Information and technology are at a critical juncture. On the one side, government agencies, businesses, and individuals are committing unprecedented volumes of personal information to computer systems. On the other, networks are broadening interconnectivity with increased speed and bandwidth. This simultaneous growth in demand and capability is opening an opportunity to aggregate and integrate information from diverse, autonomous sources. However, any mechanism that collects and stores personal information from multiple sources must consider not only the technical challenges, but also the rights of the individuals and organizations providing the data.

To achieve this, data integration must look beyond static enterprise-based systems, which are modeled on current patterns in investigation and treatment, to a dynamic services-based solution. As the “Characteristics of Software Services” sidebar describes, in this model users choose the functions they want, which leads to the selection of appropriate services at the exact time they are required. For healthcare systems, this means that widely distributed and autonomous agencies, such as hospitals and social services, can deliver functions in response to a user query, while software services help the user identify which functions (and sources) are relevant to that query.

Indeed, with information being a primary resource in healthcare, the dynamic nature of such a model is appealing to an industry that is both heavily data dependent and in constant flux. Health science developments bring increasingly sophisticated diagnostic and therapeutic techniques and demand a greater understanding of individual citizens’ past and current health conditions. The variety of data can be overwhelming, encompassing individual personal medical history, healthcare delivery actions, and knowledge-based information about best practice alternatives. Adding to the problem is the lack of user sophistication. With technology development outpacing user knowledge, professionals trained primarily in clinical skills end up using distributed data capture and presentation systems that they know little about.

This confluence of distributed unfamiliar components and the challenge of navigating extremely large electronic files raises new risks that those who need the information most will not be aware of or will incorrectly interpret patient history. Such risks directly compromise the vision that health information technology should be quality enhancing.

Motivated by both the growing risk to the healthcare industry and the potential of the services model to address it, we instigated a research project to investigate how the services model could overcome some of the enterprise-based systems’ weaknesses. Delivering data through an information broker service, for example, appeared to have great promise. Using such an approach offers the opportunity of integrating data from a range of autonomous agencies, while at the same time preserving any restrictions on access to or use of the information that either agencies or legal and ethical frameworks might impose.

Our research led us to evolve the “data as a service” paradigm and a system for using a broker to deliver services. On the basis of that paradigm, we built a demonstration prototype, the Integration Broker for Hetero-
WHY AN INFORMATION BROKER?

The information broker is a relatively new idea in healthcare. Traditionally, data acquisition has centered on the electronic health record—an enterprise-based integrated file of the patient’s history that the practitioner or provider organization generates. The EHR stemmed from a 1990 study by the US Institute of Medicine that examined the idea of using electronic records as the appropriate modern medium to accommodate the rapidly growing need for concurrent multiple professional access to increasingly complex patient records. The report has since been updated, and most countries have either emulated or aspire to emulate the EHR model, which represents a patient’s lifelong health record.

Despite its proliferation, the EHR has no blueprint in terms of record size and transaction volumes, and there are even fewer studies of its effectiveness in light of rapidly developing medical science. This is a dangerous oversight, particularly since the precedents for large integrated enterprise systems from any sector are not encouraging.

An EHR, being enterprise-based, has inherent boundaries. This goes against the nature of many healthcare scenarios, which require information from a spectrum of potentially independent and autonomous sources. In some cases, such as cancer treatment, this can mean that a specialist must drill down into, or mine, data from a variety of primary and secondary care agencies.

Other applications might be concerned more with integrating bits of data from a range of agencies. Child protection is a good example: Usually, no one agency has a definitive picture of the threat to a child, but aggregating information from primary care, accident and emergency services, social workers, and others could provide a much clearer composite profile.

For real-time data—particularly important in child protection—an information broker has a distinct advantage over data-copying techniques, such as mining. All too often, retrospective studies of tragedies involving serious or fatal neglect show that individuals from various agencies had some information about apparently minor incidents, but saw it in isolation. Had data on all these individual concerns been brought together at a single point in time, the true picture would have rapidly emerged, and the child might have been protected.

For all these reasons, large-scale enterprise healthcare systems look increasingly inappropriate, and their continued use might result in serious problems. Ultimately, they are incompatible with the idea of integrating records across a patient’s life, including that person’s genetic Information Sources (IBHS). In feasibility tests, we used IBHS to aggregate information across six use scenarios developed with operations staff of the UK’s National Health Service. We successfully drew this information from sources distributed across three university sites that used a mix of database technologies.

Characteristics of Software Services

In one guise or another, software services are emerging as a potentially important paradigm for software development, although as with all emerging paradigms, a clear and widely accepted definition of concepts and terminology can be elusive.

One researcher has suggested that services are essentially an evolutionary step onward from the object paradigm, through which distributed objects can be easily reused. Others view services as being largely Web or business services. Indeed, service-oriented architecture has become as much of a buzz phrase as object-oriented architecture was in the 1980s—and is about as well defined.

The one thing that most of the software industry agree on is the set of characteristics that all services must have, regardless of the particular model:

- They must be used rather than owned, and such use generally does not require processing on the part of the user’s system.
- They must conform to a document-style interface, being independent of language or platform.
- They must be stateless and not preserve user-related knowledge across episodes of use.

A service must also provide a contractual interface. At a minimum, it must specify what is provided and be extensible to include any terms applying to its use.

From these core characteristics it is possible to adapt services to provide a particular model. For example, you can bound services statically (before execution time) if you know in advance which services are to be used. Statically bound services offer benefits over traditional programming forms, such as being platform independent. The most significant benefit of the service model, however, is the ability to loosely couple dynamic services, finding and binding these as needed.

References

The prototype was based on the system as empowering healthcare technology. This in turn will boost user confidence in and trust of thus make it easier to become familiar with the system.

will let users customize the way they query records and from external structures. Equally important, the broker development of new file structure and storage types for new petence in recording specialist data and permit the devel-

smaller departmental systems maintain their local com-

In contrast, adopting a broker approach will let smaller departmental systems maintain their local competence in recording specialist data and permit the development of new file structure and storage types for new diagnostic or therapeutic techniques without constraints from external structures. Equally important, the broker will let users customize the way they query records and thus make it easier to become familiar with the system. This in turn will boost user confidence in and trust of the system as empowering healthcare technology.

as a “trusted intermediary”—a phrase found in most definitions of information broker. Our vision was to provide a way for the user to assemble patient data on the fly, allowing the query’s nature and the user’s access rights to determine the usable sources. Thus, IBHIS should let clinicians, individual healthcare workers or teams, and even patients assemble needed information from sources such as hospitals, primary care practitioners, and social services. Obviously, a service model, in which information provision is the service offered, would be the only efficient way to assemble such varied information from so many diverse sources.

We began the project by working with National Health Service staff to identify information associated with six use scenarios that represent a range of data needs:

- child protection,
- disabled children with complex needs,
- mental health of looked-after children,
- single assessment of a person with multiple health and social care needs,
- intermediate care, which lets patients leave the hospital earlier and receive home care using multiple providers, and
- mental health of children and adults.

For our first prototype, which we used partly as an exploratory vehicle, we confined the use of the service model to IBHIS itself. Essentially, IBHIS was a fixed set of component services, which we deployed as we would components in a statically bound component-based software engineering solution. For the demonstration prototype, we moved to a more extensive use of the service paradigm, in which IBHIS selected data services dynamically as the original information request warranted.

The prototype was based on Java 2 Enterprise Edition (J2EE) Web services. We used IBM’s WebSphere environment both to develop and run the prototype along with three database technologies running on both Windows and Linux platforms. The three technologies were distributed across the institutions involved in the IBHIS project: Durham University, the University of Manchester, and Keele University. Users accessed IBHIS through a dynamic Web interface.

HOW IT WORKS

Figure 1 shows how IBHIS in its current form works as part of a larger broker system. The user interface checks user authorization and then lets the authorized user create a query through pull-down menus. To ensure that the query is semantically meaningful to IBHIS, the interface’s menu entries use a healthcare ontology—in this case, a subset of Snomed-CT (www.snomed.org/snomedct/index.html). The interface also presents the query results.

Figure 2 shows one of the steps in query formulation within the IBHIS prototype.

A 45-year-old patient, Joe, has a history of heart attacks and high blood pressure—conditions that medical personnel at his acute-care hospital have investigated and are currently controlling and monitoring. He also has a history of allergic reactions to penicillin, of which his local primary care physician is aware.

On a fishing trip to a remote mountainous area, Joe experiences serious chest pains. In struggling to walk back to the fishing lodge, he falls and lacerates his face. The force of the fall also causes alternating states of confusion and unconsciousness.

When passersby alert the rescue services, Joe is not coherent enough to name his primary healthcare practitioner. The paramedics and subsequently the local trauma team have no clues other than Joe’s name and where he lives. They therefore proceed with antibiotic treatment for the lacerations and must begin diagnosing the cause of the chest pains from first principles. Antibiotic treatment triggers Joe’s penicillin intolerance, and he continues to suffer chest pains while the local physicians keep trying to diagnose his cardiac condition.

Had an information broker been in place, as well as a modest indexing infrastructure, the clinical team would have been able to identify Joe’s primary care physician and then interrogate the record for major allergies and events. They also would have been able to identify and interrogate the acute hospital system that held Joe’s cardiology and hypertension history, current medication, and most recent investigations. Instead, Joe had to suffer treatment delays and a break in receiving vital medication.

Patient Mobility: A Strong Case for Healthcare Information Brokering

A 45-year-old patient, Joe, has a history of heart attacks and high blood pressure—conditions that medical personnel at his acute-care hospital have investigated and are currently controlling and monitoring. He also has a history of allergic reactions to penicillin, of which his local primary care physician is aware.

On a fishing trip to a remote mountainous area, Joe experiences serious chest pains. In struggling to walk back to the fishing lodge, he falls and lacerates his face. The force of the fall also causes alternating states of confusion and unconsciousness.

When passersby alert the rescue services, Joe is not coherent enough to name his primary healthcare practitioner. The paramedics and subsequently the local trauma team have no clues other than Joe’s name and where he lives. They therefore proceed with antibiotic treatment for the lacerations and must begin diagnosing the cause of the chest pains from first principles. Antibiotic treatment triggers Joe’s penicillin intolerance, and he continues to suffer chest pains while the local physicians keep trying to diagnose his cardiac condition.

Had an information broker been in place, as well as a modest indexing infrastructure, the clinical team would have been able to identify Joe’s primary care physician and then interrogate the record for major allergies and events. They also would have been able to identify and interrogate the acute hospital system that held Joe’s cardiology and hypertension history, current medication, and most recent investigations. Instead, Joe had to suffer treatment delays and a break in receiving vital medication.
Role of services

Internal services process the query and responses from the data sources. The query service and ontology service map a query into a form that matches the data structure and semantics of a particular data source, according to a data description in the Web Services Description Language. The security service then matches the query with the access permissions a particular source declares and reduces the query’s scope if necessary. IBHIS then sends the resulting set of subqueries to the selected data sources for processing. The data-access service (DAS) processes the query and performs additional access-control filtering to ensure that an unauthorized user does not gain access to patient records. The security service filters the integrated result a final time to reduce the risk that a user will use inference to acquire unauthorized information.

The DAS is perhaps the key service because it represents each data source as well as being the access medium for that source. As the “Inside the Data-Access Service” sidebar describes, the DAS’s role is to translate the query into a local format (such as the Structured Query Language for a relational database), obtain the relevant data, and return the data to the broker. Each DAS must record a service description with the semantic registry, detailing the DAS’s interface and access-control policy, and the available information’s structure and form. IBHIS then consults the semantic registry when determining where to route a query’s elements.

Access control

Obviously, the integration of specific patient data raises serious concerns about custodianship and ethical access control. Because one of our primary design goals is to let agencies retain their autonomy and data control and ownership, we assume that data providers can create their own access policies using their own terms, data syntax, and so on. IBHIS would then enforce that particular access-control policy.

We also wanted to address access entitlement. Users’ entitlement to access data could depend not only on their role when
Inside the Data-Access Service

As Figure A shows, the DAS acts as a transforming agent, linking the semantic forms of the query and response description and the data’s syntactic organization.

The semantic-description file provides the service description for the DAS, expressed in terms of a healthcare ontology. (The IBHIS prototype uses a subset of Snomed-CT.) The DAS engine processes the query, sends requests to the data stores, and collates and transforms the responses to create a query response.

The metadatabase holds the mappings between the domain ontology and the forms used in the local data sources. For example, the term “address” might map to many database fields or just to one or two, depending on the data store’s organization scheme. The mappings let the DAS translate requests expressed in terms of a world view, such as Address, into the provider’s local view, such as House Number, Street, Town, City, and PostCode. The mappings also let providers apply access rules to the request.

Overall, these elements create a data service that the information broker can use without needing any detailed knowledge of how the data is organized or managed. Indeed, within a health grid, other potential users could join the broker. The DAS also preserves the data owner’s autonomy: DAS creation involves only the service provider and the data source’s administrator to configure the semantic description file. As part of the configuration process, the data owner can specify exactly who has permission to see which data.

We identified three ways to realize this concept and used all three in our IBHIS demonstration prototype. The first is to have data owners host the DAS using the same system that holds the data. The second is to have the service reside on the broker’s own system, and the third is to have a dedicated front-end processor. We found that, as a rule, the third option is the most flexible and imposes the lowest overhead.

Data storage could also take many forms: one or more databases; sequential or indexed flat files; a set of files within a folder; or a real-time data stream from a monitoring device. What makes the DAS so powerful is that knowledge of this level of detail stays with the DAS, although some aspects might appear in the service’s published description, as appropriate.

Data as a service

Although the software services model in its most accepted form has many useful applications, it largely ignores specific data requirements, such as confidentiality. The data as a service model is based on work to develop a fine-grained model that incorporates an additional service-integration layer. This demand-led service model contrasts with more familiar models, such as .NET and J2EE, which are supply led and deliver shrink-wrapped functions. The additional service-integration layer lets the model support five major functions:
• **Description** encompasses characteristics such as functionality, interfaces, nonfunctional characteristics and constraints, as well as the parameters the service provider sets for service provision.

• **Discovery** supports the location of client services, typically by identifying a list of candidate services and providers.

• **Negotiation** is the dynamic interaction between the client and one or more providers to agree on the terms and conditions for supplying a service. Negotiation usually involves matters of contract and pricing.

• **Delivery** involves the invocation, provision, and suspension of the service itself.

• **Composition** is the process by which an end user or another service creates services that are not atomic from a set of lower-level services. It either uses history to select these or identifies suitable ones on demand.

From this demand-led model, we built the idea of data as a service and used it as the basis for IBHIS. The DAS provides access to data sources through software service technologies in a way that is independent of the platform and language used to implement the functions provided.

The functions in our data as a service model are much the same as the five functions just described, with only a few exceptions:

• Data description and discovery are essentially the same.

• Delivery is very similar in the data as a service model. Indeed, many examples of functional services, such as a spell checker, will return data. There is no notion of suspension, however.

• Negotiation is concerned primarily with access-control issues, rather than contract and pricing, although these might be elements in some circumstances.

• Composition is not as frequent, since data services are mostly atomic. However, a broker in a hospital could easily provide a composite service for other brokers.

As this list shows, except for negotiation, the data as a service model is close enough to the demand-led software service model to offer many of the latter’s strong advantages.

**THE DOMAIN CHALLENGE**

Healthcare is a key domain if for no other reason than its sheer volume. In most countries, it represents an average of approximately 10 percent of both the gross domestic product and employment. It has social and consequently political significance. Healthcare systems range from totally free market, through mandatory health insurance systems, to publicly funded systems. The health sector is under constant pressure to deliver more care, more quality, more integration, and more speed. The future pattern of EHR requirements is likely to change radically in the coming decades and will create a whole new set of challenges.11

**Health provider integration.** Many health record systems are office-based or department-based. Although individual specialties provide depth of expertise, divisions between them are a serious barrier to providing holistic integrated care based on the patient’s life experience. At some point, individual hospital-based systems could merge provider-based diagnostic and treatment records, but at present wider integration is proving elusive.

**Domain boundary expansion.** The healthcare domain has traditionally been tightly defined, involving medical, nursing, and related clinical professions and their immediate support staff. That is changing. The industry is recognizing that social work, sheltered housing, and other forms of support are vital to maintaining an individual’s well-being. Boundaries of the wider healthcare support team will continue to expand, and barriers must become selectively porous to support nonclinical personnel, such as staff in schools with special-needs children.

**Pervasive and ubiquitous technologies.** These are two aspects of the same phenomenon. Pervasive technologies mean embedding microsensors and microprocessors in clothing, for example, and using them to monitor the domestic living environment. This ready availability also makes the technologies ubiquitous. Consequently, sensors will be feeding a multiplicity of data on at-risk patients, such as those who are extremely elderly or handicapped.

**Genetics.** More professionals are recognizing the importance of genetic history in determining disease susceptibility. Genetic code itself is extremely data-intensive and thus space hungry, but specific requirements are changing as knowledge and clinical responses develop. At a minimum, the individual’s genetic profile will determine both screening for and treatment of diseases as the individual matures. Current record architectures and taxonomies could find it difficult to accommodate this data volume.

**Increased use of digital scans.** Diagnostic investigation is a critical part of healthcare assessment and treatment. An increasing range and proportion of diagnostic processes are now reported and recorded digitally. Moreover, not only are the most modern tests increasingly digitally demanding, but both clinicians and patients are requesting that they be undertaken
more frequently—both CT and PET scanning, for example, are moving from an exceptional to a routine repeat scenario.

**Personalized prescriptions.** Given increased awareness that individuals have different drug dose responses, personalized prescribing will become more common. Clinicians and pharmacists will need to master new clinical systems, linked to other clinical data.

**Increasing expectations.** Both health professionals and patients have increasing expectations that data will be instantly available at the point of need, which is normally a clinical consultation or a medical emergency. People are also living longer, with exponentially increasing needs for support and monitoring of multiple health conditions. At the same time, the increasing volume of sophisticated data increases the risk of data swamping. Hence, as data on demand emerges as a viable concept in healthcare, the need for privacy and confidentiality is giving rise to increased expectations about data storage and transmission security.

**MEETING THE CHALLENGE**

Our demonstration prototype of IBHIS revealed that it is possible to draw data from distributed sources that use a mix of database technologies. The question now is, How can we move from a demonstration prototype with limited application to a full-scale operational system?

The evidence suggests that such a transition would require little modification to the IBHIS’s core structures. The underpinning services model and technologies should be able to deliver a full-scale solution. However, we have identified three areas that present immediate obstacles.

The first is the user interface. Our existing interface is highly effective for demonstrating that IBHIS has enough functionality to cover a broad scope of possible users. However, healthcare roles and tasks have critical and specific requirements, which change rapidly. This argues for a range of purpose-specific interfaces, but such a concept comes with its own set of questions, such as how would a healthcare team work with the broker and cooperate in sharing information?

The semantic registry concept also needs rethinking. The ontology in the current prototype is only a partial representation of SNOMED-CT. The registry can handle matching using this partial ontology, but searching with a full ontology could incur significant overhead. The number of available DASs might also affect the registry searching time. Additional work should focus on minimizing this overhead.

Finally, using software services in conjunction with the healthcare grid requires the right environment. Future work should investigate how to establish the best environment for a data integration broker.

Clearly, a data integration broker such as IBHIS has the potential to overcome four significant challenges to EHR systems: It will be better able to accommodate exponential growth in record volume and complexity; adapt to new concepts, taxonomies, and related specialist record components; seek data from a variety of autonomous and distributed devices in a way that is user- not product-driven; and provide for flexible and extensible access-control models.

Information brokering is a powerful concept for the healthcare domain. It promises a more robust future solution to meeting the US Institute of Medicine’s original vision than either high-volume messaging or complex EHR systems that involve many enterprises.

Although important technical challenges remain in providing a full set of information broker services, the obstacles are not insurmountable. Our prototype has already demonstrated interim solutions, so in a sense, the revolutionary element is complete. What remains are evolutionary steps toward providing healthcare users with a much more information-rich environment than has yet been possible.

**Acknowledgments**

The IBHIS project was funded by the UK’s Engineering & Physical Sciences Research Council as part of its Distributed Information Management program. The project was collaborative, involving the Universities of Durham, Keele, and Manchester (the latter formerly the University of Manchester Institute of Science & Technology), and we thank our coworkers within the Pennine Group, which comprises software engineering researchers from these three universities.

**References**

IEEE Software Engineering Standards Support for the CMMI Project Planning Process Area

By Susan K. Land
Northrup Grumman

Software process definition, documentation, and improvement are integral parts of a software engineering organization. This ReadyNote gives engineers practical support for such work by analyzing the specific documentation requirements that support the CMMI Project Planning process area. $19

www.computer.org/ReadyNotes