Automatic Programming, Vol.16 No.5-6, 1993, pp.571-606.

[STEP:92] International Organization for Standardization: *ISO 10303 Industrial Automation Systems and Integration — Product Representation and Exchange — Overview and Fundamental Principles*; Draft standard ISO, TC184/SC4, 1992.

[Tourtier:94] Paul-Andre Tourtier: "A Flexible, Facilitator-based Cooperation Framework"; Proc. Int.Symp. on Fifth Generation Comp Systems, ICOT, Tokyo, Japan, Vol.W3, Dec.1994, pp.101-110.

[TCGJSS:93] A.U. Tansel, J. Clifford, S. Gadia, S. Jajodia, A. Segiv, R. Snodgrass: *Temporal Databases, Theory, Design and Implementation*; Benjamin Cummins Publishing, 1993, pp.563-579.

[Ware:93] Willis H. Ware: *The New Faces of Privacy*; Rand Corporation, report P-7831, Santa Monica, 1993.

[WCC:94] Gio Wiederhold, Stephen Cross, Charles Channell: *Information Integration*; IEEE Educational Videotape, 2 hours, October 1994, Robert Kahrman, sponsor. IEEE, Picataway NJ.

[W:89] Gio Wiederhold: "Views, Objects, and Databases"; *IEEE Computer*; Vol.19 No.12, December 1989, pp.37-44.

[W:92C] Gio Wiederhold:"Mediators in the Architecture of Future Information Systems"; IEEE Computer, Vol.25 No.3, March 1992, pp.38-49.

[W:92I] Gio Wiederhold: "'The Roles of Artficial Intelligence In Information Systems"; *Journal of Intelligent Information Systems*, Vol.1 No.1, 1992, pp.35–56.

[W:93] Wiederhold, Gio: "Intelligent Integration of Information"; *ACM-SIGMOD 93*, Washington DC, May 1993, pp.434–437.

[W:94] Gio Wiederhold: "Interoperation, Mediation, and Ontologies"; *Proceedings International Symposium on Fith Generation Computer Systems (FGCS)*, Workshop on Heterogeneous Cooperative Knowledge-Bases, Vol.W3, pp.33–48, ICOT, Tokyo, Japan, Dec. 1994.

[W:95A] Gio Wiederhold: "Digital Libraries, Services, and Productivity"; Comm. of the ACM, April 1995.

[W:95D] Gio Wiederhold: "Value-added Mediation in Large-Scale Information Systems"; Proceedings of the IFIP DS-6 Conference, Atalanta, May 1995; to appear in Meersman(ed): 'Database Application Sematics', Chapman and Hall.

[WQ:94] Gio Wiederhold, and Xiaolei Qian: "Database Rngineering"; in J.J. Marciniak: *Encyclopedia of Software Engineering*, Vol.1, pp.269-282.

[WWC:92N] Gio Wiederhold, Peter Wegner, and Stefano Ceri: "Towards Megaprogramming"; *Comm. ACM*, Vol.35 No.11, Nov.1992, pp.89-99.