# Chapter 16

# Toward Integrating Healthcare Data and Systems:

## A Study of Architectural Alternatives

**Timoteus B. Ziminski**
*University of Connecticut, USA*

**Eugene Sanzi**
*University of Connecticut, USA*

**Steven A. Demurjian**
*University of Connecticut, USA*

**Thomas Agresta**
*University of Connecticut Health Center, USA*

## ABSTRACT

*The adoption of health information systems and the integration of healthcare data and systems into efficient cross-institutional collaboration workflows of stakeholders (e.g., medical providers such as physicians, hospitals, clinics, labs, etc.) is a challenging problem for the healthcare domain. This chapter studies the way that well-established software engineering concepts and architectural styles can be employed to satisfy requirements of the healthcare domain and ease health information exchange (HIE) between stakeholders. Towards this goal, this chapter proposes a hybrid HIE architecture (HHIEA) that leverages the studied styles that include service-oriented architecture, grid computing, publish/subscribe paradigm, and data warehousing to allow the health information systems of stakeholders to be integrated to facilitate collaboration among medical providers. To demonstrate the feasibility and utility of the HHIEA, a realistic regional healthcare scenario is introduced that illustrates the interactions of stakeholders across an integrated collection of health information systems.*

## INTRODUCTION

The healthcare domain, frequently criticized for its antiquated handling of data (e.g., by using paper-based patient registries in physician practices), has been infused with a multitude of software solutions for Health Information Exchange (HIE) that focus on the integration of patient data from multiple sources in order to improve quality of care, lower healthcare costs, and support research. Some important driving factors are programs that provide funding such as the Meaningful Use EHR Incentive Program of

DOI: 10.4018/978-1-4666-9446-0.ch016

Medicare and Medicaid (CMS, 2013) or the Strategic Health IT Advanced Research Projects (SHARP, 2013) program. For example, SHARP has already spawned highly valuable platforms such as Informatics for Integrating Biology and the Bedside (i2b2, 2004) and the Substitutable Medical Apps & Reusable Technology platform (SMART, 2011). At the same time, patient involvement has been increased by initiatives such as the Blue Button (Blue Button, 2013), which allows patients simple access to their data or the data of a cared for elderly parent or child collected from participating medical providers (e.g., physicians, nurses, clinics, hospitals, image labs, pharmacies, therapists, etc.). In addition, the fitness market has exploded with a variety of fitness devices (wearable technologies) that link to mobile applications with new initiatives by the two dominant mobile computing players, Apple and Google. Apple has proposed a new HealthKit app (Apple Health App, 2015) for a dashboard to manage health and fitness data, while Google has announced its own Google Fit fitness tracker (Google Fit, 2015). Both companies are moving strongly into the smartwatch market to track motion, heart rate, blood pressure, activity, etc. In fact, Apple just announced ResearchKit (ResearchKit, 2015), an open source framework that allows researchers/developers to create apps in support of medical research; such a transformation will strongly rely on HIE in order to gather relevant data.

Despite all of the emphasis on HIE, there have been numerous problems that have been encountered during the same time span, particularly in regards to regional or statewide networks of connected healthcare stakeholders that practice HIE. Some very promising exchanges have failed (e.g., CalRHIO (Robinson, 2010) and CareSpark (Enrado, 2011)) and the progress of the work of regional health information organizations has been described as "discouraging" and "insufficient" (President's Council of Advisors on Science and Technology, 2010). However, even though the adoption of health information systems (HIS) by medical providers is starting to approach a wider acceptance in usage, the corresponding and required integration of healthcare data and systems via HIE remains a challenging problem, technologically as well as politically. On the side of technology, factors that limit adoption of HISs and HIE have been identified

(Gomes, Ziviani, Correa, Teixeira, & Moreira, 2012): a high development cost associated with HIE; a lack of agreed upon open-standardization particularly in regard to the sharing and exchange of data; a focus on brute force technology solutions rather than a healthcare process orientation that considers the needs of patients and providers and high data availability across HISs; and, the difficulty in maintaining HIE across multiple HISs that have the potential to evolve with new capabilities. On the political side, there is concern by major medical providers (e.g., hospitals in a particular region) that sharing data may lead to losing business (patients).

The healthcare domain has significant complexity and poses unique challenges that require novel approaches as well as leveraging existing solutions to have the potential to remedy the four aforementioned factors. This chapter studies established software engineering concepts, architectural alternatives, and best practices and investigates the way that they can be utilized in support of potential HIE solutions that integrate medical data and systems. We study architectural alternatives with a four step process by exploring the varied and complex requirements of the healthcare domain for supporting HIE of HISs, and matching these requirements to solutions from the software engineering domain. The first step provides an overview of software/system architectural alternatives that can be chosen for structuring an HIE system that integrates multiple HISs. The second step describes a detailed and realistic regional healthcare scenario with multiple entities that defines the scope of stakeholders that include a sole-provider practice, a community practice, local and regional hospitals, testing laboratories (blood, scanning, etc.), pharmacies, a university academic medical center, etc.; this scenario was done in collaboration with our co-author at a medical school. Within the scenario, the collaborative links between the involved entities are identified to serve as requirements of the domain. The third step proposes a hybrid HIE architecture (HHIEA) that leverages components from all of the identified alternatives and highlights strengths for particular use cases. Finally, the fourth step maps HHIEA into the assumed scenario including the way that it satisfies identified links to demonstrate the way that the architecture could be realized in a system.

This chapter is structured in 6 sections. In the *Background* section, context information is provided on varied HISs with a focus on the benefits of HIE and on the unique challenges that an HIE system architectures has to surmount. The *Architectural Alternatives* section explores high-level alternatives for system organization from the software engineering and architecture domains (federation, replication, and centralization) and a variety of instances of those alternatives such as service-oriented architecture (Rosen, 2008), grid computing (Foster, 2002), publish/subscribe paradigm (Eugster, Felber, Guerraoui, & Kermarrec, 2003), and data warehousing (Zeh, 2003). The *Regional HIE Scenario* section details a realistic regional HIE scenario with a selection of identified stakeholders, their capacities as medical data providers and consumers, and the collaborative links that exist between them. The *Hybrid HIE Architecture (HHIEA)* section proposes an approach that leverages the studied architectural styles to address both informational and functional requirements of HIE; in the process, the architecture is aligned to the realistic regional scenario. Then, the *Future Trends* section presents future efforts and directions in HIE that are emerging that may have an impact, including app-centric plugin architectures such as the aforementioned SMART platform, abstract architecture specifications for the construction of health applications such as the Open mHealth architecture (Open mHealth, 2011), and the integration of genetic analysis and results into Electronic Medical Records and Genomics (eMERGE, 2007). Finally, the *Conclusion* section draws this chapter to a close.

## BACKGROUND

This section provides background material that is required for the remainder of the chapter. To begin, health information exchange concepts are introduced and placed into context that includes their usage with other health information systems (HISs). Using this as a basis, the challenges facing health information exchanged are identified and briefly reviewed.

### The Role of HIE

Medical and health data obtained at the point of care presents both a challenge and an asset. Through transformation, aggregation, and analysis, medical/health data usage can range from billing and reimbursement through insurers to clinical decision support to automatically monitor patients and issue alerts to the foundation for research to establish new knowledge and improved procedures (Shortliffe & Cimino, 2006). To support all of these possibilities, HIE has emerged as one of the major means for electronic transfer of medical data among distinct healthcare organizations and their health information systems (HISs). The goal of HIE is to make medical and health data available to healthcare stakeholders (e.g., healthcare providers, researchers in academia and industry, insurers, patients, etc.) in an efficient, cost-reducing, timely, and safe manner (Kuperman, 2011). Monetary savings can be, for example, achieved by avoiding duplicated laboratory tests and the reduction of administrative overhead. HIE also has the potential to significantly improve the quality and safety of patient care by enabling healthcare providers to react faster to a patient's needs and avoid misdiagnosis, mistreatment, and adverse effects caused by incomplete knowledge about a patient's medical and health history.

Foundationally, HIE is based on the communication of a variety of complementary HISs employed by domain stakeholders. First, electronic health records (EHRs), which are a computerized record of a patient's health-related information (e.g., physician's observations, laboratory results, treatments, etc.) which can be created, managed, and consulted by authorized clinicians and staff within one or across multiple healthcare organizations. Their purpose can be manifold (support of research, education, etc.) but usually focus on supporting continuity of care. Second, electronic medical records (EMRs), offer similar functionality as EHRs, but are usually designed to function within one larger healthcare organization such as a hospital, clinic, etc. There are various EHRs and EMRs on the software market such as the publicly available Veterans Health Information Systems and Technology Architecture (VistA, 2003) developed by the US Department of Veterans Affairs, the open source OpenEMR (OpenEMR, 2012) and OpenMRS (OpenMRS, 2004) or a vast number of commercial products (ONC, 2015). Third, Personal health records (PHRs) are electronic records of health-related information on an individual that can be drawn from multiple sources while being managed, shared, and controlled by the individual patient or their representatives. Thus, they will frequently contain a special subset of the data available in EHRs, but can be also a rich source of information such as the nutritional supplement use of a patient. However, since their control lies with the patient, PHRs do not have the best reputation among physicians in terms of the reliability of the entered data. A prominent PHR is the Microsoft HealthVault (Microsoft HealthVault, 2007) platform with many insurance companies doing their own implementations.

Other HISs of note are: Medical Laboratory Information Systems (MLIS) for supporting the laboratory workflow from the test request to the specimen labeling to the creation of the lab report; Data Repositories or Data Warehouses for structured information storage and support of research surveys in academia and industry; and, Decision support systems (DSSs) for assistance with clinical decisions through evaluation of evidence-based knowledge in the context of patient specific data. Examples of functionality are drug interaction alerts or reminders for specific guideline-based interventions during healthcare (e.g., reminders for vaccine shots during a child's physical) and the care of patients with chronic disease (e.g., reminders to check blood pressure or glucose levels). Information can be presented by a DSS in a patient-centric view of individual care or in an aggregate view in order to support population-wide health management. Practice Management Systems (PMS) are for processing financial, demographic, and non-medical information about patients as well as scheduling, claim submission to payers, and other tasks; Electronic prescribing (e-prescribing) systems for reviewing drug and formulary coverage and transmitting prescriptions electronically to a local pharmacy. Systems can be integrated into clinical information systems to also screen drug interactions and allergies; Billing systems supporting payers such as insurance companies or employer to process claims and bill patients.

Healthcare is a fundamentally collaborative discipline in which patients, particularly with chronic medical conditions, are treated by a wide range of medical providers in both outpatient (medical office or clinic) and inpatient (hospital) settings. This involvement poses a unique set of problems for data exchange including: a large number of stakeholders, workflows that routinely cross institutional boarders, and a lack of data/information standardization. Consider a case in which a patient who suffered from a heart attack presents himself with flu symptoms at a physician's office. When the physician attempts to retrieve an overview of the patient's medical history, their HIS might simultaneously obtain records from the hospital's emergency room (ER), where the patient was provided with care, the hospital clinic, where he was an inpatient during his recovery, and from his insurance company, which covered the cost of his treatments. While the record from the ER might be coding the patient's condition as *Myocardial Infarction*, the inpatient record can use the nonstandard description *Heart Attack* in combination with a decimal code, and the billing information might refer to a proprietary coding system that identifies the condition through an alphanumeric string. The significant challenge for HIE is to recognize that those three records belong to the same single individual and describe the same event (through a mapping of terms and semantics) in order to ensure that the medical history retrieved by the physician treating the patient's current flu symptoms lists the heart attack as one single incident, despite the incoherence of the source data.

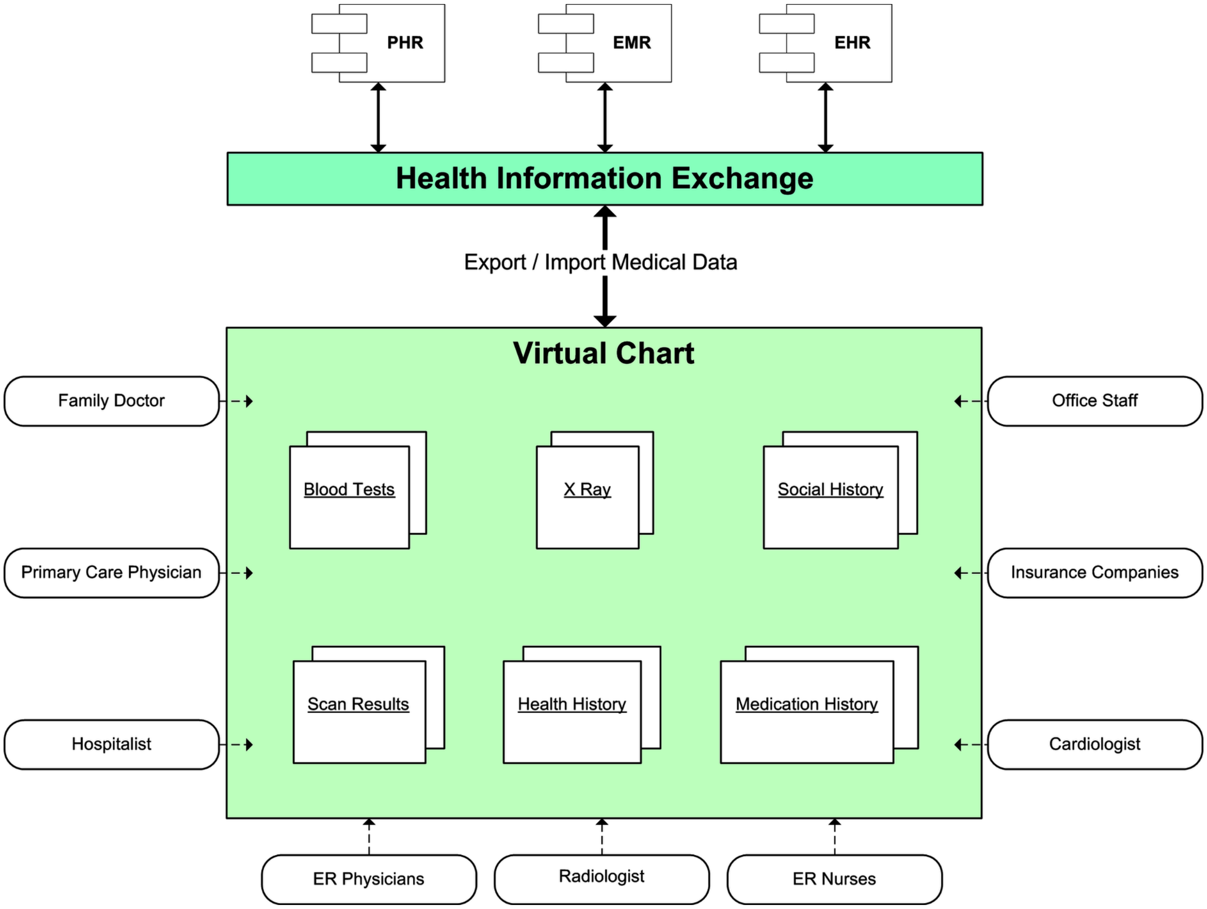
This implies that in order to ensure correct and safe data transport between stakeholders, an HIE solution first has to overcome technical communication barriers between the participating systems, which can be only achieved by the establishment of nationally recognized standards for the transfer of medical data from both syntactic and semantic perspectives. In addition, an integrated HIE must not only to enable the transport of the physical data, but also to ensure that the meaning of exchanged information is maintained throughout the communication process. This second step can be achieved through a uniform utilization of medical ontology frameworks and common term databases which can translate heterogeneous names, codes, and identifiers to one common meaning (Demurjian, Saripalle, & Berhe, 2009).

Medical facilities which currently use HISs and would be involved in the above example are often not sufficiently connected to one another due to lack of HIE standards, competitive roadblocks to sharing data across enterprises, associated costs for HIE without proven cost benefits, and so on. As a result, while data is instantly available inside the physician's office system, collaboration with external providers (and their systems) is haphazard and ad-hoc. When a patient visits the majority of providers today, a hard copy paper patient chart is maintained, containing the medical records for all of the patient's visits over time. This chart grows over time as record deletion is extremely rare (unless incorrect data or lab results have been inserted); the time-oriented content (from present to past) is vital to clinicians who must not only treat the current ailment but also look back in time to have a full history of the patient and their conditions and treatments. On the path towards fully adopting HISs in all phases of a patient's treatment, the need to have a fully integrated, electronic version of this medication history (current and past) will be critical. As a result, the creation and availability of a virtual chart (VC)

(Kenny, Parsons, Gratch, & Rizzo, 2008) has been proposed to provide a consistent, complete, and historically accurate patient medical record through automatic, HIE based integration of data from various source repositories. In Figure 1, the source repositories are a patient's PHR, a physician's EHR, and a hospital's EMR; other sources would include prescription records at pharmacies, billing records of insurers, imaging/scanning and laboratory test results, etc.

In summary, the envisioned virtual chart: provides an individual patient EHR that gathers data from multiple HISs; allows for retrieval of individual EHRs facilitating communication between providers; supports extraction of anonymized patient data from for aggregation to support data mining; has data monitoring capabilities for event tracking; implements security and access control enforcement and storage of audit trails that includes privacy rules for all parts of the stored data, data de-identification, etc.; issues alerts and preventive information that is in relation to a given chart instance; and, supports system personalization for adjusting to patient and provider needs.

*Figure 1. Healthcare Stakeholder Collaboration via HIE and Virtual Chart*



## Challenges for HIE Architectures

Software designers and engineers have worked on solutions for cross-institutional communication, (legacy-) system integration, and data exchange for decades and have created a large selection of solutions, templates, and alternatives that are routinely being leveraged by software architects in various fields on a daily basis. These well accepted methods meet a significant challenge in their usage for HIE due to the extremely complex and heterogeneous nature of the healthcare domain. In a typical scenario, each group of stakeholders utilizes different types of HISs; for each of these system types there are multiple products offered by different vendors and equipped with different communication mechanisms. Software architectures for HIE must support incremental development, since systems of this scale cannot be built in one effort. Similarly, HIE must support adaption to evolving requirements, since the list of requirements are not always fully known at the start of the system construction. Furthermore, scalability in throughput and size must be at the core of a suitable architecture, since processing will be increasingly challenging with HIE growth and adaption of new technologies involved in the HIE. Finally, the chosen architectures for HIE must support heterogeneous environments, since the current HIS landscape is highly fragmented into many different types of HISs which in turn are available from different vendors. This typically hinders interoperability between systems and domain stakeholders.

Two types of interoperability are required by HIE. *Syntactic interoperability* enables interaction on a technical level (i.e., the format of the interaction) while *semantic interoperability* enables interaction on the content level (i.e., the meaning of the interaction). For example, for connecting the HISs of a physician's office, hospital, and insurance company for simple data exchange, one might establish a message format that allows the exchange of events, containing a field "event name" (syntactic interoperability). However, the same event might be recorded differently, as a cardiac arrest, a myocardial infarction, and a billing code, respectively in those HISs. The goal of semantic interoperability is then to recognize all three records as one event. To support this, a medical ontology would know that a myocardial infarction is synonymous to heart attack. However, different HISs use different ontologies and may organize their ontologies differently, thus causing a need for a HIE architecture to be able to manage different ontologies and translate between them for semantic interoperability (Demurjian et al., 2009).

Integration of patient data requires an ID Management system with the ability to identify the patient across many disparate HISs. Since patients may see many providers and their information may be stored in different HISs, the HIE architecture must provide a method to uniquely identify a patient in order to retrieve their data form multiple systems. A lookup service allows an authorized HIS to query with a patient's demographic information to obtain a global identifier that other systems can recognize. Finally, HIE architectures also require a strong support for privacy and security, since they are tasked with processing highly sensitive data that is subject to regulations such as HIPAA (HIPAA, 1996) and FERPA (FERPA, 1974). The architecture should implement a single point of entry to handle security, utilizing fine grained role and permission management for medical documents (De La Rosa Algarín, Demurjian, Berhe, & Pavlich-Mariscal, 2012; De La Rosa Algarín, Ziminski, Demurjian, Kuykendall, & Rivera Sánchez, 2013).

## ARCHITECTURAL ALTERNATIVES

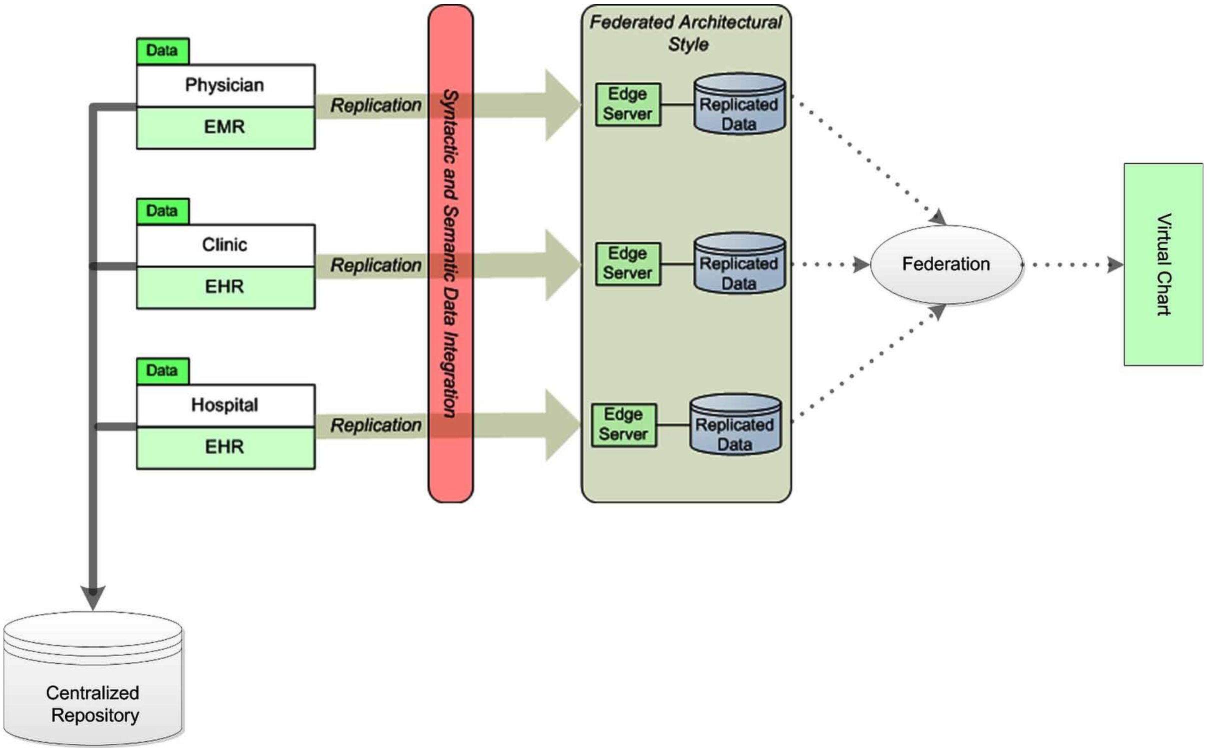
To set the context for the proposal of a hybrid HIE architecture, this section presents architectural styles that have the potential for usage in data integration and HIE. The first three, are collectively reviewed, namely: federation that directly utilizes in real time the data of HISs that is available for providing care; replication that off-loads data from HISs to dedicated edge servers that are then in turn federated; and, centralization that extracts data from multiple HISs into a single shared repository. Using this as a basis, a set of well-proven architectural styles and their suitability for HIE are analyzed, namely: service-oriented architecture (SOA) and grid computing, publish/subscribe architecture, data warehouse, and cloud computing.

### Federation vs. Replication vs. Centralization

In the federated architectural style, shown in the right side of Figure 2, medical data remains at its HIS source(s), and is made remotely accessible on demand. In this classic approach, a global database query is submitted, broken down into constituent local queries, processed at the remote sources, with results collected and formatted for presentation to the user. There are many advantages to this style. First, since medical data remains under the governance of the HIS source(s), the central portions of such architectures can remain relatively lightweight and process ad-hoc data queries on demand. As a result, federated solutions lower the amount of sensitive data that is shared and avoid making a priori assumptions about the value of data that an exchange participant may offer. Second, the data that is provided by a federated solution should be the most up-to-date version available since the federated user is accessing the federation in real-time. Third, the nature of federated solutions promotes scalability in terms of adding new HIS sources to an exchange network. There are also some significant disadvantages. First, the availability of data depends on the availability of the providing HIS source(s) and may therefore be unpredictable., e.g., patient records stored in a physician's office EHR may be unavailable after office hours. Second, performance bottlenecks are possible, e.g., the major hospital's EMR in a region may be accessed from numerous physician offices on a daily basis; if such access impacts performance to delay patient care, the result can be catastrophic. Third, security and monitoring of activities may be challenging in federated solutions, especially controlling and restricting access to sensitive data required distributed security models.

To overcome some of the fundamental issues of the federated approach, it can be extended by replication, as shown in the middle portion of Figure 2, which introduces an additional repository for each HIS source housed on an edge server. Since each repository is periodically updated, one advantage is that there is no impact on patient care on the HIS source. A second advantage is that the data integration to the edge servers gives each participant fine-grained control about which data is shared in an exchange, therefore preserving local governance choices. Third, the replicated repositories can be optimized to-

*Figure 2. The Federated, Replication, and Centralized Architectural Styles*



wards performance and availability, with minimal impact (typically only during offloading time) on HISs that are actively used for providing care. However, there are some disadvantages. First, the data that is offloaded from HIS source(s) may not always be up-to-date; periodic updates may be limited in order to minimize impact on the usage of the HIS source(s); for patient care, this is a problem since medical providers always require the latest patient data. Second, the information may be out of sync, e.g., the same patient that has data in two or more different HIS(s) may have an inconsistent view if the update periods differ per source. Third, security and monitoring remain just as challenging as with a federated approach.

Lastly, the centralized approach in the bottom left of Figure 2 provides a common location—a new central participant in the data exchange—to operate as a main, shared repository. Designing and implementing a centralized architecture requires the extraction and integration of existing data from the HIS source(s), either stepwise or in one major effort; this is true for both initialization and periodic updates to the repository's content. Through this integration process, a centralized approach makes it possible to attain syntactic and semantic interoperability, as discussed in the prior section. A first advantage is that one administrative governance can be utilized to control access to the shared information. Second, since the data in the central repository is available independent of HIS source(s), these source(s) are no longer impacted by external access. Third, contributors have significant control in terms of the patient that is to be shared and in what way (security). However, there are some disadvantages. First, there is a capacity factor to be considered since the data from potential hundreds of providers (e.g., hospitals, clinics, practices, labs, etc.) in a geographical region must be collected and combined; a country-wide integration would be even more difficult to achieve in practice. Second, the integration process from multiple HIS source(s) would require the need to reconcile all of the patient data to insure that the same John Smith's data has been collected from all sources without error. Third, from a system perspective, there is a probable performance bottleneck as HIS source(s) increase, the potential for fatal events if the repository goes down, and attacks or data theft that now impact a larger body of data.

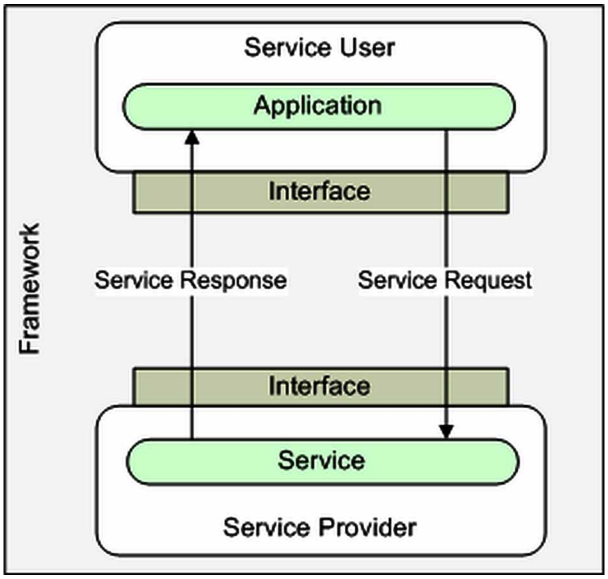
## Service-Oriented and Grid Architecture

The service-oriented architecture (SOA) as shown in Figure 3 is an architectural alternative which serves to construct extensible and inexpensive software support for business processes and workflows in order to supply a framework for administration and combination of software services which reflect processes. The services are loosely coupled, which means that they communicate over platform independent interfaces and assume their communication partners to be black boxes. Loose coupling also means that a component is self-sufficient except for its awareness and usage of the other components. Components offer their functions to other components in the form of services, which are similar in concept to publishing methods of an application programming interface (API). Services hide technical details (black box) and are defined functions which can be used on their own or as part of a larger task. The components connect via a mechanism which allows them to be aware of other components and their services while hiding the details of the component communication.

SOAs can be realized based on web services, which are defined by the World Wide Web Consortium (W3C) as "a software system designed to support interoperable machine-to-machine interaction over a network"

(Haas & Brown, 2004). Web services are commonly realized with HTTPS (Hypertext Transfer Protocol Secure – basic protocol for secured passing of web service data from machine to machine over the web), XML (Extensible Markup Language – a widely accepted standard for information exchange),

*Figure 3. Service Communication in a Service-Oriented Architecture*

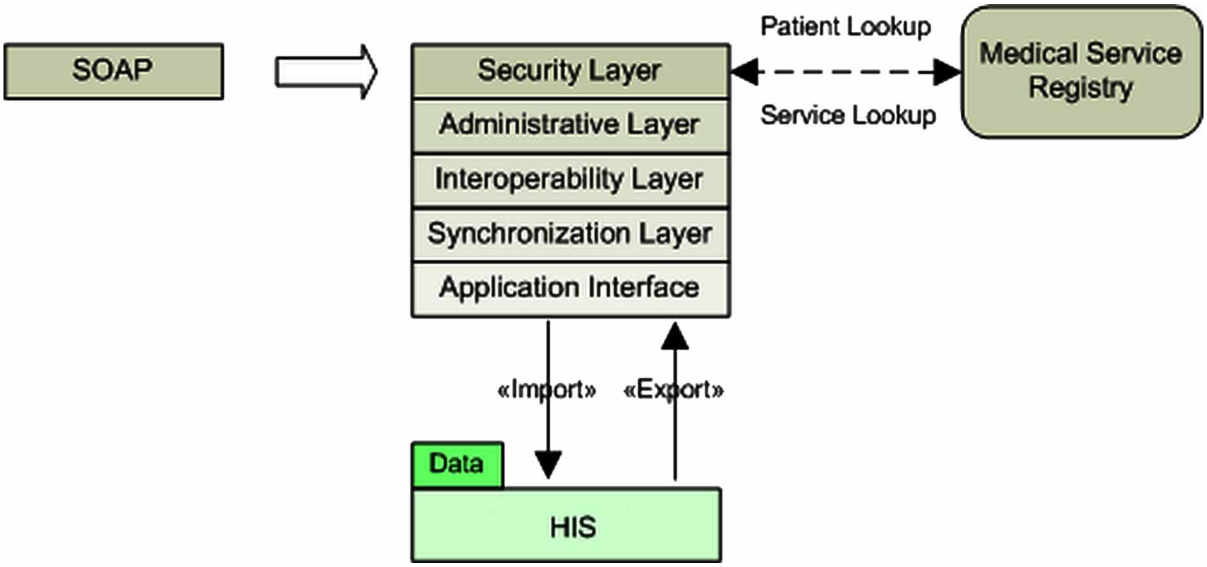


WSDL (Web Services Description Language – used to describe the functions of a web service in a machine-readable way, which allows the programmatic localization and utilization of a web service), SOAP (Simple Object Access Protocol – a lightweight communication protocol for message exchange and remote procedure calls which builds on HTTP and XML), and UDDI (Universal Description, Discovery and Integration registry – a registration directory for web service environments). Comprehensive introductions to all mentioned technologies are provided by the W3C (W3C, 2008).

SOAs manage the data of the participating components based on the federated/replicated approach. For the purpose of a healthcare SOA, this component needs to provide a set of lookup services: a *registry* for medical services, which stores references to the services that can be used through the SOA (may be based on WSDL and UDDI); a *patient identification mechanism*, which is effectively a *master patient index (MPI)* for identifying individual patients across the participating HIS sources; and, a *medical record lookup facility*, which stores of location of data, the patients index, and meta-data about the nature of the stored patient data (e.g., laboratory results, filled prescriptions, EMR, etc.) along with administrative data (e.g., creation date, last update, etc.). The meta-data will also contain information about the syntactic and semantic formats of the stored record; this allows accessing HIE participants to determine whether they can process the information automatically or not.

The lightweight central architecture means that each participating HIS source needs to implement an interface towards the SOA that is structured as depicted in Figure 4. The security component communicates with the central component of the SOA for authentication and authorization clearance. The administrative layer creates logs and audit trails for satisfying legal requirements. Transformations for semantic and syntactic interoperability are executed by the interoperability layer based on the meta-data provided with every medical record passed through the SOA. The synchronization layer is responsible for buffering or saving data for local use, signalling updates to other SOA participants, and reacting to data updates signalled from the SOA. Finally, the application interface interacts with the underlying HIS

*Figure 4. Logical Components Encapsulated in the HIE Interfaces of HIS sources*



source and triggers import and export of data as well as data processing. Clearly, there is a significant set of layers that must be provided for an HIS source to successfully utilize SOA.

To deal with performance issues in SOA, for complex science and research environments with high resource demands (e.g., CPU cycles, data storage, etc.), grid computing can be utilized. Grid computing describes sharing otherwise unused resources in a cluster (grid) of independent computing nodes. This happens in a transparent way, in which the additional resources from the created virtual super computer are available, just as electricity is in a power grid, to each of the connected nodes.

(Foster, 2002) defines the grid system based on following characteristics: “A Grid is a system that coordinates resources that are not subject to centralized control, using standard, open, general-purpose protocols and interfaces to deliver nontrivial qualities of service.” For the purpose of this chapter, grid computing is a realization of SOA with a selection of extremely fine-grained services. While the coarse-grained services of a SOA seem a more suitable solution for the workflow-oriented medical domain, there are emerging initiatives to utilize grid computing for healthcare improvement, medical research, and collaborative care (caBIG, 2004; World Community Grid, 2004), including: medical image processing and analysis, pharmaceutic research/development tasks, complex modelling and visualization jobs, and genomic applications. As personal genomics moves into the forward with genetic information linked to medical patient data, grid computing may be necessary to handle the potential higher volume of data.

## Publish/Subscribe Architecture

Publish/subscribe is an asynchronous messaging paradigm describing the relationship of senders (publishers) and receivers (subscribers) with: *publisher* which sends out categorized messages containing relevant data and do not contain a specific target address; *subscriber* which allows a subscription to one or several feeds that cover message classes without having knowledge on publishers that process all received messages according to their needs; and, *broker* which are optional components for mediating between publishers and subscribers by encapsulating the classification of messages, the subscription process, and the message dispatch. In an HIE system that connects multiple HIS source(s), a publish/subscribe architecture could be used for: the exchange of arbitrary medical data between the stakeholders of the domain, health status and advisory notifications such as epidemic alerts, and feedback mechanisms such as drug reaction reporting.

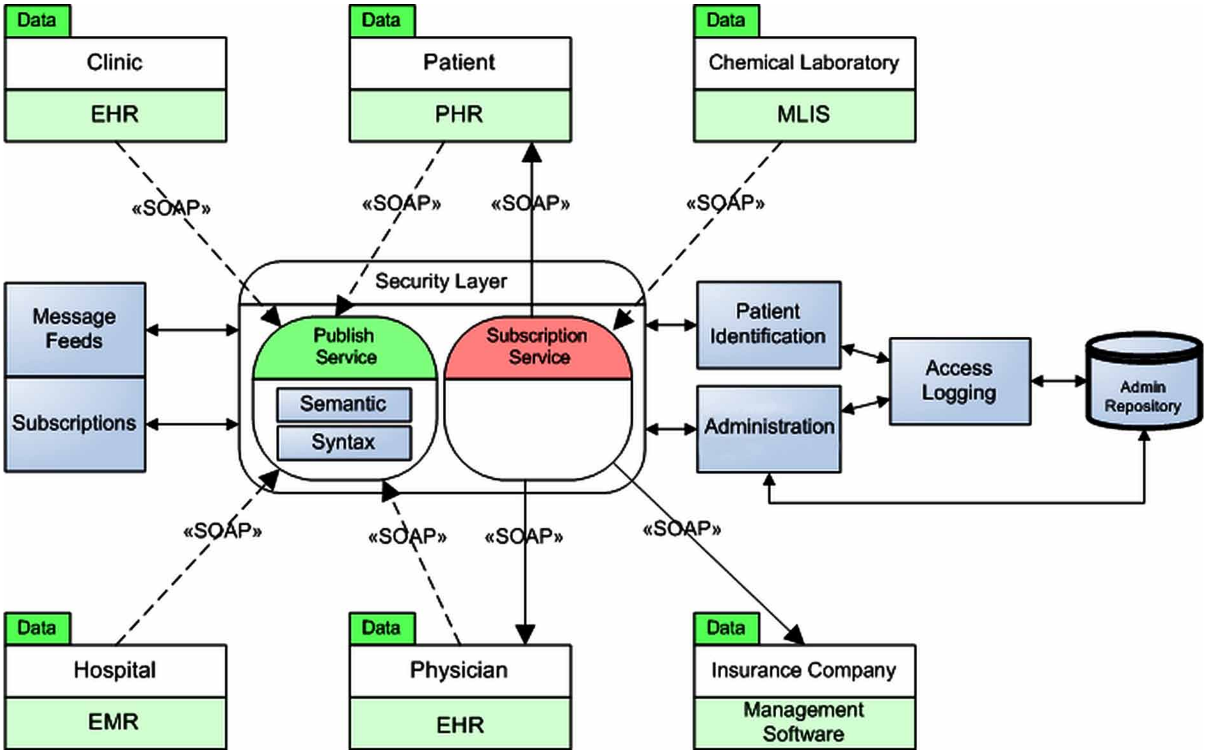
The publisher/subscriber architecture could be applied to healthcare to provide data from multiple HIS source(s), as shown in Figure 5, by implementing the federated data storage approach without central data storage using an *enterprise service bus (ESB)* for the broker. ESB allows message-based communication and integration between components connected to the bus. This is achieved through a set of services provided by the bus, such as routing services to pass messages from a sender to one or several receivers, and transformation services which can encapsulate syntactic and semantic transformations that are needed in healthcare. Using an ESB for healthcare allows for a higher level of central governance and interoperability control, since all of the communication happens over one component, while keeping the central architecture relatively small. Several ESB products, mostly based on sets of various middleware components, are available from different vendors, in both open and closed source commercial distributions (OpenESB, 2012; WebSphere ESB, 2008; BizTalk Server, 2006; Oracle ESB, 2012; Apache ESB, 2008). Publish/subscribe for healthcare, as depicted in Figure 5, contains the following components: *Patient Identification* implements an MPI in order to identify patients across HIS source(s) by creating an index for a given patient with each source responsible for its correct usage; *Administration* and *Access Logging* contain the addresses of the registered HIS sources, their meta-data, and access rights information to log which HIE participant received which messages in the publishing process and stores records to meet legal audit requirements; *Message Feed Administration* and *Subscription Administration* that contain a list of all message feeds and their current subscribers; *Publish Service* that is the communication interface for publishers, contains syntactic and semantic transformation services for a central interoperability control, and assures that messages distributed over the ESB are readable by all related subscribers; and, *Subscription Service* that is the communication interface for subscribers, which notifies all of the subscribers of a message feed on the arrival of new related messages and maintains the messages until they are delivered to the subscribers.

In summary, the publish/subscribe architecture implements the federated approach to data storage (the broker does not persist any data after a successful dispatch), and the architecture shares the basic advantages and disadvantages as in the case of SOA. Two potential problems are the unnecessary data replication across the system and the balance which the HIS source(s) must keep between (a) signing up for too many feeds (risk of information overflow) and (b) signing up for too few (loss of relevant information). Too high a number of feeds may slow down the receiving system and have negative impact on patient care; too few feeds may mean a healthcare provider has incomplete data and cannot make a diagnosis or a researcher arrives at an invalid conclusion in a study. The central broker component of the publish/subscribe architecture can remain relatively lightweight and inexpensive, yet still incorporate a good handling of logging, access rights management, and interoperability control.

## Data Warehouse

The data warehouse collects data from multiple sources to provide a uniform view on data for querying, analysis, and decision making tasks. Data warehouses (Inmon, 2005) are data collections with the following key characteristics: *subject-oriented* which describes the way that data for the warehouse is

*Figure 5. Publish/Subscribe HIE*



chosen, where a possible subject for the healthcare domain would be the patient or the physician; *integrated* which refers to the common schema in which the data (extracted from heterogeneous sources) is stored; *time-variant* which refers to the long-term storage of data, allowing analysis related to time; and, finally, *non-volatile* which means that once stored, data remains in the data warehouse (i.e., there are no delete or overwrite operations). For example, in healthcare, a data warehouse for emergency room patient data from hospitals throughout a country could be utilized to identify and track diseases, epidemics, etc.

Establishing a data warehouse for healthcare data includes two main tasks: extract and integrate data from multiple HIS source(s); and, make the integrated repository available to all eligible HIE participants via a query interface. Data extraction from HIS source(s) occurs periodically through scheduled pull operations (ideally in the after office hours in order to minimize impact on performance of systems used for providing care) or as push operations initiated by the sources (e.g., when low system load is detected). The extraction and integration process is complex and requires the following subtasks

(Inmon, 2005): *converting* the data into a common format (e.g., HL7 CDA (HL7, 2007) or CCR (CCR, 2012) in healthcare) with syntactical and semantic checks as well as by transformations as needed to match the utilized storage formats; *cleaning* the data of irregularities such as data entry errors, e.g., heights in meters instead of feet and inches; *integration* of the different data sets to suit the data model of the data warehouse, e.g., clearly identify the same John Smith in all sources; and, *transformation* of the data through summarizing and creating new attributes, e.g., aggregating certain medical data that may make an automated clinical decision such as finding multiple high blood pressures for the same patient over time with no appropriate medication prescribed. Note that these four steps are often a semiautomated processes requiring significant human interaction and intervention which is a negative for usage with healthcare. The healthcare data in a warehouse must separate identity with encrypted storage of patient identifiers for safe retrieval of anonymous data and identification of individuals for authorized entities, e.g., contact of suitable patient cohorts for research studies.

The data warehouse has advantages that include room for optimization, acceptable and predictable performance, and administration of security and interoperability matters under one governance structure. However, the actuality of a real-time data warehouse might be not realizable in a satisfyingly performant way for a regional or country-wide HIE. For example, in healthcare, there is a need to have high availability of data as it impacts patient care, which is not as critical for an e-commerce application. This will necessitate both frequent uploads and synchronizations which may impact HIS source(s) performance; if not, the data would be worthless to the treating medical provider. Further, a country-wide data warehouse may simply not be feasible with the large scale of systems to gather data from and the resulting volume of data. The usage of data warehouses for healthcare may be limited to non-patient care situations but might be possible to construct very large scale warehouses for offline medical and healthcare data analysis.

## Cloud Computing

The formal definition of cloud computing (Mell & Grance, 2011) sums up this paradigm: “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”. Cloud computing is an architectural concept which allows the usage of abstract resources that hide computational details and complexity from their users, much like SOA. A key feature of cloud resources is their rapid elasticity (Armbrust et al., 2009) that allows the company to grow and shrink capacities according to requirements that potentially change in real time, catering towards application fields like e-commerce with very high punctual peaks in infrastructural requirements (e.g., sales on Black Friday or Cyber Monday). Cloud computing typically provides “on-demand self-service” (Armbrust et al., 2009), which means that adjustments to the service (e.g., registration, configuration, extension) are mostly automatic and can be executed programmatically. Service will usually depend on broad network access through standard mechanisms (e.g., web service technologies) which requires a constantly available network connection with sufficient quality of service. Resource pooling and virtualization plays an important role in the background of a typical cloud computing service. However, the user deploying a service to the cloud or using it through a thing client (e.g., a smartphone) has, in most cases, no control of any physical resource instances and is protected from the complexity of that the process.

Cloud-computing solutions for enabling health information collection, exchange, and collaboration are appearing in the marketplace, e.g., Microsoft Cloud Services for Health (Microsoft Cloud Services for Health, 2015), the IBM Collaborative Care project (Andrews and Mack, 2011), Logiworks Healthcare Solutions (Logicworks, 2015), VMware vCloud (VMWare vCloud, 2015), etc. Since cloud computing platforms typically foster modular approaches, this facilitates configurable plug-in architectures with support of various extensions for healthcare tasks: decision support, note processing, voice recognition, and patient education. Furthermore, the versatility of the deployment model is capable of streamlining the realization of HIE and the integration of multiple HIS sources by installing regional clouds and enabling communication through hybrid clouds is a big plus. For example, cloud EHRs such as Practice Fusion

(Practice Fusion EHR, 2015) could be used to efficiently implement the application service provider model as described by

(Wager, Lee, & Glaser, 2009), which is important in the context of implementing EHRs in small practices that do not have the financial means to hire technical know-how for software and hardware maintenance. A third advantage is, building the virtual patient chart, as introduced earlier in this chapter, could largely benefit from a platform with some of the characteristics provided by the cloud services. There are some disadvantages. First, while HIS sources may implement federation/centralization on top of the cloud paradigm, their foundation is an abstract system under the governance of the cloud provider. This has two results: negative implications for regulatory issues since cloud providers do typically not specify where data centers are located, so that patient data may be stored outside of the patient's country and become subject to foreign laws; and, uncertainty for security particularly if sensitive data become accessible outside of an HIS source through shared hardware and faulty instance encapsulation.

## A REGIONAL HIE SCENARIO

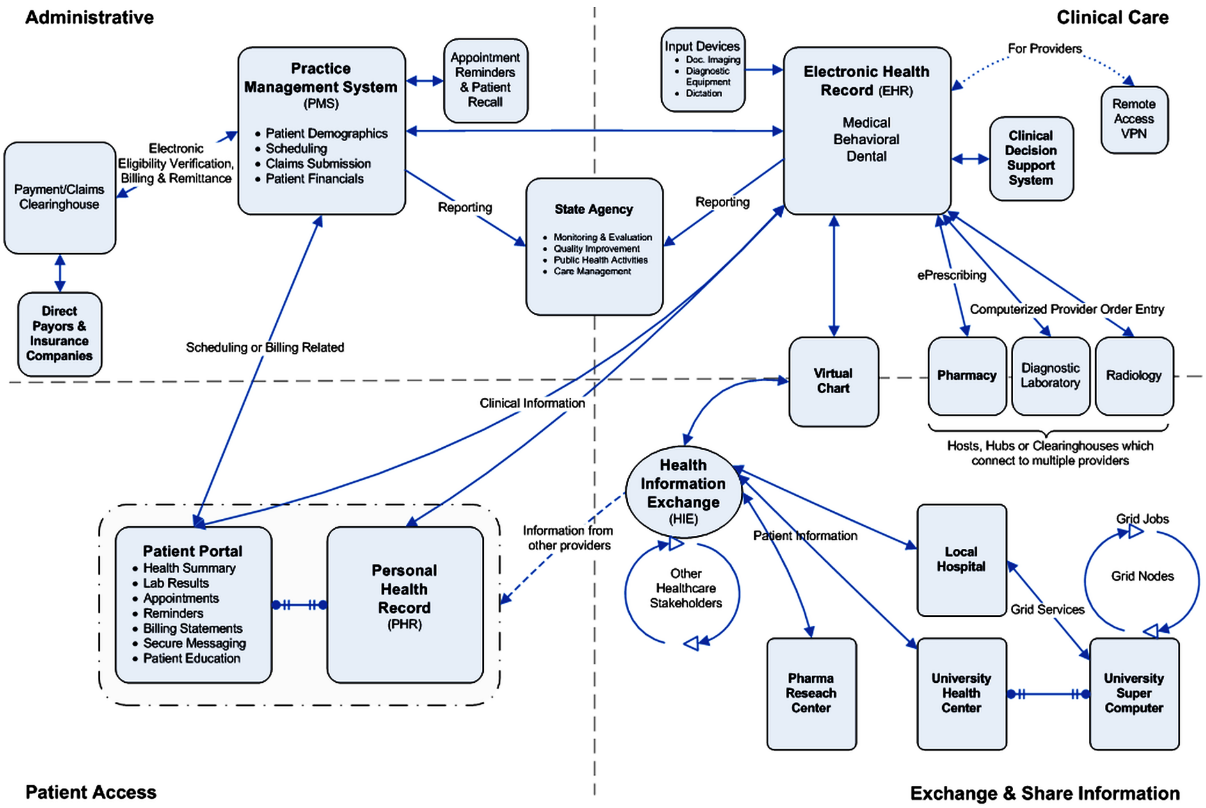
To define the canvas against which the features and components of the proposed hybrid HIE architecture to be explained in the latter portions of this chapter, this section introduces a set of HIS sources that participate in a regional HIE scenario, lists the employed HISs, and identifies selected examples for collaboration between sources. For the purpose of this discussion, a source is classified either as a data supplier or as a data consumer. Note that in order to maintain clarity, a simplified model is used, with a limited amount of sources and under the assumption that each source fits exclusively into its category. In an actual HIE scenario, suppliers naturally will consume data at the same time, and vice versa. In order to place the data suppliers and consumers into a concrete context, Figure 6 illustrates a real-life regional HIS infrastructure. The example is centered on the HISs of a physician's practice (e.g., a community practice) and explores the support for administrative tasks, clinical care, patient access, and the exchange of information in the process of providing patient care.

The core of the HIE infrastructure showcases the use of a selection of HIS sources as discussed earlier in this chapter: Administration of the practice, to manage patient demographics, appointments, billing tasks, etc. are supported by the patient management system (PMS), which interfaces with insurance companies and other payers. Clinical care is provided with the support of an EHR system in combination with a decision support system and various input devices; PMS and EMR automatically report to healthcare-related agencies. The EHR supports an e-prescribing feature as well as the electronic communication with various medical laboratories. Patients are able to access information from the practice's systems via a web-based patient portal and the PHR. The EHR system accesses information from other external providers via HIE to access a hospitals EMR or a laboratory system, and allows reviewing a patient's medical history via a virtual chart. Note that the parenthetical notation in the remainder of this section is referring to the location of each HIS source in Figure 6.

To begin, the data suppliers that are shown in Figure 6 are reviewed.

- **Community Practice (Upper Half):** A medical practice operated by several physicians (e.g., a general practitioner, a pediatrician, an internist, and a radiologist) and their staff. The HISs used in the practice are a decision support system (DSS), a practice management system, a web-based

*Figure 6. Overview of a Health Information System Infrastructure*



patient portal and an EHR. The practice collaborates with the virtual chart, pharmacies (e-prescriptions), and state agencies (reporting of selected diseases mandated by law).

- **Local Hospital (Lower Right):** A hospital providing healthcare to inpatients and outpatients from the local population. The hospital staff (i.e., nurses, physicians, ER personal, etc.) files patient data with an EMR and uses a DSS, where required by patient care. The hospital's radiology department participates in a large-scale breast cancer screening program and stores the resulting images in a custom database (the beginning of a data warehouse) for automatic analysis. The hospital collaborates with a university supercomputer center (automated x-ray analysis), the virtual chart, and state agencies.
- **University Health Center (Lower Right):** A research focused healthcare facility with a limited patient cohort (two hundred inpatients), maintaining an infrastructure for clinical studies, and running a human tissue and specimen bank. A medical chemistry laboratory equipped with a medical laboratory information system (MLIS) is also part of the center. The health center staff (i.e., physicians, clinical researchers, nurses, etc.) utilizes an EHR system and custom databases for the management of data related to inpatients and study participants. The EHR system supports automatic drug tolerance reporting to the pharmaceutical research center of a sponsor. The center collaborates with a university supercomputer center (for genomic research), state agencies, and a pharmaceutical research center.

- **Personal Health Record (Lower Left):** A web-based application used by patients to maintain their medical and health histories, including conditions, received treatments, current and past medications, allergies, food supplements, health markers, fitness data, etc. Significant parts of the stored data originate from patient entries (manual or recorded through smartphone apps or other consumer-grade tracking devices such as wearables). The application also supports a calendar function for the scheduling and planning of practice visits. The record is connected to the virtual chart.

To complete the discussion, the data consumers that are shown in Figure 6 are reviewed.

- **Local Pharmacies (Middle Right):** A group of pharmacies offering a variety of over-the-counter and prescription drugs. The main HIS used is a pharmacy management system (management of prescription histories, e-prescribing interface, business processes). Each pharmacy is connected to the community practice (e-prescriptions), the local hospital, and the university health center.
- **State Agency (Upper Middle):** Institution such as the Department of Mental Health and Addiction Services, the Department of Children and Families, or the Centers for Disease Control and Prevention. The agencies provide various healthcare-related services to the population such as issuing reports on health-trends, monitoring of the population-wide health status, executions of screening programs (e.g., alcohol abuse), and tracking of health-related events (e.g., in the case of a pandemic outbreak). The agency collects data generated by the community practice, the local hospital, and the university health center.
- **Insurance Company (Upper Left):** A company providing a variety of health plans for public agencies, and private and corporate customers; handles eligibility verification and billing for the local hospital and the community practice.
- **Pharmaceutical Research Center (Lower Right):** A research facility operated by a pharmaceutical company. The facility collaborates with the university health center (collection of anonymized data for research purposes) as well as with the local hospital and the community practice (recruit patients and retrieval of results from pharmaceutical studies).
- **University Supercomputer Center (Lower Right):** A virtual supercomputer operated by the university's center for transitional science. The system is based on a grid architecture, connecting grid nodes at the university and nationwide.
- **Virtual Chart (Middle Right):** An HIE application providing consistent, complete, and historically accurate patient medical records through automatic integration of data from various source repositories. The chart is linked to all HIE participants that process medical histories (such as the local hospital or the community practice).

## A HYBRID HIE ARCHITECTURE (HHIEA)

The architectural alternatives studied in this chapter have many disadvantages that make their usage unsuitable for an HIE solution using only one of the alternatives. This section proposes a hybrid HIE architecture (HHIEA) at a design level, by proposing a combination of the studied architectural styles which both balances and mitigates the advantages and disadvantages. The resulting HHIEA establishes a more effective and flexible architecture that facilitates the exchange of information between multiple

HIS sources and provides a comprehensive and integrated view on healthcare data. HHIEA combines: data warehouse (Prokosch & Ganslandt, 2009; Bauer & Günzel, 2013), SOA (Rosen, 2008; Ryan & Eklund, 2008), grid computing (Dickmann et al., 2012; Foster, 2002), and the publish/subscribe paradigm (Eugster et al., 2003; Singh, Vargas, Bacon, & Moody, 2008). The proposed HHIEA, as shown in Figure 7, is presented in five logical groups in separately labeled subsections, namely: the *Data Layer* where all of the data suppliers and data consumers are logically part of this group; *ID Management* that is utilized to both identify and differentiate patients and organizations participating in the HIE; *HIE Management* that is composed of tasks related to the maintenance of medical record references; *Security* that contains components that are related to the management of audit trails, patient consents, and participant authentication; and, a *Health Service Bus* that is responsible for ensuring the precise passing messages between the HIS sources. The discussion on HHIEA is completed with a section that utilizes the realistic healthcare scenario in order to fully demonstrate the ability of HHIEA to attain and support HIE.

## Data Layer

The *data layer* of the proposed HHIEA, as shown in the middle of Figure 7, is the location where information that has been extracted from each HIS source is utilized with a replicated storage style with edge servers in order to acquire information from the HIS sources. The processes for the initial loading of each HIS source (data supplier) to the replicas and subsequent incremental updates at periodic intervals are left to the administrators of the deployed HHIEA. If data of a HIS source is requested by a data consumer then it is served from the dedicated edge server within which the HIS source replicates its contents. The physical location of the replica can be at the HIS source or at a neutral location, where data from all source(s) is assembled; the location is transparent to all users. As shown in Figure 8, the *data suppliers* in the data layer are intended to provide the appropriate services that are needed in order to create both the initial replica and incrementally update from the HIS source(s). The *data consumers* in Figure 9 in the data layer provide the appropriate services for the stakeholders to access the aggregated information that spans these multiple HIS source(s) that are now replicas in the data layer. As a result, the data layer of HHIEA exploits the replication style to create and organize the replicas of the HIS source(s), leverages the publish/subscribe architecture in order to stage data in and out of the replicas, and provides data consumers with the means to make use of health information (e.g., for aggregation in a data warehouse).

This approach to the data layer insures that governing HIE participants have full control over which data is shared via HIE while simultaneously decoupling performance and security issues from the operative systems of the replicas that contain data from the HIS source(s). The combination of replication and publish/subscribe into the data layer promotes the definition of security policies, privacy regulations, and permissions at this common layer. In setting up each replica from a specific HIS source, the data owner (e.g., a hospital with an EHR) can determine what information to share, as depicted in Figure 9, where information from the EMR, patient portal, practice management system, etc., can be shared. The edge servers can enable interoperability by allowing them to enforce common standards for storage (e.g., syntactic interoperability) and to manage the stored data with common ontologies (e.g., semantic interoperability). Replication to the edge server is solved separately by each HIS source which allows for a high level of customization in order to meet special characteristics of the legacy systems. Therefore, the process of edge server deployment brings the opportunity to enforce a canonical message format and a common semantic foundation for exchanging data.

*Figure 7. High Level Overview of the HHIE Architecture*

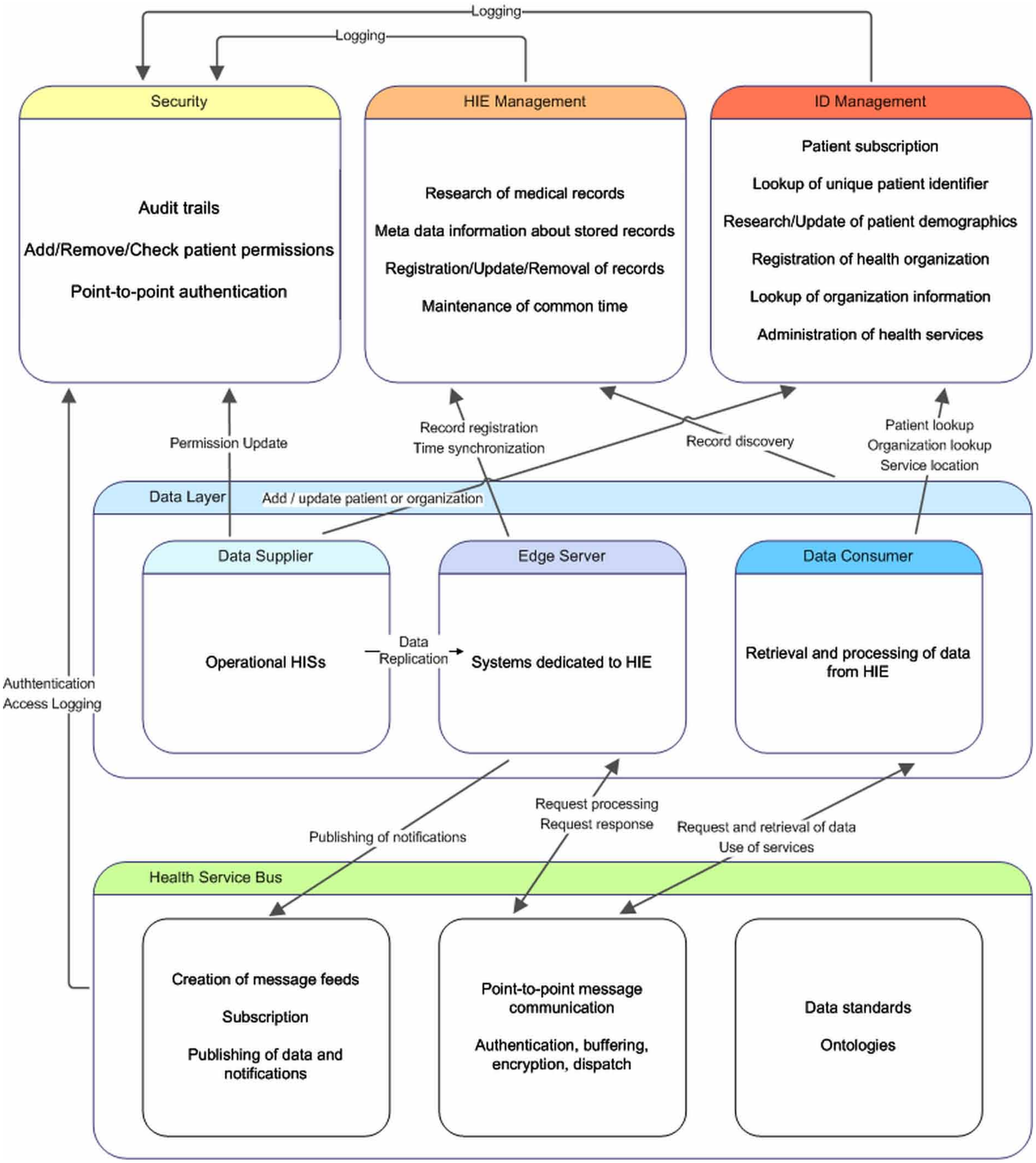
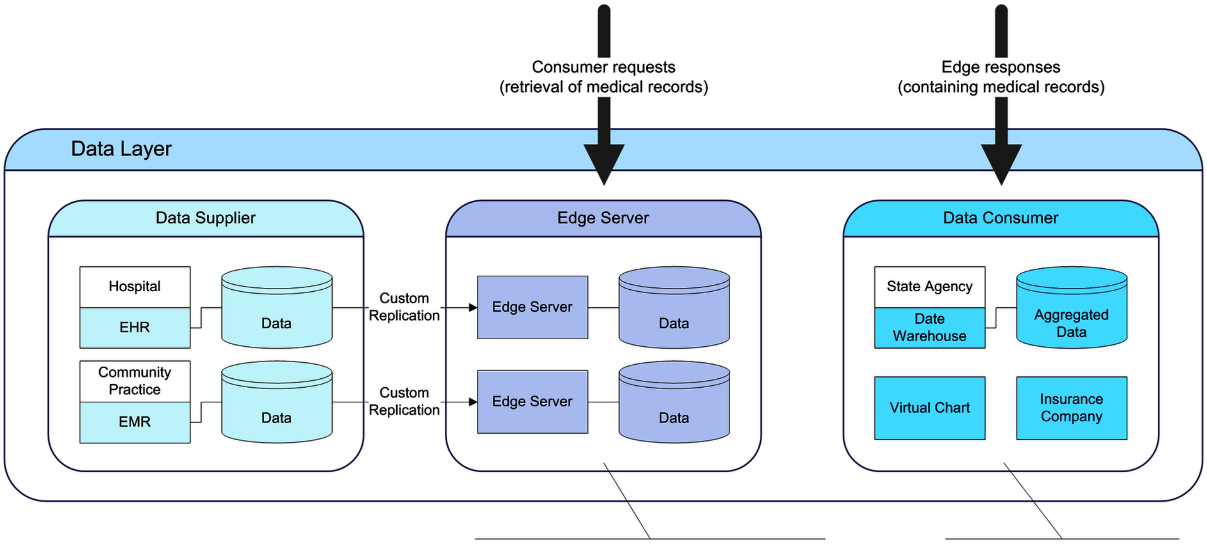


*Figure 8. The Data Layer with Data Suppliers, Edge Systems, and Data Consumers*



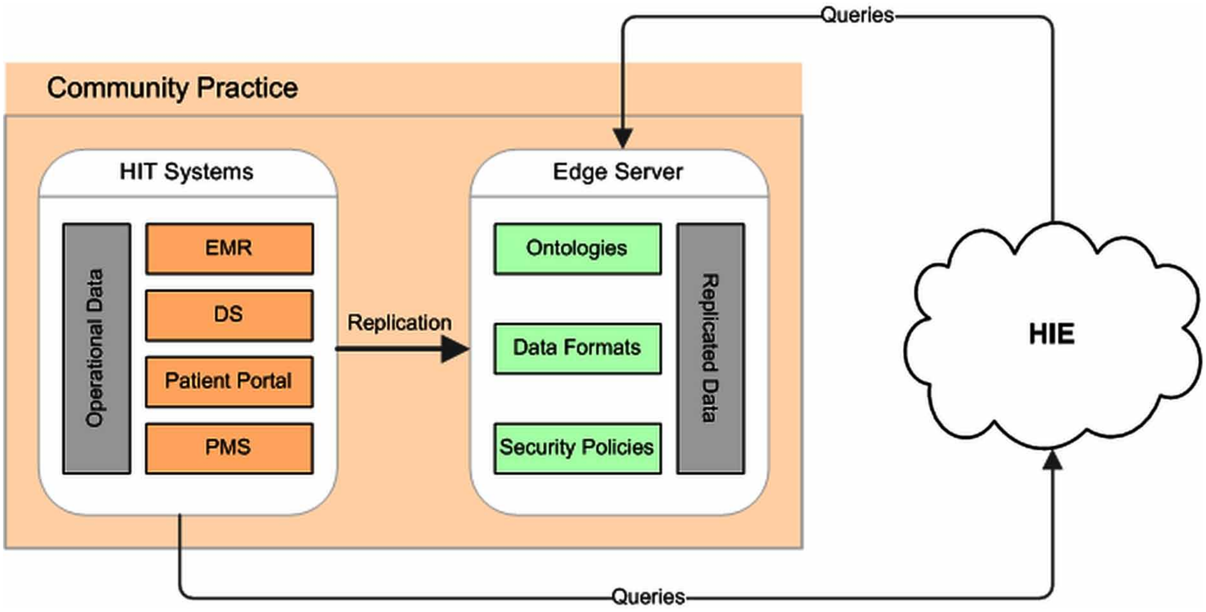
Uniform data storage systems

Physically separated from operational care delivery HIS sources

Logically outside organizational borders

Retrieving of data from the HIE: processing, aggregating, analyzing etc.

*Figure 9. A Community Practice with Edge Server System and HIE Interactions*

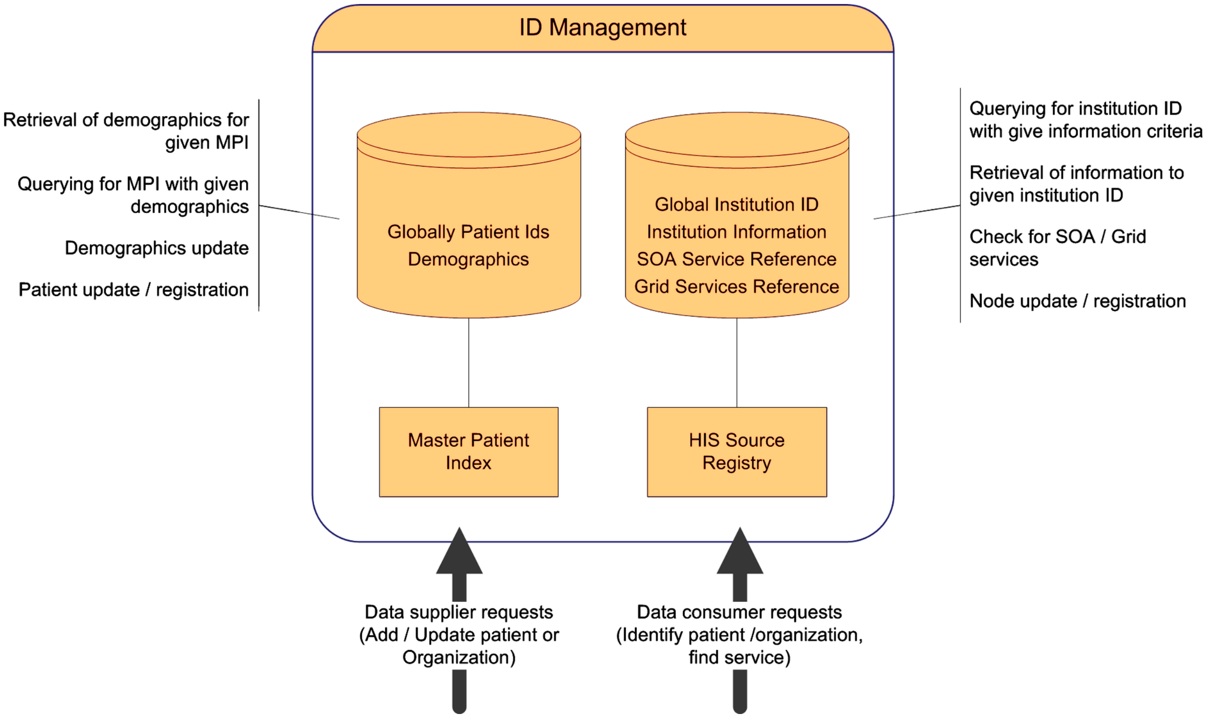


## ID Management

The components in the *ID management* group, as illustrated in the upper right of Figure 7, provide the means for HHIEA to maintain cross-organizational identification of HIE participants. For the identification of patients, a *master patient index (MPI)* component must be implemented, shown in Figure 10, in order to allow this component to register, for every patient participating in the HIE, a unique ID and storing it with a set of demographic information. Furthermore, the MPI must support two types of queries: queries parameterized with a valid global patient ID, and queries parameterized with a set of demographic information (e.g., first name, last name, date of birth, and birth place). The first type returns the demographics for a given patient for identity verification purposes and for updating demographic information (e.g., changed place of residence). With the second type, the unique identifier of a patient can be retrieved for further use (or created if the patient is not registered in the HIE yet).

The MPI component will typically be queried by data consumers of the data layer in order to retrieve the MPI for a patient, which is needed to locate patient records. Data suppliers will access the component to sign up new patients (request a new MPI) and to update existing demographics. A similar component, the *HIS source registry,* is utilized for the identification of healthcare organizations (i.e., hospitals, clinics, labs, etc.) participating in the HIE. The registry maintains a globally unique ID for each HIS source, which is distributed to an organization during an adequately controlled registration process in order to avoid service violations and to ensure a secure domain. In addition to the HIS source ID, the HIS source registry stores general information about an organization (e.g., organization type, physical address, etc.) in query language form. Furthermore, each organization can offer SOA and grid services

*Figure 10. The Identity Management Component*



via this component; complete and machine-readable descriptions to those services are also kept in the HIS source registry. The ID Management group will, in most cases, be queried by data consumers for the lookup of services. Data suppliers will access the registry to update information about their organization.

## HIE Management

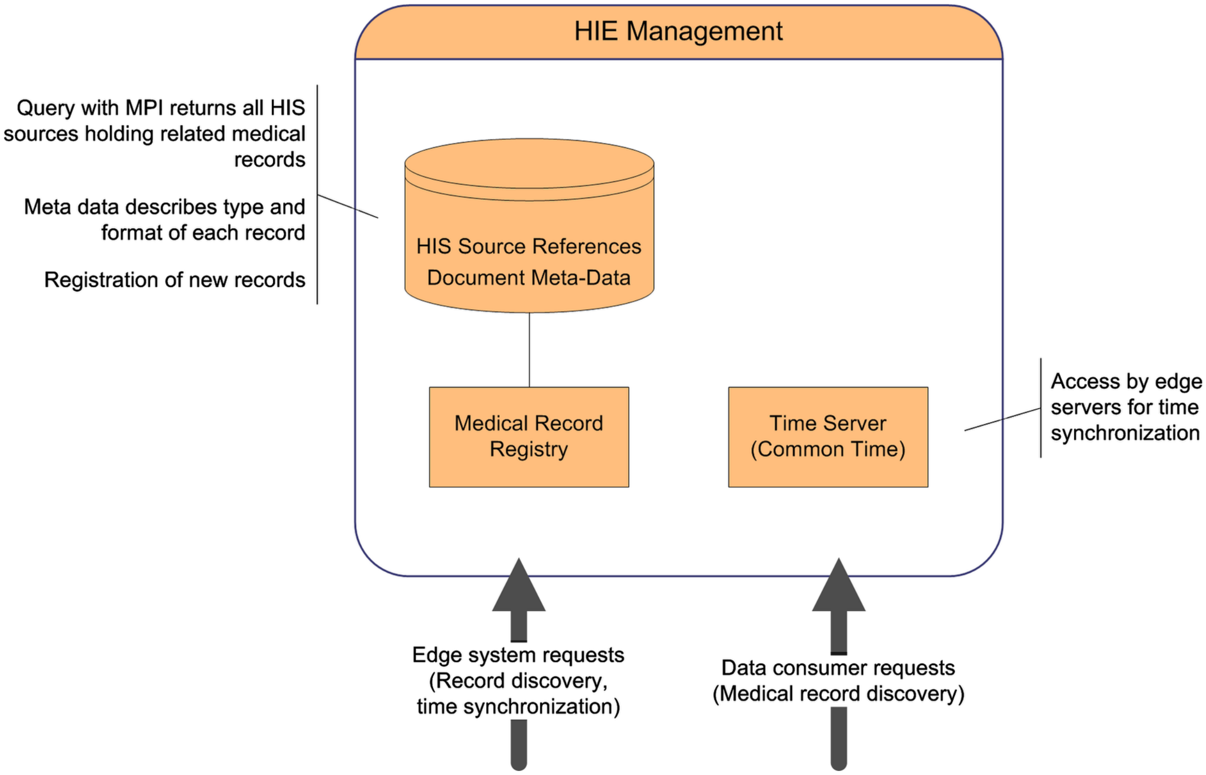
The *HIE management* group of the proposed HHIEA, as shown in the upper middle of Figure 7, consists of two components: a *medical record registry* that is capable of mapping the MPI to all of the multiple replicas that may contain data for a patient; and, a *global clock* for time-stamps in order to eliminate inconsistencies in data. As shown in Figure 11, the medical record registry identifies medical records across the HIE by processing queries from data consumers; these queries must contain a valid MPI value. After receiving a query from an eligible data consumer, the registry will return a list of HIS source identifiers which reference all of the replicas that are storing records related to the MPI in question. For a baseline, a simple retrieval of all available related HIS source identifiers can be implemented that would bring in entire medical records for a patient from the HIS source replicas. A more comprehensive querying approach may be smart enough to limit the retrieval to only recently added records for the patient. With the list of identifiers, a data consumer is able to retrieve the contact information of the data suppliers storing the records and request them via an appropriate service (which will finally be served by the replica of the data supplier). In addition to the identifier, the registry stores metadata on the data formats available for the record as well as the ontologies necessary to interpret the record's content. Based on this meta-data, a requesting data consumer can determine if the processing of the record can be managed automatically or if it will require human intervention. The medical record registry will be, in most cases, accessed by data consumers during the search for records related to a patient; edge systems will also access the registry and add information each time a new record becomes available or at some periodic interval. The global clock component addresses problems arising from asynchronous/ faulty timestamps in medical records. For example, if two records related to one incident are available on two different replicas and the timestamp of one of them is offset by a year due to a faulty set system time, an automatic interpretation of the records will result in the assumption that the incident occurred twice. To avoid such situations, all HIS sources and their replicas must synchronize their timestamps with the help of the global clock.

## Security

In order to enable the exchange of sensitive medical data, a *security* component of the proposed HHIEA is required, as shown in the upper left of Figure 7. Security and access control in the heterogeneous and federated healthcare domain is a highly complex topic and, thus, implementation details cannot be provided in the scope of this chapter. However, to support the key requirements, a security group for the HHIEA containing the set of components illustrated in Figure 12 is proposed: an *authentication* component, an *access authorization* component, and an *audit authority* component. For secure point-to-point message communication between the participating data suppliers and data consumers, means for identity authentication of communication partners must be provided. This task is addressed by the authentication component, which is utilized as a central trusted third party.

In a healthcare setting, a medical provider (e.g., physician) would typically be authenticated to their EMR and have access to an EHR. In HHIEA, the replicas bring together multiple HIS sources, and the

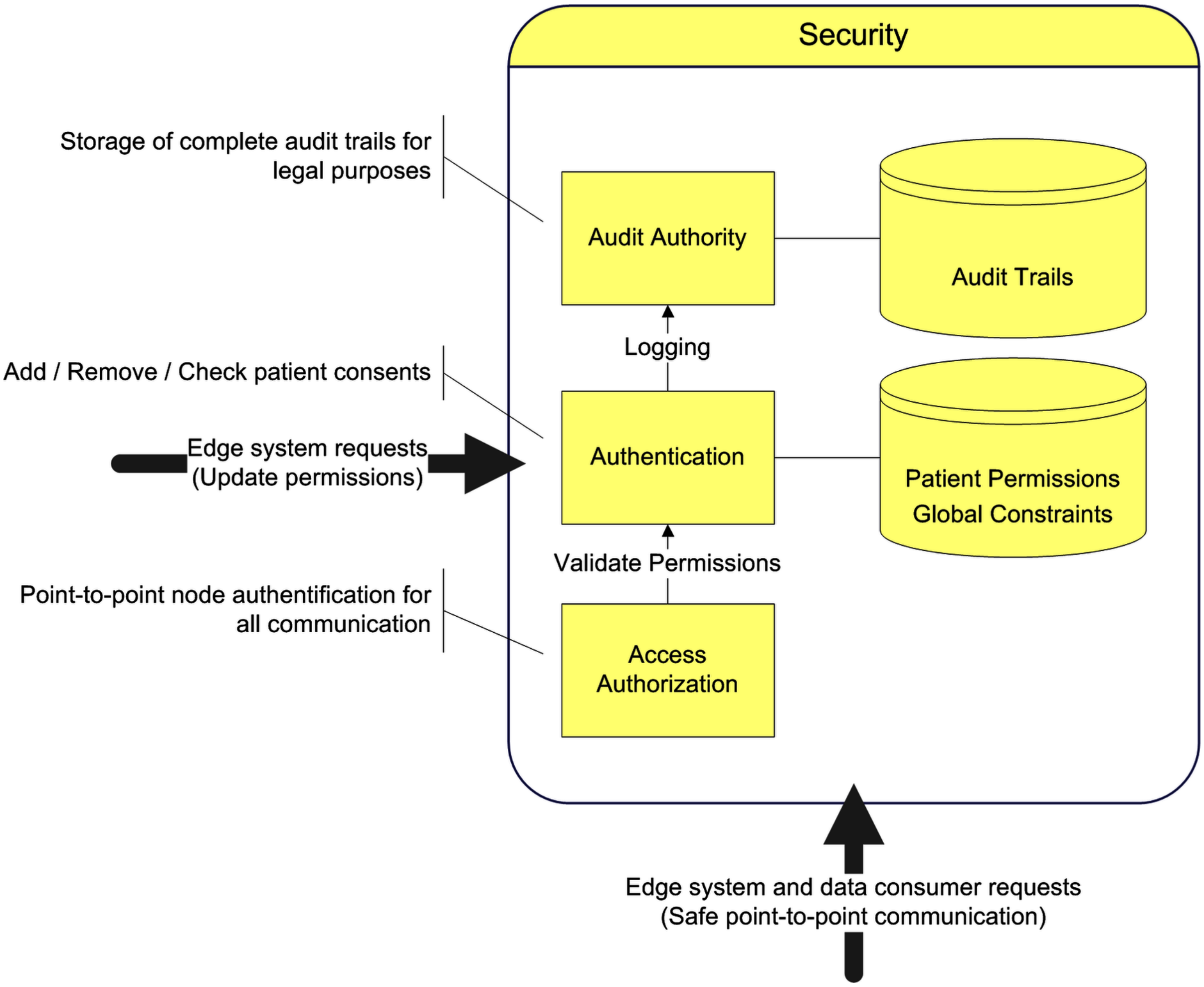
*Figure 11. The HIE Management Component*



likelihood is that the physician has not been authenticated to all sources. As a result, the authentication component must be capable of reviewing a physician's authorized privileges on the HIS sources they are explicitly authorized on (EMR and EHR) and utilize this information to determine the other sources (replicas) that would also be accessible. For the authentication component, a solution based on federated business SOA approaches built around certificates and public/private key encryption could be adopted. In addition to authentication, adequate access control is a mandatory requirement for the communication between HIS sources. This would include techniques such as: role-based access control (RBAC) (Ferraiolo, Sandhu, Gavrila, Kuhn, & Chandramouli, 2001) that focuses on the responsibilities of the users for an application per role; discretionary access control (DAC) (Na & Cheon, 2000) to allow both authority and privileges to be passed from one user to another; and mandatory access control (MAC) (Bell & LaPadula, 1973) that assigns sensitivity levels to subjects (clearances) and objects (classifications) to control access to information. RBAC can set up different privileges for different categories of users such as roles for physician, nurse, therapist, etc., that differ in privileges (what can be read and/or written). DAC can allow privileges to be passed among users such as the case where a physician delegates access to the records of their patients to the on-call physician (for nights/weekends). Lastly, MAC can classify patient data, differentiating between more generally available data (demographics, medications, tests, etc.) from very sensitive data (e.g., psychiatry records).

The access to medical records must frequently be checked against permissions given by the patient, requiring the access authorization component to store patient consents. For each request involving patient data, the authorization component has to check if the communication needs to be explicitly permitted,

*Figure 12. The Security Management Component*



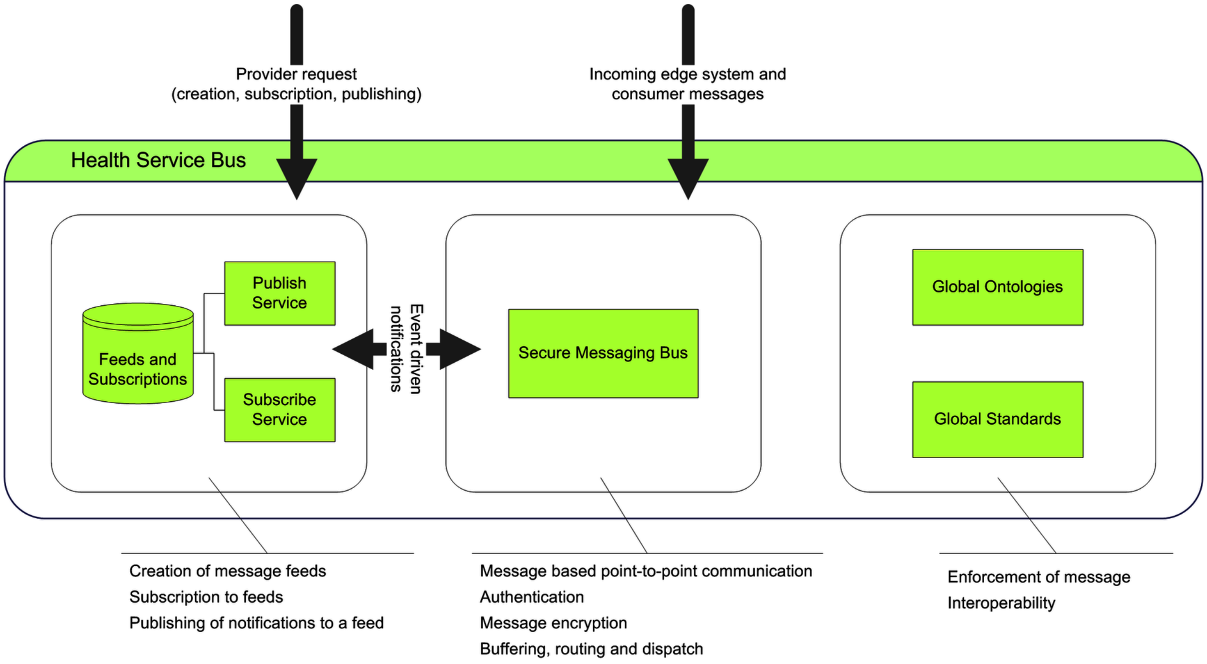
and if this is the case, it has to verify the compliance with the stored consents and global constraints. Furthermore, edge systems will access the component to update stored permissions. In this infrastructure, the security group is the point of HHIEA which is aware of all of the communication across the multiple HIS sources that form the HIE. Thus, the authentication component and the access authorization component are coupled with an audit authority, which records complete audit trails of connections between data providers and data suppliers as well as details about executed data transfers. The audit authority component provides an interface for the retrieval of those audit trails, so that they may be used to meet regulative requirements or for system maintenance and analysis. For example, while two physicians may have the same role, they may be authorized to be in charge of different patients at a hospital. As a result, most hospitals audit patient access records to insure that their employees have not been accessing patients to whom they have not been authorized (which historically has happened when a famous person is in the hospital). Note that establishing adequate security for communication in the HHIEA has significant overhead and a negative impact on the overall system performance; an estimate on complexity and the costs of this task could be attained by examining a secure multi-enterprise SOA. In addition, the Extensible Access Control Language (XACML, 2003) can be utilized to specify security policies.

## Health Service Bus

The *health service bus* component of the proposed HHIEA, as shown in the bottom of Figure 7, contains components relevant for communication of data suppliers and data consumers. The key component is the *secure messaging bus*, which enables message-based, asynchronous, point-to-point communication between the participating HIS sources. With the combination of the functionalities of the other HHIEA groups presented in the previous sections, the HIS sources are able to form messages containing all of the necessary information required for the messaging bus. The service bus receives and buffers those messages, clears security and logging issues with the components of the security component, and dispatches the messages to their targets, as shown in Figure 13. The prototypical secure messaging infrastructure can be based on different approaches such as web services as interfaces for the HIS source replicas, SOAP (Simple Object Access Protocol) as the basic message format, and an Enterprise Service Bus (ESB) solution for business SOA as the connecting element (Curbera et al., 2002). The term ESB roughly describes the concept of a middleware component which allows message-based communication between entities connected to the bus.

Custom ESB extensions for the support of grid services will be required in order to minimize security overhead for messages containing anonymized and non-sensitive data. More comprehensive ESBs will require syntactic and semantic interoperability among the replicas through the support of standards and ontologies from the message bus side. This is the location to reconcile differences in healthcare nomenclature such as the need to reconcile the various medical terms for heart attack: Myocardial Infarction,

*Figure 13. The Health Service Bus*



cardiac arrest, coronary infarction, etc. This means that the health service bus maintains a set of global standards and provides services for format transformations between those standards. Equally, a global ontology, alongside an ontology mapping engine, can translate between the different local ontologies utilized by the HIS sources. Finally, the health service bus can be equipped with a publisher service and a subscriber service. Both services are connected to a repository storing data related to message feeds and subscriptions, and can be used to implement the functionalities of a publish/subscribe architecture: creation of message feeds, subscription, and publishing of messages via feed (message dispatch can be realized through the secure messaging bus).

## HHIEA and the Regional Scenario

This section concludes the presentation of the HHIEA by examining the identified collaborative linking of HIS sources and stakeholders, where data from each HIS source is extracted to form a replica that operates as a data supplier in the HIE, and each stakeholder is interested in accessing patient data across multiple edge servers and operates as a data consumer. The goal is to demonstrate the way that the full HHIEA architecture as shown in Figure 14 can efficiently manage an HIE scenario by providing the combined functionalities of multiple architectural styles. Figure 15 visualizes the process by complementing the previously HIS infrastructure with details from the proposed HHIEA.

HHIEA components in Figure 15 are the publish and subscribe services, which are used to realize automated reporting, and the secure messaging bus, which implements the extendable message-based communication between the various HIS sources and allows for the execution of SOA/grid services. Furthermore, each HIS source is equipped with an SOAP interface for the generation of outgoing messages and the processing of incoming messages. Data warehouses add support for data aggregation and analysis at eligible HIS sources, and may also be accessed via message bus and by services offered by the HIS source. In this infrastructure, concepts from multiple system architectures are used to resolve identified collaborative links. The following reviews the usage of the various architectures in support of the realistic scenario.

- **Service-Oriented Architecture:** By utilizing SOA, the HHIEA is designed to support service discovery, patient lookup, medical record localization, and secure point-to-point communication. In addition, a full realization of the health service bus includes comprehensive support for interoperability. In combination, these functionalities provide the means for retrieving all of the available medical records related to a certain patient and, thus, for a virtual chart. This resolves the links Community Practice – Virtual Chart, Local Hospital – Virtual Chart, and PHR – Virtual Chart. Furthermore, the SOA functionalities cover all of the links which rely on the transfer of arbitrary data, such as Community Practice – Local Pharmacy, Local Pharmacy University Health Center, Local Pharmacy – Local Hospital, Insurance Company – Community Practice, and Insurance Company – Local Hospital.
- **Publisher/Subscriber:** The proposed health service bus is equipped with a publish/subscribe service, the infrastructure for the administration of message feeds, and the means to dispatch messages containing arbitrary data. This resolves all of the links related to event-driven tasks (e.g., reporting of medical cases) such as Community Practice – State Agency, Local Hospital – State Agency, University Health Center – State Agency, Pharmaceutical Research Center – Local

*Figure 14. A View of the Detailed Hybrid HIE Architecture*

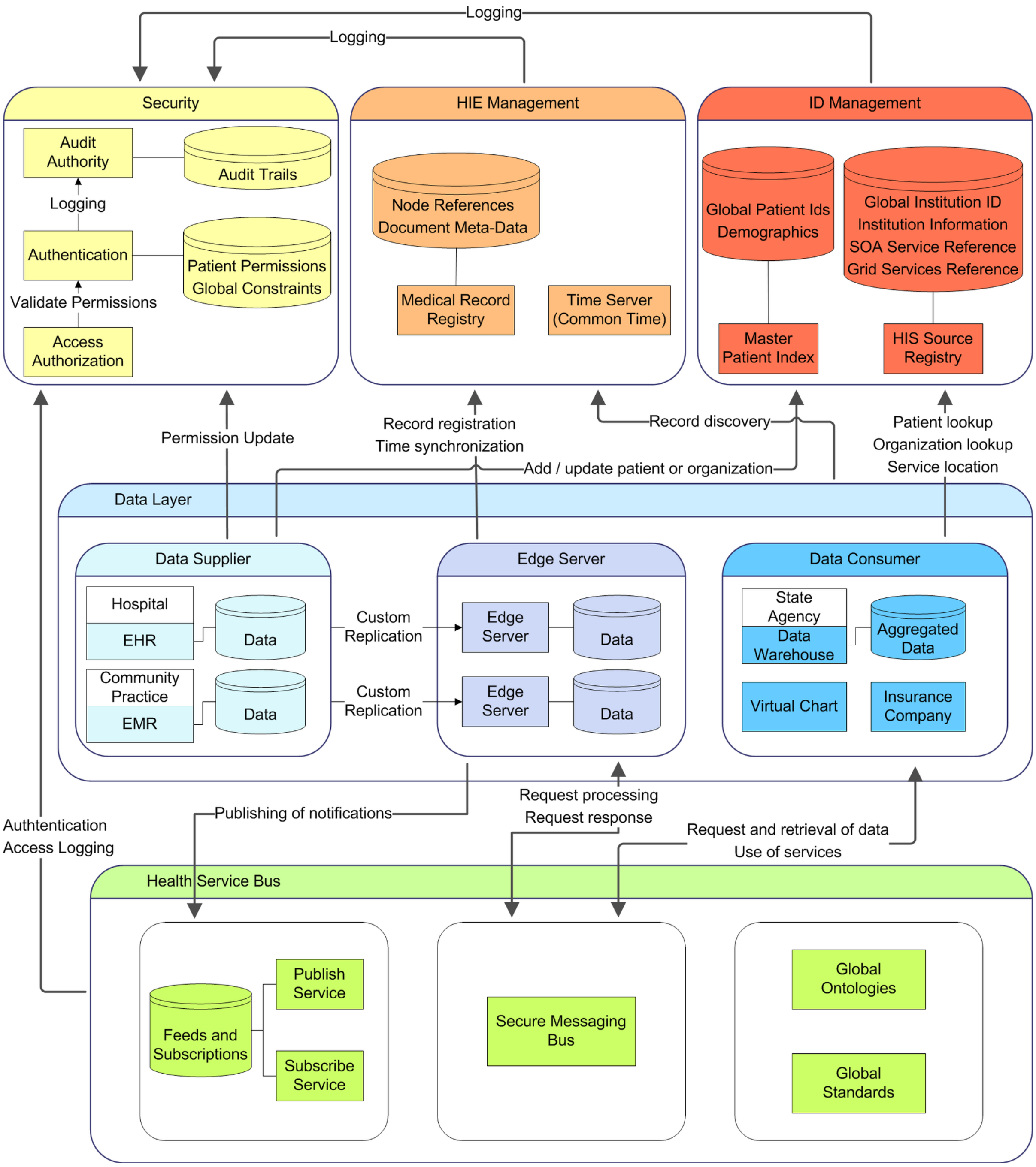
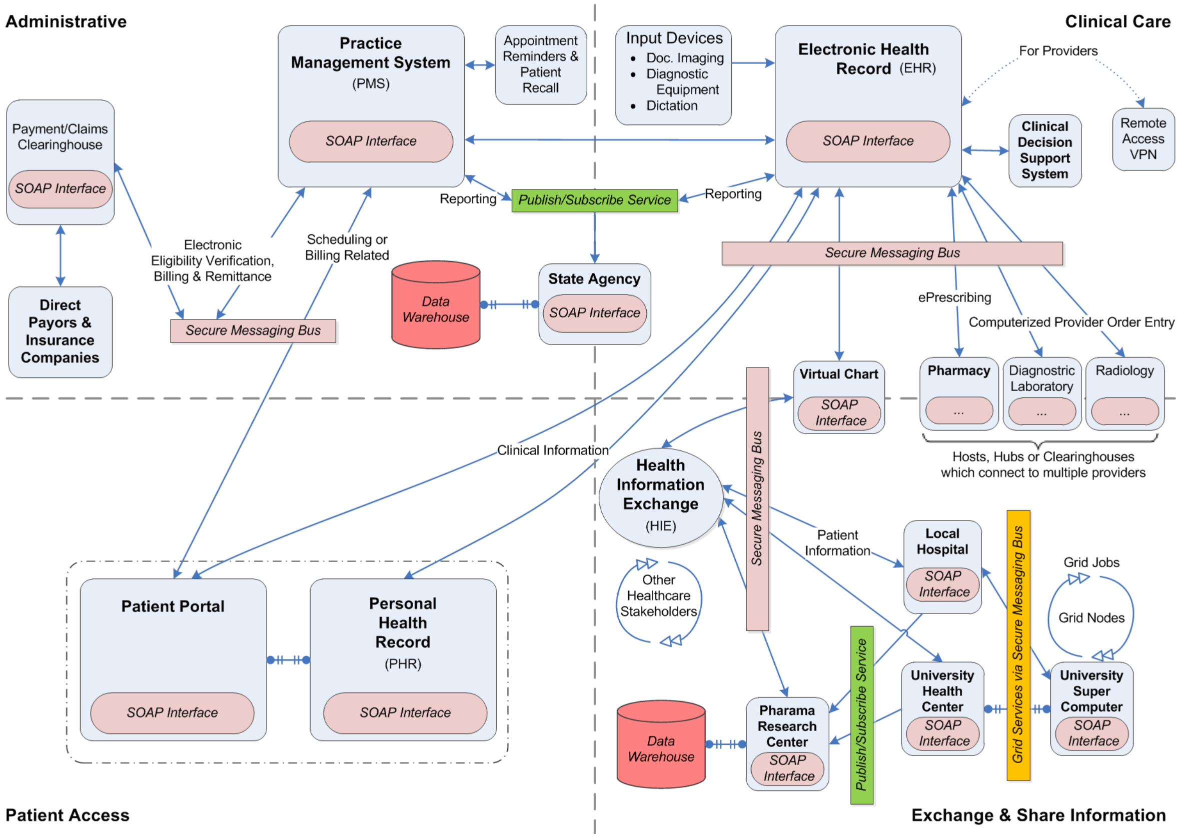


*Figure 15. HHIEA Applied to HIS Infrastructure*



Hospital, Pharmaceutical Research Center – Community Practice, and University Health Center - Pharmaceutical Research Center.

- **Cloud Computing:** The HHIEA contains multiple components that are suitable to benefit from the cloud computing abstraction. Most importantly, the proposed edge server infrastructure can run on a cloud infrastructure with an adequate deployment model and largely benefit from its elasticity (i.e., on-the-fly relocation of resources to replicas that experience increased system load) and low up-front cost (i.e., decrease of initial implementation barriers). Further targets for the cloud model are the installed data warehouses as well as any of the dedicated HISs (e.g., EHRs, EMRs, PHRs, PMSs, etc. that are integrated).
- **Grid:** The HHIEA provides all of the means for the execution of grid applications: grid service discovery, grid service descriptions, and a secure virtual domain with point-to-point messaging. This resolves all of the links which are related to applications requiring large amounts of computational or data storage resources, such as Local Hospital - University Supercomputer Center and University Health Center – University Supercomputer Center.
- **Data Warehouse:** Since the proposed infrastructure effectively supports reporting, as well as the transfer of arbitrary medical data, data warehouses can be installed at all of the HIS sources which require the collection and analysis of large data sets, such as Stage Agency and Pharmaceutical Research Center. Note that the creation of a data warehouse can occur through incremental event-

driven reporting via publish/subscribe or through mass extraction of data from data suppliers (HIS sources) of the HIE via an adequate service. Access of the data warehouses can be either limited to the governing HIS source or made available to the whole HIE via service.

## FUTURE TRENDS

This section identifies and reviews trends that will impact the evolution of HIE technologies and solution approaches. There is a two-fold emphasis for the discussion. First, the trend for modularization of healthcare applications is presented, as there are technologies that are nearing fruition towards a viable solution for HIE. Second, the need for integration genomic testing and research data with clinical care EHRs is discussed, which is more speculative in nature and represents an emerging need as genetic testing is utilized by medical providers as part of the diagnosis and treatment processes.

### Modular Application Architectures

Recently there have been noteworthy efforts in developing modular frameworks for healthcare applications that are coming to the forefront as possible solutions for HIE. The Substitutable Medical Applications, Reusable Technologies (SMART, 2015) platform chooses an app-centric approach to modularization, which is inspired by the use of apps on smartphone platforms. The approach is based on the abstract specification of a SMART container that encapsulates healthcare data and exposes a well-defined application programming interface (API) that allows interaction with the data. In theory, any HIS source containing relevant data (e.g., EHRs, EMRs, PMSs, PHRs, etc.) can be turned into a SMART container by implementing the container API on top of the used HISs. The API is designed to be used by SMART apps, which perform tasks based on the containers data (e.g., detection of drug-drug interactions or the visualization of genomic test results). The apps are reusable in the sense that they only conform to the container API and are unaware of the details of the underlying HIS source and therefore can be executed by any container. They are substitutable, since the container abstraction decouples functionality from data and, thus, allows competition between different apps (e.g., there can be multiple apps for visualizing blood glucose levels). This approach lowers the complexity for software developers and fosters the creation of a quickly growing app collection (Mandl et al., 2012). From the perspective of HIE between multiple containers, the SMART architecture has to rely on functionalities implemented in its apps and will eventually require another level abstraction/synchronization (ensuring that container A and container B run compatible apps for a given task) to establish collaboration. In the SMART world, each EHR/EMR vendor could be asked to provide a SMART app to their product, where the apps would all have identical APIs in terms of the services. This would allow a developer to access multiple HISs through a similar interface thereby supporting both reuse and HIE.

Another modular approach for the healthcare domain is Open mHealth (Open mHealth 2015), which provides an architecture description based on abstract components that can be used to build healthcare applications. These components are classified as data visualization units (DVU), data processing units (DPU), and cache units (CU). By providing specifications for implementing the external behavior of the components, Open mHealth aims to define the interaction of the components. This abstraction allows for separately developing components, such as, a DVU for presenting bar charts of a fitness marker can be developed separately from the DPU aggregating the data a fitness tracking service and a PHR. As a result, these components can them be combined, such as a DVU presenting the fitness marker results as a spread sheet can replace the bar chart DVU. Here the collaboration of multiple HIS sources that are built according to the Open mHealth architecture relies on the capability of expressing HIE requirements within the specifications.

### Integration of Genomics into EMRs

With the progress of genomic analysis methods and the decrease of cost for genotyping procedures, genomic data and knowledge is becoming increasingly comprehensive and available. As a result, its usage for patient diagnosis, treatment, and even medication selection, may now be based on genetic information. In support of this, the physicians must be brought into a process so that they can effectively utilized genetic information for patient care. In fact, there needs to be an HIE of genetic data with the diagnostic data found in an EHR/EMR. The major initiative of note in this area is the Electronic Medical Records and Genomics (eMERGE, 2007) Network that is striving towards the goal of providing approaches for combining and reconciling data from specialized DNA repositories with data collected in EMRs. This integration of research data and clinical data allows for advancing genetic research on the one side and significantly improving clinical care on the other side.

Genomic research investigations can benefit from this linkage by exploring a broad data set that is already available as a result of providing clinical care to patients (i.e., there are no additional costs for data collection) and contains information required for phenotype analysis (Denny, 2012). This facilitates studies that attempt to link certain genes to diseases or drug intolerances. Physicians equipped with EHRs that integrate genetic profiles have the means to provide care based on this additional dimension of knowledge. This includes options such as: better risk assessment and early prevention if the genetic profile of a patient indicates traits that research linked to diseases (Chute et al., 2013); improved diagnosis through decision support systems that make use of genetic profiles (Overby et al., 2013); and, accurate and personalized treatment based on results from pharamacogenomic research (Relling & Klein, 2011).

Integration of genomics data into EHRs and EMRs creates a new set of requirements for HIE architectures. First, data stored in the clinical care records has to be made available for data mining in the genomics context. This includes safety and privacy preserving access to the data as well as preparation of the data by establishing common formats and extraction procedures (e.g., extraction of data from free text notes) of data relevant for genomic research. Second, EHRs/ EMRs have to be redesigned to match genomics enabled workflows (e.g., preemptive ordering of tests based on a patient's health history and health markers) and to process genetic test results (frequently large and highly complex data sets); this may require sophisticated user interfaces for medical providers accompanied by education and training.

## CONCLUSION

This chapter has studied the alternative approaches to data and system integration in the healthcare domain utilizing a select set of well-established software architecture styles. Towards this goal, the concept of health information exchange (HIE) has been introduced as a means to integrate data from multiple HIS sources (e.g., EHRs, EMRs, PHRs, PMSs) which required the reconciliation of a set of diverse requirements to HIE-enabled systems and a review of the challenges that HIE architectures need to overcome in order to establish interactions between these systems that facilitate the collaboration of medical providers. In support of this goal, a selection of software architectures (service-oriented architecture, grid computing, publish/subscribe paradigm, and data warehousing) was studied, presented and discussed in the context of these requirements, which allowed us to highlight strengths and weaknesses that each architecture exhibits for a given aspect of HIE. To allow a further inspection of the capabilities of the different introduced architectural styles, a regional HIE scenario was detailed around a real-life HIS infrastructure that connects multiple and diverse healthcare stakeholders and is based on significant input from our medical collaborator. From this scenario, a set of collaborative links was extracted that can be used as a benchmark for the capabilities of an HIE architecture. Utilizing the architectures and the scenario as a basis, elements from SOA, grid, publish/subscribe, data warehouse architectures, and a replication approach to data storage were subsequently utilized to propose a hybrid HIE architecture (HHIEA), which has been shown to meet the established requirements and the identified links. The usage of the studied architectural styles for the proposed HHIEA is coupled with a focus on service-oriented components which is easing the process of integrating new HIS sources over time. This chapter was concluded through a discussion of modular application architectures for healthcare that promote both abstract application development and HIE, coupled with the emerging need for integrating genetic and patient data for more effective treatment.

## REFERENCES

Andrews, C., & Mack, R. (2011). IBM to Collaborate with Nuance to Apply IBM's "Watson" Analytics Technology to Healthcare. *IBM News Room.* Retrieved from http://www-03.ibm.com/press/us/en/pressrelease/33726.wss

Apache, E. S. B. (2008). *Apache Software Foundation Synapse ESB*. Retrieved from http://synapse.apache.org

Apple Health App. (2015). *Apple Health App.* Retrieved from https://www.apple.com/ios/ios8/health/

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R. H., Konwinski, A., . . .. (2009). *Above the clouds: A berkeley view of cloud computing*. EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28.

Bauer, A. & Günzel, H. (2013). *Data-Warehouse-Systeme: Architektur, Entwicklung, Anwendung*. dpunkt. verlag.

Bell, D. E., & LaPadula, L. J. (1973). *Secure computer systems: Mathematical foundations*. DTIC Document.

Booch, G., Rumbaugh, J., & Jacobson, I. (2005). *Unified Modeling Language User Guide*. Addison-Wesley Professional.

i2b2. (2004). *Informatics for Integrating Biology & the Bedside (i2b2).* Retrieved from https://www.i2b2.org/

W3C. (2008). *World Wide Web Consortium (W3C) tutorial site*. Retrieved from http://www.w3schools.com/

BizTalk Server. (2006). *Microsoft BizTalk Server*. Retrieved from http://www.microsoft.com/biztalk/

Blue Button. (2013). *Blue Button*. Retrieved from http://www.healthit.gov/patients-families/blue-button/

caBIG. (2004). *The Cancer Biomedical Informatics Grid (caBIG)*. Retrieved from https://cabig.nci.nih.gov/

CCR. (2012). *Continuity of Care Record (CCR)*. Retrieved from http://www.aafp.org/practice-management/health-it/astm.html

Chute, C. G., Ullman-Cullere, M., Wood, G. M., Lin, S. M., He, M., & Pathak, J. (2013). Some experiences and opportunities for big data in translational research. *Genetics in Medicine*, *15*(10), 802–809. doi:10.1038/gim.2013.121 PMID:24008998

CMS. (2013). *Centers of Medicare and Medicaid Electronic Health Records Incentive Programs*. Retrieved from www.cms.gov/EHRIncentivePrograms

Curbera, F., Duftler, M., Khalaf, R., Nagy, W., Mukhi, N., & Weerawarana, S. (2002). Unraveling the Web services web: An introduction to SOAP, WSDL, and UDDI. *IEEE Internet Computing*, *6*(2), 86–93. doi:10.1109/4236.991449

De La Rosa Algarín, A., Demurjian, S. A., Berhe, S., & Pavlich-Mariscal, J. A. (2012). A security framework for XML schemas and documents for healthcare. *Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference on* (pp. 782–789). IEEE.

De La Rosa Algarín, A., Ziminski, T. B., Demurjian, S. A., Kuykendall, R., & Rivera Sánchez, Y. (2013). Defining and Enforcing XACML Role-Based Security Policies within an XML Security Framework. *Proceedings of 9th International Conference on Web Information Systems and Technologies (WEBIST 2013)* (pp. 16–25). doi:10.5220/0004366200160025

Demurjian, S. A., Saripalle, R., & Berhe, S. (2009). An Integrated Ontology Framework for Health Information Exchange. *SEKE*, *09*, 575–580.

Denny, J. C. (2012). Mining electronic health records in the genomics era. *PLoS Computational Biology*, *8*(12), e1002823. doi:10.1371/journal.pcbi.1002823 PMID:23300414

Dickmann, F., Falkner, J., Gunia, W., Hampe, J., Hausmann, M., Herrmann, A., & Sax, U. et al. (2012). Solutions for biomedical grid computing - case studies from the D-Grid project Services@ MediGRID. *Journal of Computational Science*, *3*(5), 280–297. doi:10.1016/j.jocs.2011.06.006

eMERGE. (2007). *Electronic Medical Records and Genomics (eMERGE) Network*. Retrieved from http://emerge.mc.vanderbilt.edu/

Enrado, P. (2011). Why shuttered RHIO CareSpark's chairman is not giving up. *Government Health IT*. Retrieved from http://www.govhealthit.com/news/why-shuttered-rhio-caresparks-chairman-not-giving

Eugster, P. T., Felber, P. A., Guerraoui, R., & Kermarrec, A.-M. (2003). The many faces of publish/subscribe. *ACM Computing Surveys*, *35*(2), 114–131. doi:10.1145/857076.857078

FERPA. (1974). *Family Educational Rights and Privacy Act (FERPA)*. Retrieved from http://www.ed.gov/policy/gen/guid/fpco/ferpa/

Ferraiolo, D. F., Sandhu, R., Gavrila, S., Kuhn, D. R., & Chandramouli, R. (2001). Proposed NIST standard for role-based access control. *ACM Transactions on Information and System Security*, *4*(3), 224–274. doi:10.1145/501978.501980

Foster, I. (2002). What is the grid? A Three point checklist. *GRID Today, 1*(6).

Gomes, A. T. A., Ziviani, A., Correa, B. S. P. M., Teixeira, I. M., & Moreira, V. M. (2012). SPLiCE: a software product line for healthcare. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium* (pp. 721–726). ACM. doi:10.1145/2110363.2110447

Google Fit. (2015). *Google Fit.* Retrieved from https://developers.google.com/fit/

HL7. (2007). *HL7 Clinical Document Architecture (CDA)*. Retrieved from http://www.hl7.org/implement/standards/

Haas, H., & Brown, A. (2004). *Web Services Glossary*. World Wide Web Consortium (W3C). Retrieved from http://www.w3.org/TR/ws-gloss/

HIPAA. (1996). *Health Insurance Portability and Accountability Act (HIPAA).* Retrieved from http://www.hhs.gov/ocr/privacy/

Inmon, W. H. (2005). *Building the Data Warehouse* (4th ed.). New York, NY: John Wiley & Sons, Inc.

Kenny, P., Parsons, T., Gratch, J., & Rizzo, A. (2008). Virtual humans for assisted health care. *Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments* (pp. 1–4). ACM. doi:10.1145/1389586.1389594

Kuperman, G. J. (2011). Health-information exchange: Why are we doing it, and what are we doing? *Journal of the American Medical Informatics Association*, *18*(5), 678–682. doi:10.1136/amiajnl-2010-000021 PMID:21676940

Logicworks. (2015). *Logicworks Healthcare Solutions*. Retrieved from http://www.logicworks.net/healthcare-cloud-solutions

Mandl, K. D., Mandel, J. C., Murphy, S. N., Bernstam, E. V., Ramoni, R. L., Kreda, D. A., & Kohane, I. S. et al. (2012). The SMART Platform: Early experience enabling substitutable applications for electronic health records. *Journal of the American Medical Informatics Association*, *19*(4), 597–603. doi:10.1136/amiajnl-2011-000622 PMID:22427539

Mell, P., & Grance, T. (2011). The NIST Definition of Cloud Computing (Draft) Recommendations of the National Institute of Standards and Technology. *Nist Special Publication, 145*, 1–2. Retrieved from http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

Microsoft Cloud Services for Health. (2015). *Microsoft Cloud Services for Health*. Retrieved from http://www.microsoft.com/health/en-ca/initiatives/Pages/cloud-services-for-health.aspx

Microsoft HealthVault. (2007). *Microsoft HealthVault Personal Health Record*. Retrieved from https://www.healthvault.com/

Na, S., & Cheon, S. (2000). Role delegation in role-based access control. *Proceedings of the fifth ACM workshop on Role-based access control* (pp. 39–44). ACM. doi:10.1145/344287.344300

ONC. (2015). *Office of the National Coordinator for Health Information Technology Product List.* Retrieved from http://oncchpl.force.com/ehrcert

Open mHealth. (2011). *Open mHealth.* Retrieved from http://openmhealth.org/

OpenEMR. (2012). *OpenEMR electronic health record*. Retrieved from http://www.open-emr.org/

OpenESB. (2012). *Sun Microsystems OpenESB*. Retrieved from http://open-esb.dev.java.net/

OpenMRS. (2004). *OpenMRS electronic medical record system platform.* Retrieved from http://openmrs.org/

Oracle, E. S. B. (2012). *Oracle ESB*. Retrieved from http://www.oracle.com/appserver/esb.html

Overby, C. L., Kohane, I., Kannry, J. L., Williams, M. S., Starren, J., Bottinger, E., & Hripcsak, G. et al. (2013). Opportunities for genomic clinical decision support interventions. *Genetics in Medicine*, *15*(10), 817–823. doi:10.1038/gim.2013.128 PMID:24051479

Practice Fusion EHR. (2015). *Practice Fusion EHR*. Retrieved from http://www.practicefusion.com/electronic-health-record-ehr/

President's Council of Advisors on Science and Technology. (2010). *Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward.* Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf

Prokosch, H., & Ganslandt, T. (2009). Perspectives for medical informatics. Reusing the electronic medical record for clinical research. *Methods of Information in Medicine*, *48*(1), 38–44. PMID:19151882

Relling, M., & Klein, T. (2011). CPIC: Clinical pharmacogenetics implementation consortium of the pharmacogenomics research network. *Clinical Pharmacology and Therapeutics*, *89*(3), 464–467. doi:10.1038/clpt.2010.279 PMID:21270786

ResearchKit. (2015). *ResearchKit.* Retrieved from https://www.apple.com/researchkit/

Robinson, B. (2010). CalRHIO shuts down after missing out on HIE bid. *Government Health IT*. Retrieved from http://www.govhealthit.com/news/calrhio-shuts-down-after-missing-out-hie-bid

Rosen, M. (2008). *Applied SOA: service-oriented architecture and design strategies*. Wiley.

Ryan, A., & Eklund, P. (2008). A framework for semantic interoperability in healthcare: A service oriented architecture based on health informatics standards. *Studies in Health Technology and Informatics*, *136*, 759. PMID:18487823

SHARP. (2013). *Strategic Health IT Advanced Research Projects (SHARP)*. Retrieved from http://www.healthit.gov/policy-researchers-implementers/strategic-health-it-advanced-research-projects-sharp

Shortliffe, E. H., & Cimino, J. J. (2006). *Biomedical informatics: computer applications in health care and biomedicine*. Springer Verlag. doi:10.1007/0-387-36278-9_2

Singh, J., Vargas, L., Bacon, J., & Moody, K. (2008). Policy-based information sharing in publish/subscribe middleware. *Policies for Distributed Systems and Networks, 2008. POLICY 2008. IEEE Workshop on* (pp. 137–144). IEEE.

SMART. (2011). *Substitutable Medical Apps & Reusable Technology (SMART).* Retrieved from http://smartplatforms.org/

Vist, A. (2003). *Veterans Health Information Systems and Technology Architecture* (*VistA) health record and information system.* Retrieved from http://www.worldvista.org/

VMWare vCloud. (2015). *VMWare vCloud for Healthcare*. Retrieved from http://www.vmware.com/industry/healthcare/

Wager, K. A., Lee, F. W., & Glaser, J. P. (2009). *Health care information systems: a practical approach for health care management*. John Wiley and Sons.

WebSphere ESB. (2008). *IBM WebSphere Enterprise Service Bus.* Retrieved from http://www.ibm.com/software/integration/wsesb/

World Community Grid. (2004). *World Community Grid*. Retrieved from http://www.worldcommunitygrid.org/

XACML. (2003). *Introduction to XACML.* Retrieved from https://www.oasis-open.org/committees/download.php/2713/Brief_Introduction_to_XACML.html

## KEY TERMS AND DEFINITIONS

**Architectural Alternative:** A software architecture style that is suitable for a given task (e.g., data integration) under a set of domain specific requirements.

**Cloud Computing:** An architectural style based on the provision of abstract computing resources, which are capable to dynamically grow and shrink depending on changing requirements.

**Collaborative Link:** A relationship between two domain stakeholders, in which collaboration requires the integration of systems or data.

**Data Warehouse:** An architectural style that collects data from multiple sources to provide a uniform view for querying, analysis, and decision making tasks.

**Health Information Exchange (HIE):** The electronic transfer of medical data among distinct healthcare organizations and their health information systems. HIE makes medical and health data available to healthcare stakeholders and enables collaboration.

**Health Information System (HIS):** A system that captures, stores, processes, or shares information about the health of individuals or supports processes and workflows of healthcare domain stakeholders.

**Hybrid HIE Architecture (HHIEA):** An HIE-specific system architecture incorporating multiple architectural alternatives.

**Publish/Subscribe Architecture:** An architectural style describing the asynchronous message passing between publishers, subscribers, and optional brokers.

**Service-Oriented Architecture (SOA):** An architectural style based on loosely coupled software services that collaborate through a connecting framework.