A Consultation Paper for the
State of Utah
Department of Health

Evolution of Public Health Information Systems: Enterprise-wide Approaches

July 10, 2007

by
Noam H. Arzt, Ph.D.
With contributions by Susan M. Salkowitz
HLN Consulting, LLC®
8449 Christopher Ridge Terrace
San Diego, CA 92127
858/538-2220 (Voice)
858/538-2209 (FAX)
info@hln.com
http://www.hln.com/

Acknowledgement

This white paper is commissioned by the Utah Department of Health's Office of Public Health Informatics as a reference for strategic planning of the Utah Public Health Integrated Data Resource.

Disclaimer: This white paper expresses the opinion of the authors only and should not be construed as Utah Department of Health policy or position.

Table of Content

Summary

Utah Department of Health intends to develop a Strategic Implementation Plan for Utah Public Health Data Integrated Resource, informed by other states' experience and lessons learned in this undertaking. State efforts span a time period during which there has been an evolution of technological advances as well as new understandings of data driven enterprises made possible by new technology and the value they may impart to the organization's mission. During any time period, states go through changes in political and organizational leadership, reorganization and realignment of public health functions and their relationship to Medicaid, changing budgetary resources and governmental priorities. It is therefore important to view states' health integration efforts with an understanding of how they may develop and adapt their strategies to these changing realities and how the perception of *success* or *failure* may be defined by the value that they can create under a changed set of realities.

This article reviews the evolution of public health systems, models for integration, and key concepts related to integration strategies and deployments. Examples are provided from a number of leading public health agencies around the country and indicate how their strategies have managed and/or survived change.

From Stovepipes to Integrated Systems

Introduction

Public health systems have evolved steadily from fragile, program-specific products to robust, often integrated systems. Public health agencies need to consciously develop strategies for agency-wide system integration, and consider a wide set of situational attributes in determining the proper course of action. As the broader healthcare community on which agencies increasingly rely for their data continue to get more sophisticated in their health information exchange capabilities and approaches, public health agencies must understand how they can participate and integrate with this new world.

CASA	Clinic Assessment Software Application (1992)
LIMS	Laboratory Information Management System
PHLIS	Public Health Laboratory Information Systems (1989)
VACMAN	Vaccine Management System

Figure 1 – Sample CDC Applications

Over the past several years, public health systems have evolved significantly, both from a technical and programmatic standpoint. Many of these systems began in the 1980's (or even earlier) as program-specific, stove-pipe systems often based on aging mainframe or early, weak standalone personal computer technologies (Figure 1). These systems originated in a variety of places. The Centers for Disease Control and Prevention (CDC) provided many such applications to health agencies thirsting for automation solutions. Users of these systems were often epidemiologists and others with public health analytical skills which they used to perform "programming" functions to tweak system functionality and performance to match the requirements of their jobs and to supplement the agencies' limited information technology staff. These agencies often had little or no internal capability or expertise to develop core information systems, as the information services function of these agencies was quite new and more focused on basic computing needs: desktop support, personal productivity applications like text processing and electronic spreadsheets, and basic network connectivity to support file and printer sharing. In these early years the Internet was just emerging as a mainstream, general purpose information highway; email was still a distant dream. Unfortunately, the CDC itself had limited funds and limited expertise in software development. What began as innovative applications over time atrophied into slow-moving, non-scalable limited systems that did not keep up with changes in hardware and operating system capabilities, nor at times keep up with changing functional requirements.

As personal computers became more powerful and operating systems became more useable with the advent of Microsoft Windows, two things began to occur. First, for a lucky subset of systems, CDC was able to update the products to make use of these more modern features and capabilities. Software was updated from DOS to MS-Windows. Additional printer support was added. In some cases, networking features were added to allow simple multi-user access. Second, public health agencies themselves began to recognize that information technology was a legitimate target for investment to improve their ability to perform core public health functions. Agencies began, on their own, to upgrade, replace, or create new systems that were more robust and specialized using modern database management systems and tools on more reliable platforms (Figure 2). The Internet began to come into its own, and the CDC promoted its first wide area communication and system integration projects through its Information Network for Public Health Officials (INPHO) initiative in 1993.

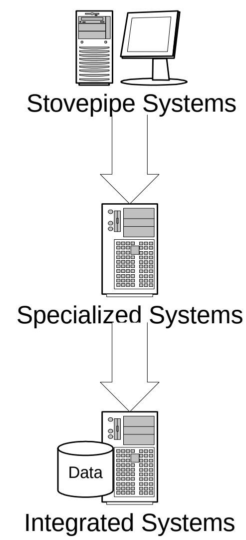

Figure 2

One of the interesting changes that took place with the evolution of technology was the evolution of technical support function within agencies. These changes paralleled developments in other industries. Computing technology moved fairly quickly from the glass confines of the dedicated computer room to the general-purpose office with the advent of personal computers, smaller and less expensive servers, and networking "for the masses" in the form of local area networks and the Internet. As these technologies proliferated, so did a tendency for computing support to move from the professional IT organization out into the field, either as supplementary IT staff or decentralized/distributed central staff. For many managers, this provided a greater sense of control over IT staffing resources which were becoming increasingly important for the delivery of the agency's mission as information management became more central to the agency's activities.

As the backlog of desired IT applications grew and grew, these local IT professional were called upon more and more to fill the gap between needed staff and available, formal IT resources. A pattern soon emerged, and depending on the prevailing political winds, IT staff organization swung between two extremes and back again: as local IT staff grew, central administrators moved to collect these individuals together into a central IT group (either in the agency, or across the larger jurisdiction) under the guise of cost savings or development of a more professional organization. Typically, program managers felt a loss of direct support as they now had to contend with a larger IT organization at least one step removed (or more) from its programmatic functions, and had to fit into a larger prioritization for IT resources with which they may or may not agree. Over time, if this centralization persisted, new local IT staff invariably crept into the programs either as employees with non-IT titles or as contract consultants and the cycle began all over again. Many programs had some autonomy, usually based on non-state funding, to engage consultants as application developers and/or to purchase applications and services with nominal Central IT oversight.

It is against this backdrop that applications evolved, sometimes under the direction of trained IT staff, sometimes under the direction of local IT support personnel who may or may not have had the appropriate technical background or training. Either way, these individuals usually showed great dedication toward trying to achieve their goals with the tools at their disposal. Each swing of the organizational pendulum, however, increased the instability of some of these applications as well as the risk to the agency. Moreover, as the responsibility for applications, and application architectures, and standards moved with each swing, remaining staff in the programs may or may not have shared the opinions of the central IT organization about the appropriate strategic direction. The potential for organizational disruption – overt or covert – was not always recognized or managed as agency-wide decisions about IT were made and implementation plans developed.

An additional note needs to be made about standards. Over the past thirty years, standards for technology have become an important fixture of organizational computing life. In some jurisdictions, detailed standards for IT are developed centrally and promulgated throughout the organization. In other organizations, more general guidelines are developed and individual projects or programs have more autonomy about the choices they make. Enforcement of standards can also vary greatly depending on the political support for the IT organization and its leadership, the mechanisms through which budgets are awarded, approved, and monitored, and the practical implications of *non*-compliance (for instance, a program needs to comply with organization-wide networking standards if it chooses to be able to connect and function within that larger technical environment). It is not uncommon for some jurisdictions to tie compliance with standards to fiscal management. What often results is a system whereby procurement using jurisdiction funds are scrutinized for agency standard compliance, but funds from grants or other sources are simply outside of the purview of this review. Given the proportion of public health funds typically received from Federal or other external sources, this only causes additional complexity and confusion in some jurisdictions.

Two Types of Integration

As the years went on, some agencies recognized the limitations of deploying systems purely within individual programs when the information related and their limited funds for technology could be better spent if leveraged across multiple projects. As applications became more network-aware and network-dependent, the need to leverage *network* investments became critical. Similarly, as systems moved to use more sophisticated relational database management systems (RDBMS) the pressure to share these expensive software licenses increased. These agencies developed a broader vision and some of their systems evolved into integrated systems supporting a wider variety of patient-centered or case-centered functions. The success or failure of their integration efforts, however, rested not only on the integration strategy they selected, but also on organizational understanding of a shared vision and readiness to engage and cooperate, as the examples below illustrate.

But there were other reasons for integration – business drivers that allowed agencies to see the potential benefit to a more planned and sustained investment in their systems. For some agencies, systems integration allowed them to understand more comprehensively who among their citizens were receiving services from the agency. By bringing together information from many programs, these agencies hoped to be able to better serve their citizens by reducing the oppressive burden of forms and paperwork faced by many people that became barriers to receiving service in the first place, and increase the coordination across agency programs to ensure that only eligible people were receiving the appropriate services. These efforts paralleled the activities in hospitals to implement unified patient intake and scheduling systems as a precursor to computerized patient order entry (CPOE) systems for inpatient facilities or comprehensive electronic health record systems for both inpatient and ambulatory settings.

Not only were agencies interested in increasing the coherence of their service to the *outside* world, they were also interested in integrating aspects of their service delivery programs *within* their agencies. This was yet another driver for systems integration, and in some agencies this was completely transparent to the citizens who ultimately were the target of these services. Information integration had the potential to enable cross-program cooperation by offering something the programs could not easily create individually. By anticipating these new capabilities programs could rethink the way they approached service delivery and organization.

Two types of integration can be considered (Figure 3):

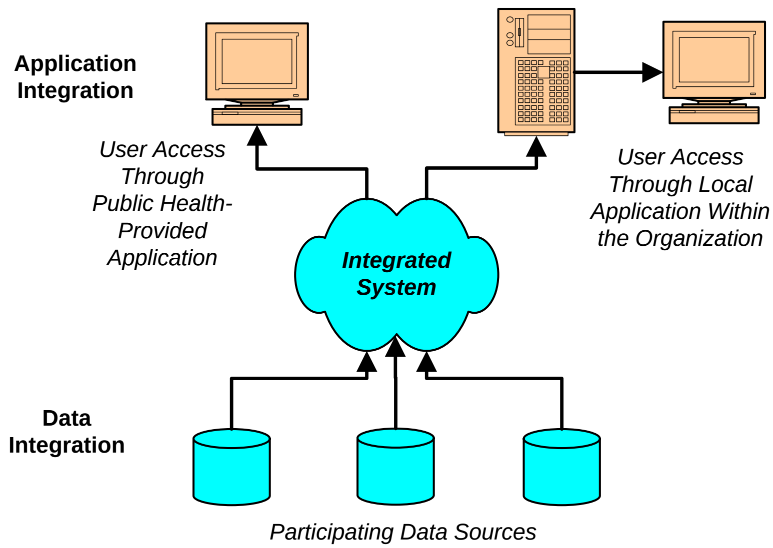

Figure 3 – Types of Integration

Data Integration involves forming valid relationships between data sources. *Application Integration* involves making data available from different sources by presenting a unified view of the data to a user through a computer application ("computer" being broadly defined as anything from a personal computer to a web browser to a smart card). These two types of integration ultimately come together in the tools, applications, and data that the end user can access and use. It is possible for a system to employ one or both of these strategies.

Let's look at a few examples. In New York City, the executive leadership of the Department of Health and Mental Hygiene wanted to integrate the Citywide Immunization Registry (CIR) and the LeadQuest (LQ), the system that supports the Lead Poisoning Prevention Program's (LPPP) blood lead level test tracking and lead abatement efforts. Managers in NYC realized that both of the systems contained records for the same children, yet they were completely separate. Healthcare providers in the community who had access to the CIR's web-based application could not see when blood lead level tests results were missing or abnormal. LPPP case workers using LQ did not know anything about children who were missing tests and did not know the immunization status of the at-risk children whose cases they were managing. The programs saw potential synergy in more field staff being able to promote mutual program goals for at risk children. The CIR was population based since it was initialized with vital records; LQ was not. Both systems had levels of data duplication that were unacceptable – CIR because of the many diverse sources of its immunization data, LQ because its primary source of data, laboratory slips, are notorious for having fragmented and inaccurate demographic data. A decision was made to integrate these two systems, there was a requirement that each system needed to maintain its own operational integrity and programmatic independence.[1] All these systems used custom-developed software owned by the agency.

A system was architected to meet all these goals. Its high-level architecture is presented in Figure 4:

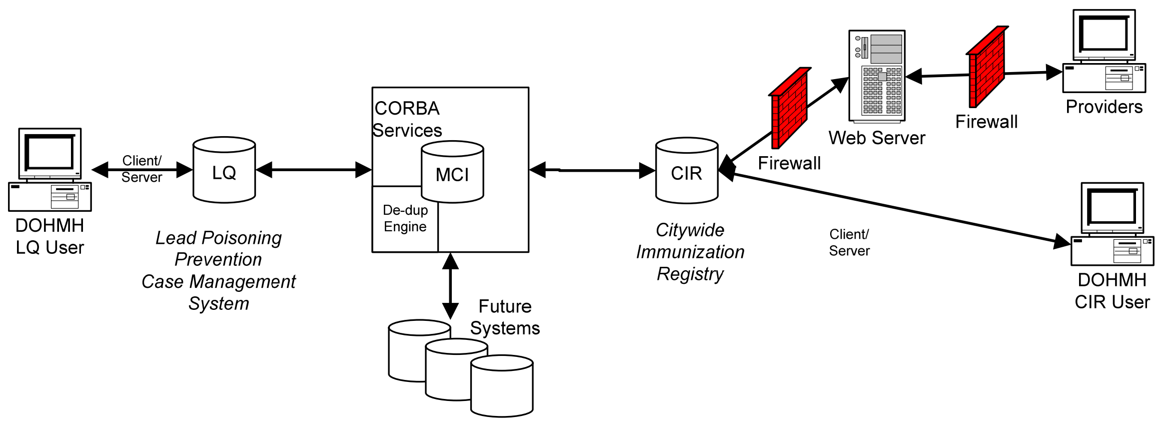

Figure 4 – New York City Integration Architecture

On the left side is a representation of LQ. LPPP users access the database via a client/server application that they execute from their office locations. CIR has internal Immunization Program users who access their data via a different client/server application, and providers in the community who access via a special web application provided for their use. Both CIR products (and a set of additional tools) access the same database. To achieve integration, a new construct was added in the center. The Master Child Index, or MCI, is a unified person database to which both the CIR and LQ register their child records and use for identifying and resolving duplicate person records both within and between their systems. Medical data about the children is not stored in the MCI, so it is not a replacement for the two participating systems. The MCI database is accessed through a standard services interface. This interface also allows access to the person record de-duplication engine. These services allow one system to query data from the other: LQ uses a service in real time to request immunization data from the CIR and present it on a LQ screen to its users, and the CIR uses another service in real time to request blood lead level test information (or the absence of information) from LQ and present it through its web application to its users.

TABLE 1 ● Number and percentage of matching results of the "initial load" data by system

	Within system		Between system	Within and between system
	CIR	LQ	MCI	CIR, LQ, and MCI
Pre-MCI, *N*	2,426,369	2,184,216	4,086,865*	4,610,585
Post-MCI, *N*	2,065,230	2,021,635	2,977,290	2,977,290
Merged, *N*	361,139	162,581	1,109,575	1,633,295
Merged, %	14.9	7.4	27.1	35.4
Human review, *N*	74,798	56,747	95,886	227,431
Human review, %	3.1	2.6	2.3	4.9

*This number represents the sum of records in each data system after MCI's internal de-duplication, ie, 2,065,230 + 2,021,635 = 4,086,865.
CIR = Citywide Immunization Registry; LQ = Lead Quest; MCI = Master Child Index.

[1] For a more detailed discussion see from Papadouka, Vikki et al, "Integrating the New York Citywide Immunization Registry and the Childhood Blood Lead Registry, *Journal of Public Health Management and Practice,* November 2004 (Supplement), pp. S72-S80.

Solving the duplicates problem within and between these two systems was the primary motivation for the deployment of the MCI. Table 1[2] shows the number and percentage of matching results from the initial load completed as the MCI became live in early 2004. Initially, 523,000 patient records were eliminated because they were duplicates. Matches were automatically merged together – 1.1 million initial patient records (37.3%) were also automatically merged – while possible matches were first reviewed by public health workers in a human review application from which they make the final determination. Since CIR records were loaded first, Table 2 shows the number and percentage of subsequently loaded LQ records that were merged with CIR or vital records during that initial load.

This system was developed cooperatively between the two programs in NYC as a custom solution, and continues to operate and flourish today. Based on this architecture and the above definitions, NYC is achieving both data integration (through the merger of person records in the MCI) and application integration (through the display of data from one system within the application screens of the other). Features were put into place to ensure that each system, LQ and CIR, could operate independently in the event that the MCI was not available or that the other participating system was not available. This allows for an architecture that at first appears tightly coupled to allow these systems to operate in a loosely-coupled fashion when there are problems or outages with major components.

TABLE 2 ● Number and percentage of Lead Quest records merged with Citywide Immunization Registry or vital records

Birth cohort	CIR	LQ	Integration merges	LQ records merged with CIR records, %
<1996 (no vital records)	851,460*	1,235,734*	494,595†	40.0
1996	157,818	133,368	105,280	78.9
1997	159,194	126,373	100,336	79.4
1998	154,415	124,180	99,236	79.9
1999	146,339	116,795	94,532	80.9
2000	150,899	107,048	87,802	82.0
2001	151,601	95,044	79,979	84.1
2002	148,015	74,892	63,228	84.4
2003	142,675	7,985	6,437	80.6
1996–2003	1,210,956*	785,685*	636,830†	81.1

Organizationally, the operation of the MCI and its services is vested in the CIR technical staff (and their contractors), but the two programs jointly administer the policies and procedures related to the system's operation. Though it was envisioned that additional systems would join the deployment in the future, that has not as of yet taken place because this has not been a priority of the current commissioner who succeeded the commissioner who sponsored the initial project, and the events of 9/11 redirected resources to other surveillance programs. There is no mandate in this agency that all or even any system register its patients with the MCI or use the MCI's services in any way. It is worth noting, however, that a communicable disease case tracking system was deployed using a clone of the MCI and its services to provide de-duplication services in that environment without risking divulging protected information about communicable disease patients had their records been present alongside other patients in the MCI. New DOHMH emphasis on chronic disease and the expansion of the CIR for adult vaccines may result in a revisit of the policy and yet add more programs to the extensible architecture.

Let's look at another example, this time from Rhode Island. Rhode Island's KIDSNET integrates data from multiple public health programs, including Immunization, Lead, WIC, Newborn Screening, Hearing Screening, Early Intervention, Home Visiting and Risk Response, and Vital Records. Developed in the mid-1990's, KIDSNET uses a mixed strategy: some of the participating programs have independent data systems that periodically supply data to KIDSNET through data exchange interfaces (*e.g.,* hearing screening); other programs use the KIDSNET database as their primary data

[2] Tables are from Papadouka, Vikki et al, "Integrating the New York Citywide Immunization Registry and the Childhood Blood Lead Registry, *Journal of Public Health Management and Practice,* November 2004 (Supplement), p. S77.

storage system, and KIDSNET applications directly for data viewing (*e.g.,* immunization). There are many inbound interfaces to participating systems for data collection, though most data sources upload their data file submissions through a web-based application provided by KIDSNET. Additional interfaces also exist to secondary users of data (like the CDC). KIDSNET is a custom-developed system that uses a multi-tier architecture and looks much like an internal application (Figure 5):

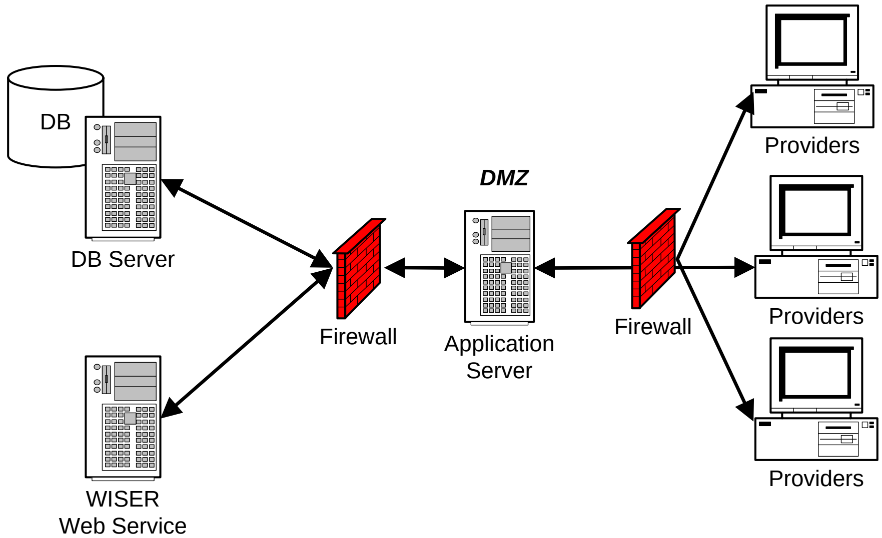

Figure 5 – KIDSNET Integrated Architecture

RI consolidated its data for many of the same reasons as NYC. There was a need to relate health data about children together in a comprehensive way to support initiatives that spanned program areas, and to provide a patient-centric view of child health to private practitioners, since in Rhode Island there are no local health departments and all primary care is delivered in the private sector. The system was also plagued by person data duplication due to the use of very conservative rules about record matching and merging, and limited software capabilities to correct the problem. KIDSNET had employed a simple deterministic algorithm for matching incoming data to existing KIDSNET demographic records. By 2004, KIDSNET had accumulated a queue of over 47,000 unmatched records. With a limited budget, RI embarked on a project to improve the matching process and to ultimately reduce the number of unmatched records. Eventually, FEBRL (Freely Extensible Biomedical Record Linkage), an open source package, was modified for use within this new framework. Probabilistic parameters were developed, and an extensive six-month testing process ensued. The solution was placed into production in May 2004.

The following graph shows the number of records "on hold" in KIDSNET during the months preceding the implementation of FEBRL in mid-2004. Several major programs are diagrammed, then the total on the top line.

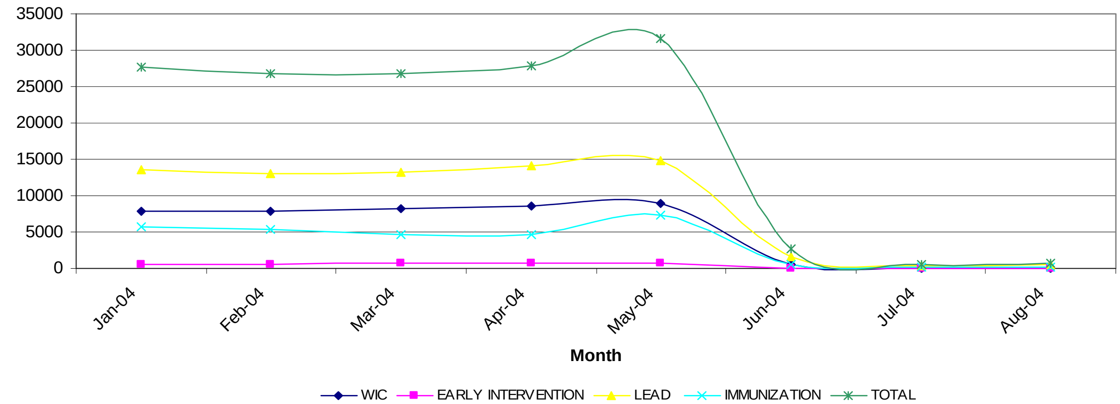

Figure 6 – KIDSNET De-duplication Results

The new process, combined with "human review" activity, reduced the number of unmatched records by over 93% in its first three weeks. Greater than 45,000 (95%) of 48,685 were removed from the "on hold" status and added to the database. Approximately 11,000 records were added to KIDSNET. Probabilistic de-duplication combined with an interactive merging interface ensures that the number of duplicates in KIDSNET will remain low even as unmatched children are added to KIDSNET. Currently, the time to resolve an error was reduced by 78%, and 90% of the data that comes into KIDSNET is immediately imported into the database and available for use.

Based on the definition above and this architecture, RI is also achieving both data and application integration, though the nature of their solution is much different than that of NYC. Unlike the MCI, the KIDSNET database contains a complete record of all the consolidated data for children in RI and adds value by relating that information together accurately and presenting it to the user through its own web-based application. In NYC, records still reside in the participating systems and are represented in the MCI only to the degree that they are useful in providing matching services. Of course, Rhode Island has an annual birth cohort of less than 13,000 children, about 10% of NYC's, and their more centralized strategy is probably more practical for them than it would be for NYC.

Enterprise-Wide Integration

Introduction

Public health agencies became more ambitious as technology became more enabling. In the commercial sector, the emphasis turned to enterprise-wide systems that offered more comprehensive and more uniform approaches to both data storage (data integration) as well as data presentation (application integration). Products from companies like Oracle, SAP, and PeopleSoft fought for opportunities to integrate more administrative functions (like accounting, human resources, and logistics). Even in large healthcare enterprises changes were afoot. By the late 1990's many hospitals abandoned their patchwork of "best of breed" applications in favor of more comprehensive, integrated, single-vendor solutions. For public health it was not so easy. On the one hand, most administrative systems were operated by larger entities (the county, the state, the Federal agency of which they were a part). On the other hand, there are few packaged application solutions focused on automating the way public health agencies did their business, though there are some exceptions, including CDP[3], KIPHS[4], and Netsmart[5] (formerly QS Technologies).

Models of Enterprise-wide Integration

These transitions were far from smooth, and not all initiatives in either the private sector or the public sector were successful. Some have yet to fully play out as these large-scale projects take years to plan and years to fully execute. How can we best think about the range of enterprise integration options? We have identified three models for enterprise-wide integration: the **Centralized Model**, the **Cooperative Model**, and the **Distributed Model**. In this section we will describe and compare each model, providing examples as appropriate.

The **Centralized Model** is typical of organizations that feel that a singular, centralized approach to data and systems is the best way to achieve operational efficiency with the highest degree of control (Figure 7):

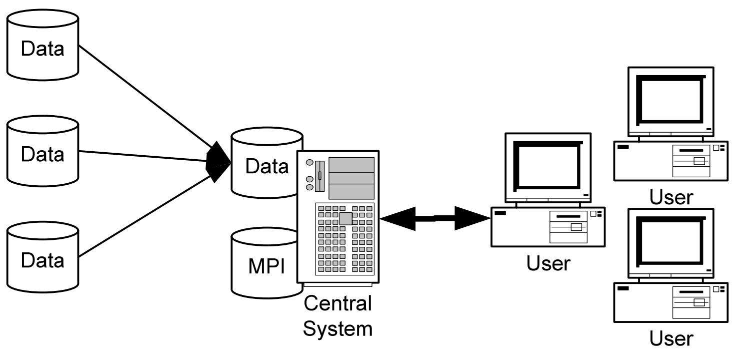

Figure 7 – Centralized Model

This model is most effectively implemented when leadership is strong, requirements are clear, and funding sources are known and available. It is also most successful when the service delivery of the agency is integrated ahead of (or in tandem with) the systems integration. When implemented correctly, the model presents a clear roadmap and can achieve large-scale results. While initially original or legacy systems may start in a distributed configuration, implementation of a centralized model typically brings disparate systems together. This model employs a centralized MPI where all person records are registered, matched, and de-duplicated. Data may be collected from a variety of sources behind the scenes, but users typically access a single consolidated application or suite of applications. Since the data and applications are more centralized data security tends to be more straightforward. Agencies interested in this comprehensive approach will likely need to make significant, sustained investments, but the potential payoff from integrated systems and data may bring significant benefits to the organization.

The centralized model does not need to be implemented agency-wide. Typically, a reasonable subset of the agencies applications, systems, and programs are selected for this approach. KIDSNET (see above) is a good example of a centralized model implemented among a reasonable subset of agency programs, in this case programs related to early childhood health and development. Another example is the

[3] See http://www.customdatainc.com/
[4] See http://www.kiphs.com/
[5] See http://www.ntst.com/solutions/public%20health/index.asp

Missouri Health Strategic Architectures and Information Cooperative (MOHSAIC), the departments integrated public health information system, which grew from a strategic planning initiative with strong executive sponsorship and developed incrementally as the plan was used to support funding sources opportunistically, starting with a CDC INPHO grant. MOHSAIC has survived significant changes in political leadership, organizational structure and the departure of its original architects and principals because it continues to provide value and serve key functions and it has been able to adapt to the new direction. [6]

The second model, the **Cooperative Model,** may be more realistic for many agencies. While still requiring strong leadership to gain the most from leveraged investments, individual programs and systems have a fair amount of autonomy in how they develop and operate their applications (Figure 8). A typical implementation of this model still employs an MPI with matching and de-duplication services, but systems continue to operate independently or semi-independently while benefiting from these new capabilities. Users continue to access their individual systems directly (shown on the lower part of Figure 8), though one or more central applications might be deployed with consolidated data (shown on the upper part of Figure 8). Security can be a bit more challenging in this technical environment as more of the systems and data are distributed. Strong standards and a genuine desire to comply with them make this model more feasible. While a significant investment may be necessary to launch the MPI and central services, existing systems can usually continue to operate while this new infrastructure is put into place. On the other hand, modifications to existing systems may be significant depending upon the implementation strategy and products selected.

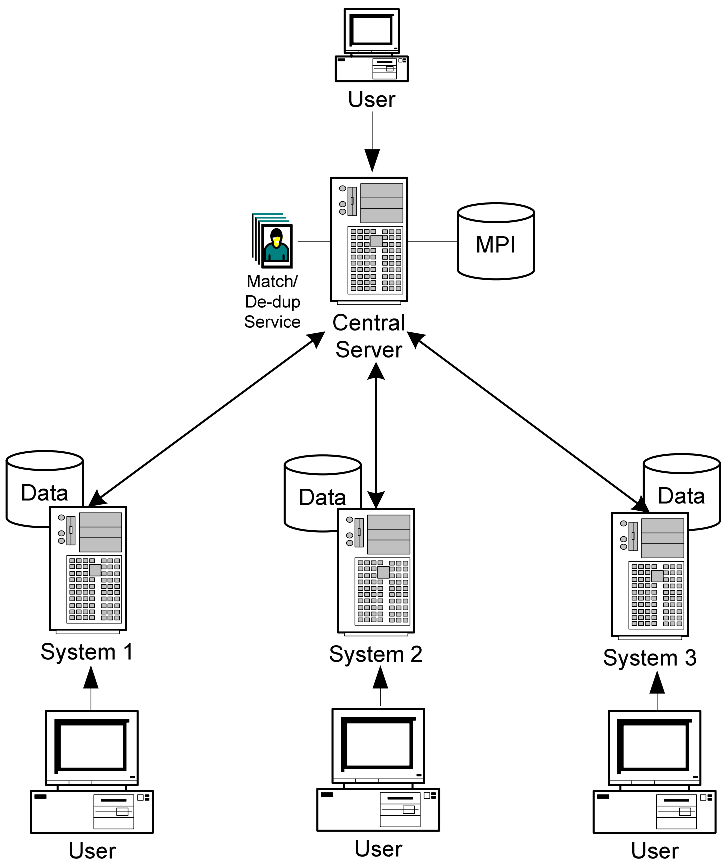

Figure 8 - Cooperative Model

The NYC MCI (see above) is a good example of the cooperative model in practice. Each program is politically independent, but they have agreed to work together and integrate their system through an MPI while maintaining the autonomy of their individual systems. Significant modification was required (and continues to be required) to allow CIR and LQ to make use of the central services, and to continue to operate if one or more of the systems is unavailable. Because the component systems in NYC are spread out over several physical locations, an additional issue is the network latency and other network failures that occasionally occur and disrupt system to system communications. The monitoring that is in place to ensure that all

[6] See http://www.dhss.mo.gov/mohsaicmoa/ and http://www.phii.org/Files-AKC/project%20briefs_08-21-05/Brief_Missouri.pdf

parts of the system is working continue to get more sophisticated and better able to deal with sub-system outages.

As cooperative systems have matured, new models have developed to help reduce system to system dependency and the risk of an interruption in the services available to users. Utah has put into place a very innovative system called the Child Health Advanced Records Management, or CHARM, which acts as an electronic broker for coordinating access by one system to another system's data.[7] Systems are not required to participate, and its implementation is both modular and incremental. Participating systems do not have much to do other than install a CHARM "agent" (a piece of specialized software) to make data available to other CHARM-enabled systems. Though the development of CHARM has been going on for a number of years, deployment has been rather slow, however, and there has been little consistent executive sponsorship for this project.

Another increasingly popular strategy for developing more loosely coupled integration is a services-oriented architecture (SOA). The concept of SOA is not new. For years software developers have created systems with application programming interfaces (API) which define how systems and subsystems interact with one another by exchanging data in reliable, structured ways. Many of the core services that are used to operate the Internet began as functions with APIs which developed into internationally-recognized standards. In an SOA, complex systems are created which are comprised of discreet functions, or services, that make themselves available to other systems on a network and perform specific tasks (Figure 9). These services form system building blocks capable of being reused over and over again in the context of different needs and applications. Diverse systems can share important algorithms, features, and capabilities by relying on these shared services rather than reproducing this functionality each time it is needed.

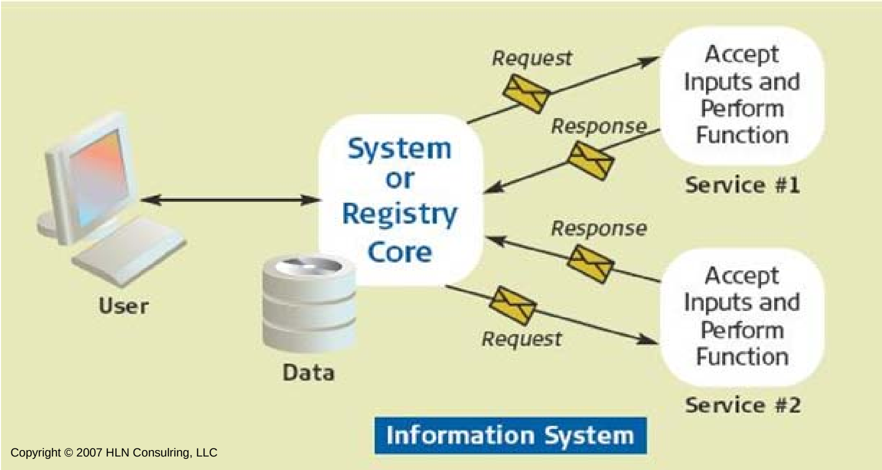

Figure 9 – Services-oriented Architecture

It is possible to use an SOA approach to integrate systems together by first breaking down key functions into service-based building blocks, then deciding how these building blocks can be assembled and used. In some cases, individual sub-systems will be able to share key services, thus saving time and funds by reducing duplication. In other cases, services will enable system to system integration by making data

[7] See http://charm.health.utah.gov/ as well as the *CHARM Data Integration Plan*, May 31, 2002. <http://charm.health.utah.gov/pdf/data_integration_plan.pdf>

available across system boundaries through a consistent and known method. Other services will develop to provide common resources to the system-as-a-whole (Figure 10): a matching or de-duplication service might be a new service made available for all participating systems to use consistently. As with CHARM's software agent architecture, SOA can in most cases effectively be implemented incrementally. Both the Massachusetts Department of Public Health and the Rhode Island Department of Health are using SOA within the architecture of their immunization registries to provide access to their immunization scheduling algorithms.[8] The Arizona Department of Health Services is also working on a "top down/bottom up" approach to implementing SOA within the agency.[9]

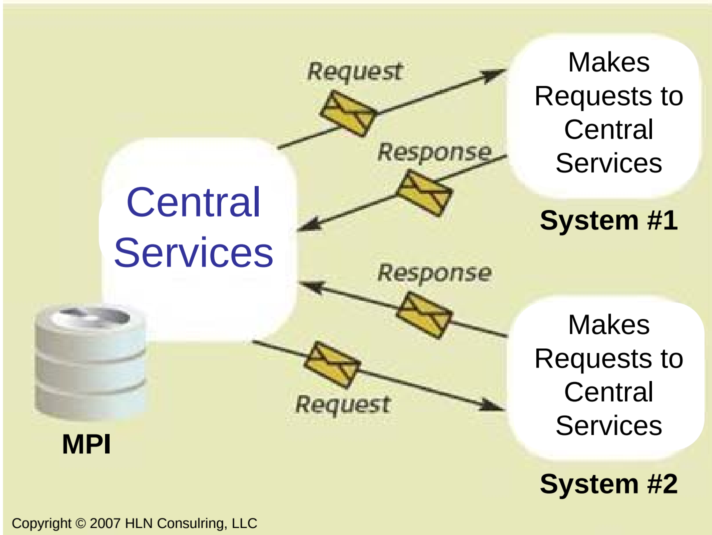

Figure 10 – SOA within a Collaborative Integration Model

The third model, the **Distributed Model** for enterprise-wide system integration, is the model of last resort. For some agencies it may represent an entirely *de facto* strategy (Figure 11):

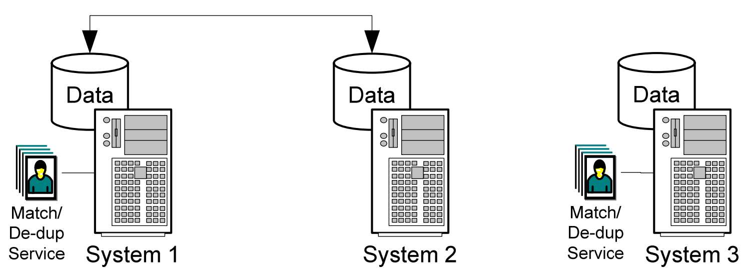

Figure 11 – Distributed Model

This model is best when there is limited, un-sustained, or uncertain leadership either at the programmatic or information technology level or both. Systems largely operate independent of one another. If data relationships need to take place they are negotiated between the individual programs and

[8] See James Daniel, "Making Your Forecasting Algorithm Available as a Web Service," 41st National Immunization Conference, Kansas City, MO, March 7, 2007. <http://cdc.confex.com/cdc/nic2007/techprogram/S7912.HTM>
[9] See John Nelson and Paul Barbeau, "Addressing Business Objectives with a Services Oriented Architecture," NAPHIT Webinar, June 8, 2007. <http://www.naphit.org/global/library/webinars/webinars_07/060807/NAPHIT_AZ_DOH_SOA.ppt>

implemented largely without central intervention or assistance. If there are matching and de-duplication services they tend to operate within the context of individual systems only – there is no consistent sharing or association of person data between systems. While standards and cooperative purchases may help to keep costs under control, there is only limited leverage possible across systems, but also limited risk of large scale deployment failure. Most agencies that are not explicitly implementing one of the first two models are almost certainly implementing this model, even by default.

Comparing Key Model Attributes

The following tables provide a summary of the key attributes of these models and compare the three models side-by-side. These tables can be used to help categorize the attributes of a particular agency or program and help assess the best fit of these three models for the setting.

The first table focuses on organizational attributes:

Enterprise-wide Integration Model Comparison: Organization			
Factor	**Centralized Model**	**Cooperative Model**	**Distributed Model**
Political Sponsorship	Strong, consistent executive sponsorship	Conceptual support but less desire or ability to plan and invest in central strategies	Issue largely left up to program directors with passive permission of agency leadership
IT Leadership	Strong PH CIO or similar position with mandated authority over infrastructure and applications	Strong PH CIO, but less clear authority over applications	Central PH IT organization primarily concerned with basic network services and desktop support
IT Staff	Strong, centralized PH IT staff; few pockets of resistance	Some centralized PH IT staff supplemented by distributed staff within programs; need to carefully manage potential pockets of resistance	Centralized PH IT staff primarily concerned with network operations; distributed staff in programs support local applications
Formal Project Management Office (PMO)	Essential for large-scale projects such as this one; uniform and industry accepted methodologies crucial	Important to follow project management best practice and industry-accepted methodologies; formal PMO certainly helpful but not required for well-disciplined organizations	Best to follow industry-accepted methodologies, but given the distributed nature of application development a formal PMO will likely have little impact or authority
Strategy	More centralized and more planned	Some centralized and some distributed elements, but careful agency-wide planning and coordination	Issue largely left up to programs with little central coordination

Enterprise-wide Integration Model Comparison: Organization			
Factor	**Centralized Model**	**Cooperative Model**	**Distributed Model**
Leverage Medicaid Relationship	More likely to occur due to central coordination and plan	More difficult due to distributed nature of management	Less likely to occur due to distributed nature of management
Data Sharing Laws and Policies	Must support data consolidation	Must at least support selective consolidation	Model can tolerate less permissive data sharing laws and policies

The second table focuses on a comparison of system features:

Enterprise-wide Integration Model Comparison: System Features			
Factor	**Centralized Model**	**Cooperative Model**	**Distributed Model**
Use of MPI	Strong feature; may be custom developed or off-the-shelf	Not mandated, but may be a feature with some system participation; may be custom developed or off-the-shelf	Likely not a feature unless implemented in the context of individual systems; may be custom developed or off-the-shelf
De-duplication Strategy	Embedded in MPI and its services	Either embedded in MPI and its services or offered as a separate service to participating systems	At best services available optionally to participating systems
Security	Easier to maintain as more resources are central and fewer interfaces exist	Somewhat more challenging as loosely coupled systems with more interfaces present additional opportunities for exposure	Security in the hands of distributed systems and their managers; training and capabilities will vary; standards are a useful start
System Acquisition Style	May be custom developed or off-the-shelf	Interface components tend to be custom developed	Interface components tend to be custom developed
Support for Analysis	Easier, as data tends to be more centralized and integration	Easier if an MPI is employed; SOA can enable this as well	More difficult, as data is more distributed and integration is loose

The third table presents a comparison of process-related attributes:

Enterprise-wide Integration Model Comparison: Process			
Factor	**Centralized Model**	**Cooperative Model**	**Distributed Model**
Agency Service Delivery	Supports a more centralized service delivery model best	Can support either a centralized or distributed service delivery model	Supports a more distributed service delivery model best

Enterprise-wide Integration Model Comparison: Process			
Factor	**Centralized Model**	**Cooperative Model**	**Distributed Model**
Technical Standards Enforcement	Easier, as more of the effort is performed centrally	Moderate, as compliance with standards is necessary to access shared services	Harder, as systems are largely stand-alone
System Requirements	More stable and clear	Moderately stable and clear	Less stable and often unclear, especially as it relates to system interoperability
System Development Coordination	Very coordinated due to strong central involvement	More independent but moderately coordinated due to some shared services	Largely independent and uncoordinated as most applications are stand-alone
Technical Innovation	Less interested	Moderately interested	Not very interested
Technical Risk	Fairly high, but so is the potential gain	Moderate	Fairly low, but the potential gains are more limited
Deployment Timetable	Incremental to a point, but requires critical mass to activate	Incremental, but still fairly coordinated	Incremental, little coordination
System Deployment Style	Tightly-coupled	Loosely-coupled; Service-oriented architecture (SOA) a useful strategy	Uncoupled; services and features usually replicated across systems
Cost	Higher up-front cost, though overall cost may be lower; software license cost may be lower due to centralized approach	Moderate up-front cost, but at the end of the day overall cost may be higher; software license cost may be minimized due to shared services	Costs distributed among participating systems; often hard to track and understand; software license costs likely higher, though coordinated purchasing may help minimize this

Integration to Interoperability

Introduction

As the years have transpired, the challenges of comprehensive enterprise-wide integration (as typified by the Centralized Model above) have become more apparent. While the technical implementations are themselves difficult, perhaps the hardest realization has been that, without clear alignment and integration of the *business processes* to be assisted by the integrated systems, the effort is almost always doomed to failure. And there have been some conspicuous failures both in the private and public sectors, and in almost every industry. One such failure was the abandonment by the Federal General Services Administration (GSA) of a new procurement system. Rather than ensure that business processes were examined and the system tailored to meet them, GSA relied on the *system* to promote changes in business practice. Four years and $77 million later GSA finally abandoned the new system and reverted to the set of legacy systems it had targeted to replace.[10]

Interoperability

Much has happened in the healthcare information technology world in the past several years. While many organizations continue to integrate their systems, the challenges they have faced have caused some to refocus the issue: rather than focus on healthcare systems integration, these organizations are focusing on healthcare systems *interoperability*. But just what is interoperability? In early 2007 the Health Level 7 (HL7) Electronic Health Record (EHR) Interoperability Work Group published a white paper[11] whose purpose was to consider the meaning of interoperability, develop a consensus definition, and discuss implications for future standards work. After reviewing and analyzing more than 100 definitions, the work group agreed upon a three-part definition:

1. **Technical interoperability** focuses on the physical transmission and receipt of health data, its transport between participating systems. Much of the work here is on message formats and reliable, secure message transport.
2. **Semantic interoperability** focuses on ensuring shared meaning between sending and receiving partners – ensuring that the meaning of what was sent is consistent with the understanding of what was received. Much of the work in this area is focused on medical terminology which can be referenced consistently by all parties.
3. **Process interoperability** focuses on higher-order workflow concepts that make data sharing a richer and more valuable experience. Work in this area tries to understand how shared health data supports the specific activities and workflow of the organizations that use it and the integration of health data into the work setting. Issues of data usability and timeliness are examples of process interoperability concerns.

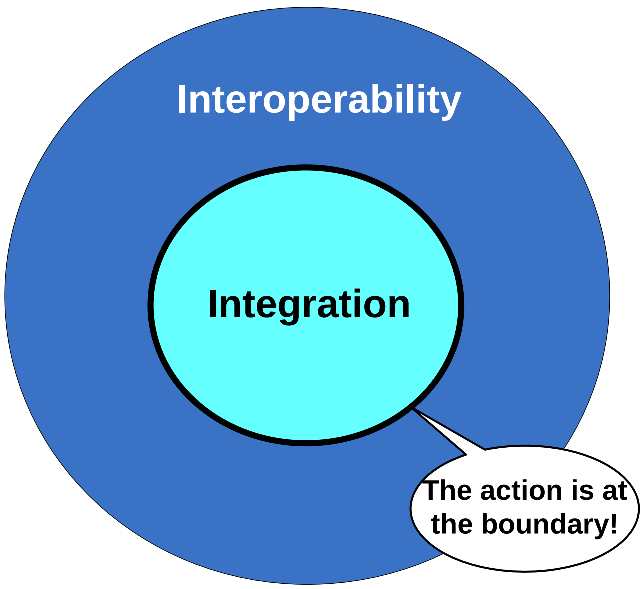

Figure 12 - Integration vs Interoperability

The HL7 EHR Interoperability Work Group went on to define an Interoperability Model Draft Standard for Trial Use[12] (DSTU) which defines the characteristics that records need to have to meet these three levels of interoperability.

---

[10] See "After $77m, GSA finally pulls the plug," *Government Computer News,* January 23, 2006. < http://www.gcn.com/print/25_2/38044-1.html>

[11] Health Level 7 EHR Interoperability Work Group, *Coming to Terms: Scoping Interoperability for Health Care,* February 2007. < http://www.hln.com/assets/pdf/Coming-to-Terms-February-2007.pdf>

[12] See http://www.hl7.org/ehr/downloads/index_2007.asp

The prospect of systems talking to one another in a more loosely-couple manner, rather than being absorbed into one another, may be a more feasible strategy for many agencies. For this reason the Cooperative Model above seems most reasonable as a middle-of-the-course option. In reality, most agencies will find themselves over the next few years pursing a mixed strategy, one of selective integration and selective interoperability. The important decision points will be determining which approach to take: decisions will lean toward integration when more of the attributes of the Centralized Model seem strong, and decisions will lean toward interoperability when more of the attributes of the Cooperative Model seem strong. The action is at the boundary between the two, when the decision criteria less clear cut based on the particular situation (Figure 12).

One additional phenomenon is fueling the interest in integration, especially for agency systems that interact with other systems outside the agency's walls: the health information exchange network (HIEN) typically operated by a regional health information organization (RHIO). These collaborative organizations focus on health data exchange on a community, county, or even state-wide basis (Figure 13). Less formal health information exchanges can develop within complex organizations, like integrated health networks or hospital systems. They have a wide and varied set of participants (providers, laboratories, hospitals, health plans, public health agencies, pharmacies, even patients/citizens). Primarily driven by private sector participants, public health is often involved in their formation and operation. While the major emphasis is on exchanging clinical data to support patient care, some health data exchanges do focus on health services data instead or in addition to their clinical needs (as is the case with the Utah Health Information Network, or UHIN[13]).

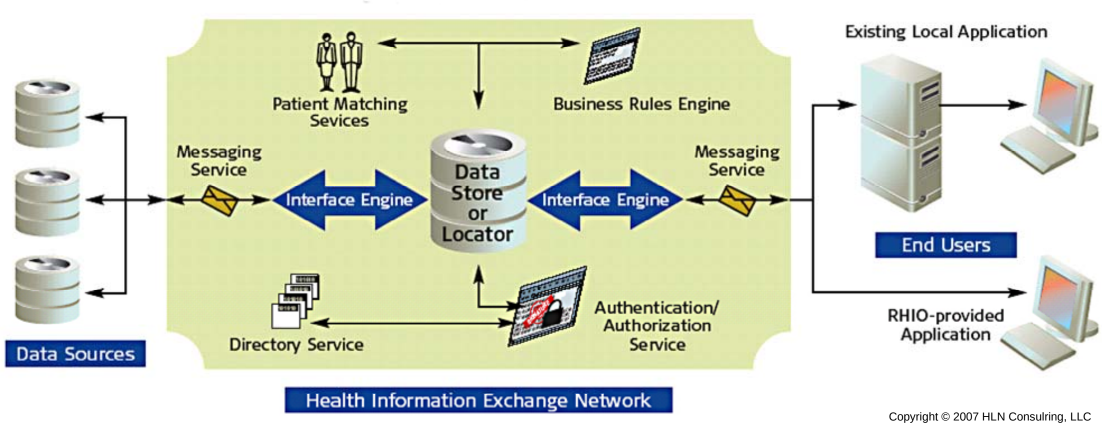

Figure 13 – Health Information Exchange Network Architecture

Momentum is growing nationally as well as in many local jurisdictions to bring some needed coherence to health information technology initiatives as well as to the entire healthcare system writ large. This effort largely coalesced as a result of President George W. Bush's 2004 State of the Union address called for the majority of Americans to have interoperable electronic health records within ten years. By executive order he formed the Office of the National Coordinator for Health Information Technology (ONC) within the Department of Health and Human Services. Shortly thereafter, the progress report issued by ONC laid the groundwork for much of what has transpired since at the Federal level: a senior level policy advisory board for the HHS Secretary (the American Health Information Community,

[13] See http://www.uhin.com/

shortly to transition to private, non-government status), work on developing technical interoperability standards to enable reliable health information exchange (Health Information Technology Standards Panel, or HITSP), four prototype projects for demonstration of HIEN concepts, and a thirty-four state and territory collaborative focusing on issues of health information privacy and security (Health Information Security and Privacy Collaboration, or HISPC). Additional programs, grants, and contract opportunities were created by the Agency for Health Care Research and Quality (AHRQ) and the Health Resources Services Administration (HRSA) within HHS.

Efforts have picked up steam in other circles as well. The Markle Foundation and the Robert Wood Johnson Foundation established the eHealth Initiative (eHI). The Certification Commission for Healthcare Information Technology (CCHIT) began certifying ambulatory electronic health record systems as being compliant with HL7 Electronic Health Record functionality. The National Governors' Association is taking a prominent role through their State Alliance for eHealth. The National Conference of State Legislatures HIT Champions[14] (HITCh) partnership provides state legislatures with information and technical assistance on important political and technical issues. Medicaid has become the anchor of HIT/HIE activities in many states and is playing a larger role in public health by focusing on disease management. All prominent medical professional societies have ongoing task forces and initiatives focused on HIT (AMA, AFPA, AAP), as do all the prominent public health professional associations (AHIMA, NACCHO, ASTHO, APHA, PHDSC) who got together in San Diego in April 2007 to try an forge a common voice for public health in issues related to health information exchange.

The advent of the HIEN can have a profound affect on the systems that agencies want to deploy to their external users. Public health applications targeted at users in provider settings may have slower adoption rates as organizations encourage (or require) users to stay with institutionally-supported applications. This is especially true in hospital and large ambulatory care settings, but this phenomenon also appears in local health departments that deploy more comprehensive services automation systems. Pressure will build for providers to interoperate *solely* through HIENs. This may affect public health data exchange partnerships as providers (or other partners) may be required to exchange data with the RHIO and may not want to exchange specific data with a public health program as well. Though many RHIOs are just beginning to focus on clinical data exchange, and public health programs are typically not among their early pilots, with sufficient momentum RHIOs will likely become *the* driving force and context for health information exchange in their jurisdictions. State Medicaid programs may be a better fit for some of these early RHIO implementations as they usually possess large warehouses of both clinical and claims data and may be an attractive target to help "jump start" and HIEN.

Richly functional public health systems run the risk of becoming used primarily as data repositories as external users lose access to more advanced features. For instance, chronic disease registries contain disease pathways that define special prevention or treatment protocols typically not found in an EHR system that might be deployed at a provider's office. If providers are prevented from accessing the chronic care registry directly, they stand to lose access to these features until their local systems can provide them. In the case of an immunization information system, which may also function as a repository, providers could lose access to vaccine algorithms, reminder/recall notice functions, and practice-level coverage assessment, which also are not typically found in their local systems. As they look to improve the functionality of their information systems in the future, public health needs to consider the best way to continue to offer these services and reach the largest number of providers effectively.

[14] See http://www.hitchampions.org/

But there is also great potential in these developments, as public health agencies have much to gain from their involvement with HIENs. Over time, public health may gain more and better quality access to data by participating in HIEN activities. The potential for leverage of technical and programmatic energy not only within the agency but within partner organizations in the community cannot be overlooked. Public health agencies also have a lot to offer: existing data sets that may be consolidated and population-based; expertise in record de-duplication, database management, web applications, and standards-based data exchange (like HL7) that may be lacking in the start-up HIEN; existing relationships with many of the key players in the community like hospital, private providers, payers, laboratories and other ancillary services, and professional associations; governance experience, including experience negotiating data sharing agreements and memoranda of understanding. Finally, issues surrounding population health are paramount on the national agenda related to health information exchange.[15]

While this is new territory for public health, there is a clear synergy between national developments and core public health requirements:

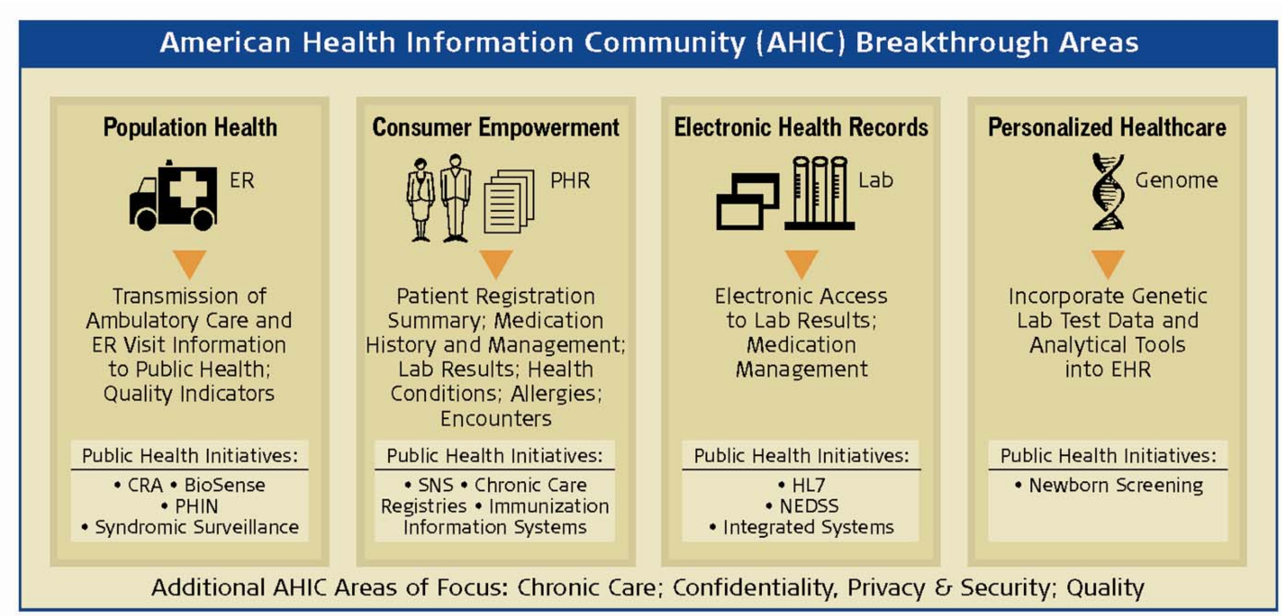

Figure 14 – Alignment of AHIC and Public Health

In September 2005, Health and Human Services launched the American Health Information Community ("The Community") to serve as a senior advisory committee to Secretary Mike Leavitt on the establishment of electronic health records for all Americans. In August 2006, the Community refocused its initial three "breakthrough areas" to guide progress over the next few years. The three breakthrough areas were further refined and are well aligned with current public health information technology projects, including the Public Health Information Network (PHIN), bio-terrorism and emergency preparedness, and Countermeasure and Response Administration (CRA). Public health

[15] See American Health Information Community (AHIC), <http://www.hhs.gov/healthit/community/background/>

projects have begun to see themselves as part of this larger framework, and the Federal government has begun to include public health in its HIEN planning and funding.[16]

Public Health Informatics Function

Public health agencies are usually quite dependent on organizations and individuals *external* to themselves to provide data that is needed for critical public health functions. Depending on the profile of applications within an agency, and most importantly on the sources and destinations of data managed by the agency, system integration and system interoperability decisions need to be made hand-in-hand. Whereas in the past, public health agencies had few standards defined for interoperability with other data systems, today national and even international standards are quickly developing to guide this functionality. Public health agencies must work within these emerging standards to ensure the least burden on its data partners and the most potential leverage of its own activities.

One of the imperatives resulting from this national direction is the need for public health agencies to organize an informatics focus in the agency to engage in and support local, regional and national initiatives. The informatics function needs to be separate from the operational responsibilities of IT as it is meant to complement and not compete with IT operations for resources and attention. Usually, informatics is defined as the use of information technology to understand and process data at up to three levels: the molecular level (bioinformatics), the patient level (medical informatics), and the population level (public health informatics).[17] But here, by "informatics" we mean the education, research, and development of a set of capacities within an agency to effectively make strategic decisions about information technology as it relates to public health. Part of that capacity is the understanding of the place that public health informatics fits within the larger health informatics world and the relevance of established and emerging standards on public health information technology.

The informatics function should report to a senior agency official and should, where possible, develop links to similar interests in academic informatics programs and appropriate national associations, standards development organizations, and informatics organizations (*e.g.,* the American Medical Informatics Association, Health Level 7, Healthcare Information and Management Systems Society). The relationship that the Utah Department of Health has with the University of Utah to develop and operate a public health informatics training site funded by the Robert Wood Johnson Foundation will contribute to the realization of this goal in that state. Similar initiatives can be found among the other RWJ grantees in this program (Columbia University, the Johns Hopkins University School of Medicine, and the University of Washington), as well as with some other agencies who have recognized this as a priority.[18] [19] In addition, the *Common Ground* initiative of the Robert Wood Johnson Foundation granted awards to state and local health departments to develop public health informatics capacity.[20]

---

[16] See CDC funding opportunity, "Accelerating Public Health Situational Awareness Through Health Information Exchanges" <http://www.fbo.gov/spg/HHS/CDCP/PGOA/ToBeDetermined/SynopsisP.html>

[17] See http://www.jhsph.edu/publichealthnews/articles/2007/healthinformatics.html

[18] See Press Release, "$3.68 Million Grant to Boost Public Health "Informatics"", June 7, 2005.
< http://www.rwjf.org/newsroom/newsreleasesdetail.jsp?id=10353&gsa=1> and <
http://www.rwjf.org/programareas/grant.jsp?id=52098&pid=1141&gsa=1>

[19] See Minnesota Center for Health Informatics and the Minnesota Public Health Information Network
< http://www.health.state.mn.us/e-health/mnphin/index.html>

[20] See http://www.rwjf.org/applications/solicited/npo.jsp?FUND_ID=55004

A number of important initiatives need to be monitored and considered as an agency determines its direction:

The Health Information Technology Standards Panel (HITSP)[21]**,** a public-private partnership funded by a Health and Human Services contract to develop interoperability standards for local, regional, and national health information exchanges. Because so many standards development organizations are involved in this effort, HITSP standards serve as a useful starting point.

The **Consolidated Health Initiative** (CHI)[22], whose objective is to enable sharing of health information between various Federal agencies by adopting existing standards.

Medicaid Information Technology Architecture (MITA)[23], which is intended to promote integrated business and IT across the Medicaid enterprise to improve the administration of the Medicaid program.

Public Health Information Network (PHIN)[24], a maturing national standard. PHIN is CDC's vision for organizing, standardizing, and managing the collection and dissemination of public health information. It requires the use of fully interoperable information systems in the many organizations that participate in public health. PHIN requires policy, technology, and vocabulary standards for interoperability between public health agencies, CDC, private health entities, and other national, state, and local organizations.

Agency and/or Jurisdiction Office of the Chief Information Officer (or equivalent) Policy and Procedures[25] serve as a useful reference in certain standards areas, and may contain requirements imposed by an agency or the larger jurisdiction of which it is a part.

Conclusion

For many agencies, the question is which strategy will likely lead them to succeed, and which strategies should they avoid as they might lead to failure. This way of approaching the key questions raised in this paper may be too simplistic to be useful. Success and failure are relative terms: relative to the investments, relevant to the risks, and relevant to the *potential* impact of the choices that are made. The passage of time also affects how organizations remember, evaluate and understand their decisions. Three models for enterprise-wide integration that were offered in this paper vary in the degree to which participation in them is mandatory or voluntary (Figure 15 below).

This governance model can change over time as the executive sponsorship, funding, and information technology capabilities of the agency change. It is up to each agency to assess its capabilities, its environment, and its political will as it plans for systems integration and interoperability. Several important rules of thumb are:

---

21 See Health Information Technology Standards Panel website, <http://www.hitsp.org/>.
22 See Consolidate Health Initiative website, <http://www.hhs.gov/healthit/chiinitiative.html>.
23 See Medicaid Information Technology Architecture website, <http://www.cms.hhs.gov/MedicaidInfoTechArch>
24 See Public Health Information Network website, <http://www.cdc.gov/PHIN>.
25 See http://its.utah.gov/policiesstandards/policiesstandards.htm

- Understand the relationship between integration/interoperability goals and the underlying strategic goals of the agency with respect to service delivery and information analysis.

- Consider the level of executive support as well as the overall political climate. The cycles of change promulgated by routine electoral change can have a huge impact on large, multi-year projects.

- Build an informatics capacity distinct from the operational information technology responsibility to help the agency keep up with national developments and to best advise its leadership and programs on informatics issues.

- Look outside of the agency to understand what is developing in the community with respect to health information technology and health information exchange in particular. Advocate for the agency to get a seat at the table as regional plans develop and are implemented.

- Develop plans, but try to be flexible to adjust to changing circumstances – technical, organizational, regulatory, financial, and political. Use standards to help form a more stable foundation and to help inform agencies when it's best to stay the course.

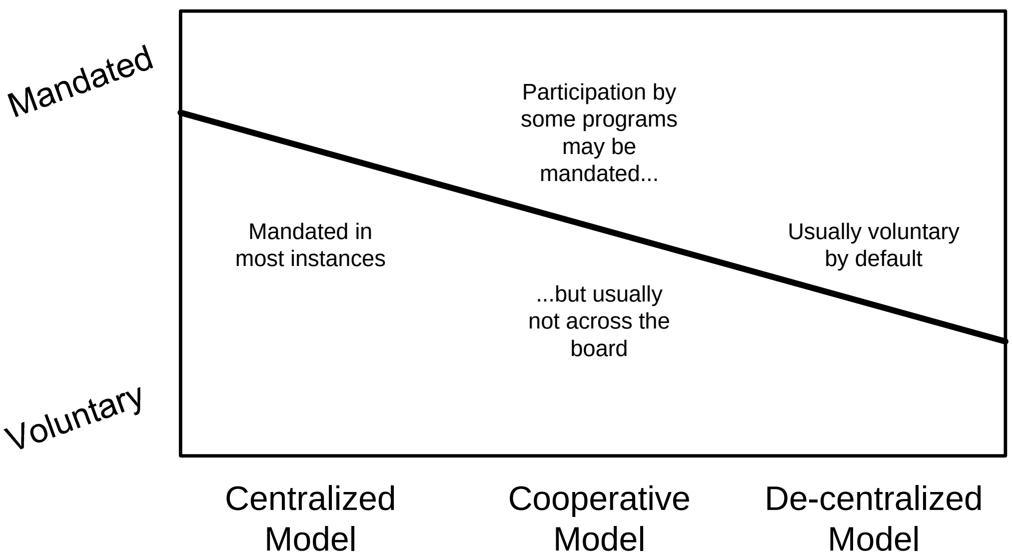

Figure 15 – Governance Approach to Integration Models

Appendix: Key Technical Challenges

As agencies make their plans for systems integration and interoperability, several key technical challenges need to be addressed. This section will discuss two key challenges: master patient index (MPI) deployment, and data-centered versus document-centered approaches to data storage and interoperability.

Master Patient Index Deployment

One of the most critical issues within a healthcare systems environment is the strategy for matching and de-duplicating person records between systems. Typically, a Master Patient Index, or MPI, is created or acquired to match records or de-duplicate sets of records that appear to have the same patient represented more than once. Systems can implement matching on the front-end, that is, when new records are being integrated into a database and one wants to determine whether the person represented by the new record is already represented in the database, or on the back-end, that is, when one wants to examine an existing database or data set and determine if it already has duplicated records within it.

There are many software products available to perform provide these services, but software acquisition is only part of the puzzle. It is estimated that only a third of the total cost of deploying an MPI pays for the software itself (Figure A1). Another third of the cost pays for the development of an appropriate architecture within which the MPI will function, as well as the integration of the MPI into the other components of the agency's systems. The final third pays for configuration and testing—by agency staff or the MPI vendor—to ensure that the software works properly with the agency's data and systems.

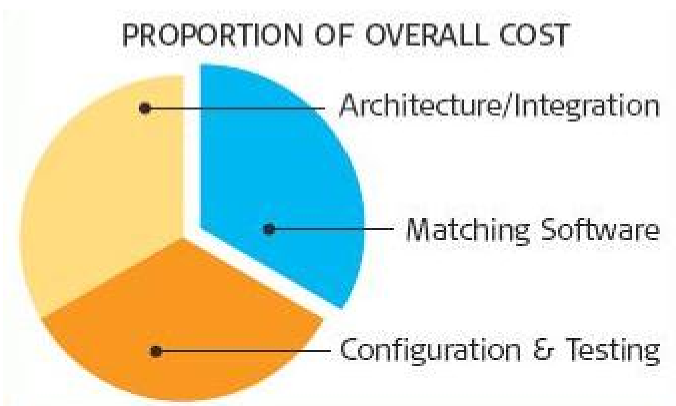

Figure A1 – Cost of Deploying an MPI

The core of any de-duplication and matching service is the type of algorithm that is used to determine if two records are the same or different. There are two basic types of matching strategies:

1. **Deterministic Matching** uses sets of predetermined rules to guide the matching process. The rules rely on a series of exact matches between data elements to identify when records match. It is most successful when the data is of relatively high quality or is dominated by reliable unique identifiers for records. Deterministic matching is less successful when the data is incomplete or inaccurate, when there are many spelling or transcription errors, or lots of inconsistencies (e.g., frequent name changes).

2. **Probabilistic Matching** is a process whereby an estimate is made of the probability that two records are for the same person based on the degree to which certain fields in the two records match. Two thresholds are then set:

   - All record pairs whose probability is above the higher threshold are considered to be matches.

- All record pairs whose probability is below the lower threshold are considered not to be matches.

The disposition of record pairs whose probability falls in between the two thresholds is considered to be uncertain and they require additional review, likely by a trained staff member.

In 2006, the Public Health Informatics Institute released its Unique Records Portfolio, a detailed guide for public health agencies to use to develop a strategy for person record matching and de-duplication.[26] It contains not only a theoretical background in the key concepts, but practical checklists for action as well as case studies describing successful efforts. In the spring of 2006 the Washington State Department of Health used the Portfolio as the basis for a two-day training seminar for agency staff to teach them about person record matching and de-duplication issues in anticipation of a broader agency-wide effort to implement a more consistent strategy to tackle this issue.

Data-centered or Document-centered?

There are two predominant styles of data storage for public health data repositories. In a *data-centered* approach, systems store data in a conventional relational database (RDBMS) with tables for different entities and rows for instances of those entities. Structured Query Language (SQL) is used to send queries to the database and extract the rows from the appropriate tables that meet specified criteria (Figure A2).

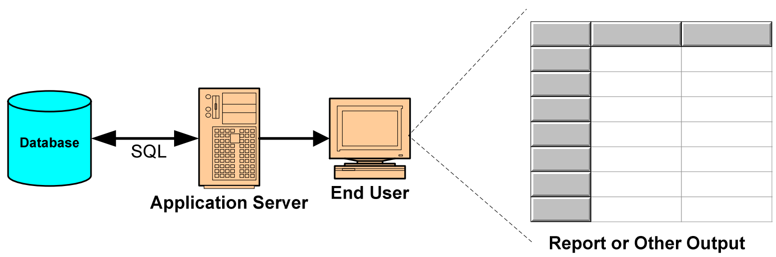

Figure A2 – Data-centered Storage Approach

In a *document-centered* approach, data is stored in a formatted document within the repository for retrieval as a unit. The contents of the document cannot usually be searched directly; a smaller amount of metadata is saved in a database at the time the document is stored or updated which contains key information useful in selecting the document for retrieval (Figure A3). Documents can be stored in many formats, including simple image files, Adobe Acrobat format, or more sophisticated XML documents.

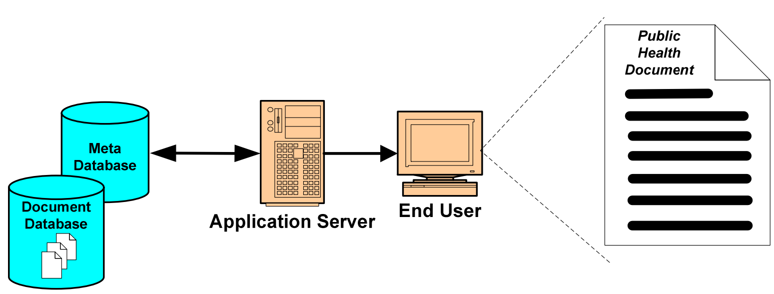

Figure A3 – Document-centered Storage Approach

[26] See http://www.phii.org/pages/Portfolio-preOrder.html

Similarly, there are two predominant styles of data interoperability *between* public health systems. A *data-centered* approach uses traditional structures to represent the data set being transported. In simple cases, this means a row in a file for a record, and either delimited or fixed length fields within the record. Metadata describing the structure of the file may be a simple header row before the data or a separate file containing more detailed field descriptions, code sets, or semantic explanations. More sophisticated examples include HL7 or X12 messages which follow a well-developed, standards-based syntax detailed in implementation guides or profiles (see Figure A4).

```
MSH|^~\&||||||VXU^V04|19970522MA53|P|2.3.1|
PID|||221345671^^^^SS||KENNEDY^JOHN^FITZGERALD^JR|BOUVIER^^^^^^M|19900607|M|||~^^^^MA^^^BDL|
NK1|1|KENNEDY^JACQUELINE^LEE|MTH^MOTHER^HL70063|
RXA|0|1|19900607|19900607|08^HEPB-PEDIATRIC/ADOLESCENT^CVX|.5|ML^^ISO+||||||||
MRK12345||MSD^MERCK^MVX|
```

Figure A4 – Sample HL7 V2.n Message

The second style of data interoperability is *document-centered*. In this case, the data is pre-arranged in a document format which is usually quite structured. Simply opening up and examining the document itself conveys its contents in an organized, labeled fashion. The best example of this approach in the clinical world is ASTM's Continuity of Care Record[27] (CCR), which contains a pre-determined set of data in a pre-determined format (Figure A5). HL7 has created a more generic architecture for creating data in this style (Clinical Document Architecture, or CDA), and the two organizations have combined the two by developing an implementation of the CCR using CDA technology called the Continuity of Care Document (CCD). Both CCR and CCD represent *summaries* of clinical information about a specific patient.

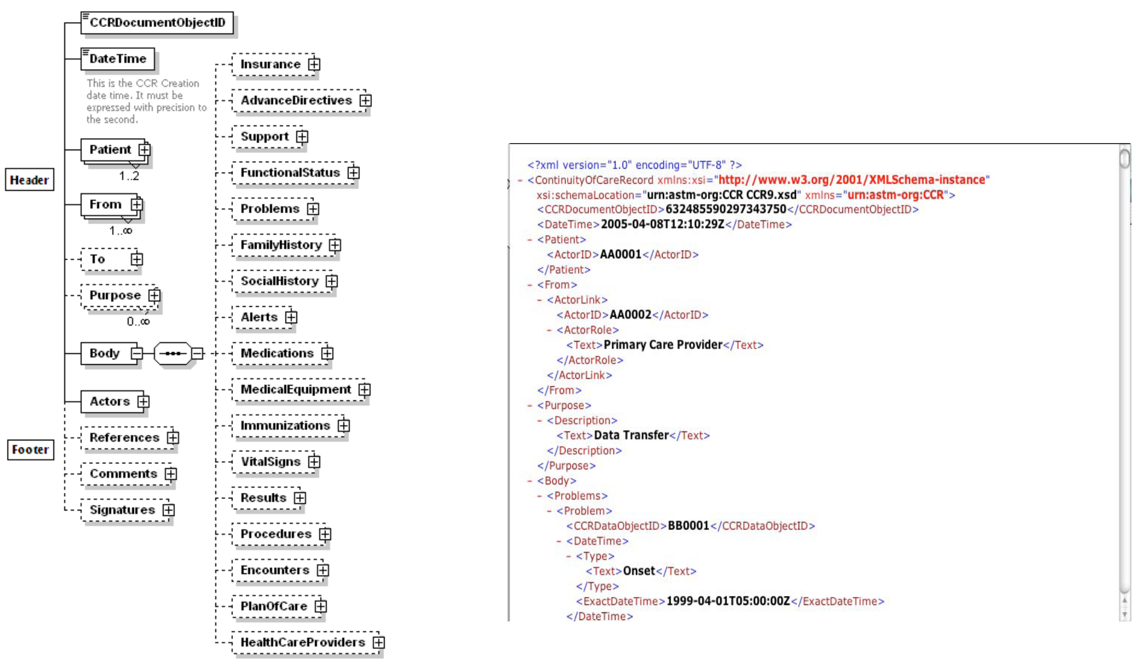

Figure A5 – Make-up of CCR Record[28] and Sample XML Representation[29]

---

[27] http://www.astm.org/cgi-bin/SoftCart.exe/DATABASE.CART/REDLINE_PAGES/E2369.htm ?L+mystore+yirv0526+1174970569

[28] Modified from diagram found at http://www.astm.org/COMMIT/E31_CCR0305.ppt

Since data exchange partners do not always control the attributes of the other's system, the style of data storage and the style of data interoperability may not be the same for a particular interaction. However, certain combinations present particular benefits or drawbacks:

Data Storage	Data-centered	Document-centered
Document-centered	May be difficult to extract discreet data from documents and assemble into the desired message or file format. Receiving data-centered messages and storing them in the databases as documents is less challenging.	Relatively easy to extract documents, transport them as such, and store them as documents in the destination system.
Data-centered	Relatively easy to extract data and assemble in the desired message or file format. Interface engines exist which facilitate parsing data from databases into messages and vice versa.	Relatively easy to extract data and assemble in the desired document format. May prove more challenging to parse documents back into discreet data elements for storage in the destination system, depending on the form of document used.

Data Interoperability

When data is made to interoperate in its native storage style (data-centered or document-centered) there are fewer challenges (lower left quadrant; upper right quadrant). When these strategies do not match, it may be more challenging on one end of the transaction or the other (upper left quadrant; lower right quadrant).

Health Level Seven's (HL7) new Electronic Health Record Interoperability Model[30] (EHR IM) can help a project understand the issues related to system-to-system data exchange by allowing the project to develop an interoperability *profile*. This profile selects a subset of the elements of the IM for implementation – systems that agree to implement their data exchange using the same profile should be interoperable with each other. In addition, a group of key health information technology organizations and industry vendors have gotten together and formed Integrating the Health Enterprise (IHE), an initiative focused on improving healthcare data interoperability through the development of precise implementation guides, or profiles, for major health data exchange transactions.[31]

While the examples above are focused on clinical systems, they are quite relevant for public health data exchange as well. For data that is coming into public health agencies from outside the predominant source is some sort of clinical system – in a provider's office for a communicable disease report, a

[29] See http://www.centerforhit.org/PreBuilt/chit_ccrnyc.pdf
[30] See http://www.hl7.org/ehr/downloads/index_2007.asp
[31] See http://www.ihe/net/

hospital for a discharge summary from which syndromic surveillance data mining is done, a laboratory for test results from a patient or the environment. This data may be data-centered or document-centered in its origin depending on the originating system. Likely there will be a mix as different data partners deploy different types of systems. The majority of the data originating from non-clinical sources within the agency is likely to be data-centered. It is important for agencies to be aware of these differences – and to follow the national standards that are developing in this area – to help inform the agency integration and interoperability strategies.