

Strategies for Person Data Matching and De-duplication

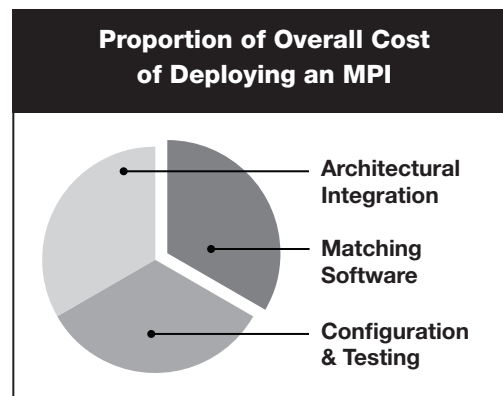
Exploring One of the Most Critical Issues Within a Healthcare Environment

ONE OF THE MOST CRITICAL ISSUES within a healthcare systems environment is the strategy for matching and de-duplicating person records between systems. This is especially true for organizations that find themselves with patient data scattered around various systems, either because they were acquired or built over time, or because they represent “best of breed” solutions deployed in their respective subject areas. Even an organization which opted to go for a more streamlined integrated system solution likely still has many interfaces to ancillary and secondary systems, even if they are outside of the enterprise. 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.

Deploying an MPI can be a costly venture. The decision deserves careful consideration as the outcome will likely need to serve the organization for many years. But the cost of deploying an MPI extends well beyond the price of the software. Only a third of the cost is typically for the matching software itself. Another third of the cost will likely go to configuration and testing (some more sophisticated products require “training” based on actual matching decisions). The final third goes to developing the architecture that will be employed and planning the integration of

the matching software into existing systems. Matching/de-duplication products likely will require custom programming to integrate the software into existing systems, or to provide interfaces (batch, queuing, or interactive) to necessary users and processes.

It is against this backdrop that the following four models are offered. The models are fairly general, and do not represent product-specific implementations. Yet the functionality that can be achieved under any of these options may differ if different matching/de-duplication products are used.



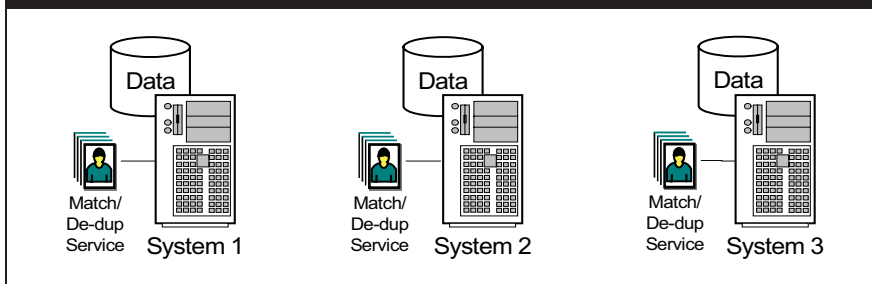
MODEL 1: DISTRIBUTED SERVICES

This simplest of models may already exist in your organization, as it can be the result of not having a strategy at all. Under this model, individual systems deploy record matching/de-duplication services locally within their systems (see Figure 2). Standardization on one product or a small number of products would allow the organization to develop better expertise in the selected tool(s) and leverage investment more wisely.

The benefit of this model is that it conforms to the reality of independent systems development. It leverages software investments if fewer products are supported, and does not create system dependencies across programs or departments. This strategy allows an organization to pilot new technologies in this arena more quickly and leverage existing successful system deployments. Most important, this strategy allows each project/system to tune matching/de-duplication service for its particular set of person records since optimal matching rules might not be uniform across all applications (for example, matching children’s records accurately often involved different data than matching adults). Batch and interactive real-time matching are both possible (depending on product capabilities), and overall technical support costs at the center of the organization are lower if a single product is leveraged.

On the other hand, individual systems may continue to operate within their traditional silos where sharing across projects is limited and requires significant effort. Software acquisition cost is higher than some other options since this strategy requires multiple software licenses across the organization. Overall support cost for

Model 1: Distributed Services



an organization as a whole may be higher, and the strategy does not promote system integration (maintains the status quo).

MODEL 2: CENTRALIZED SERVICES

Under this model, a central organizational authority provides a network service to perform matching and de-duplication services to systems that choose to invoke it (see Figure 3). Individual systems do not need to maintain or operate this service themselves. Through a standard interface, participating systems pass data to this service which processes each transaction and returns a standard result. Depending on the product selected, service requests may only be able to be satisfied one at a time, with or without queuing. Interactive, real-time matching may or may not be possible depending on product selection. The central server under this model has no persistent data storage of its own:

the service merely processes requests and returns results to the originating system and does not retain any information about the person records processed.

This option maintains some level of independent systems development, and leverages knowledge and software investments fully. It allows the organization to deploy matching/de-duplication services to new systems more quickly, and allows migration to a master person index at some point in the future more readily. This model brings lower software acquisition cost through a shared license and promotes development of specialized knowledge on matching/de-duplication issues that can be used organization-wide. Overall support costs for the organization may be lower depending on the solutions deployed.

On the other hand, this solution creates a systems dependency on the central matching/de-duplication service that may

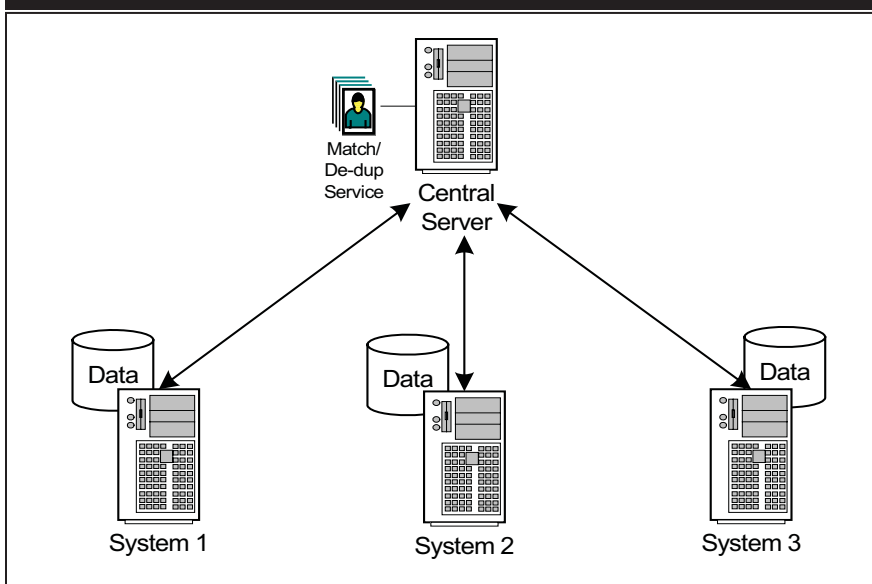
constrain or affect program operations. Current systems within departments may now need to be modified to conform to an organization-wide service offering. Significant technical support costs now exist at the center of the organization. Each system may need to compromise on tuning matching/de-duplication services for its particular set of person records. This option does not especially promote system integration, and interactive, real-time matching may not be possible with some products.

MODEL 3: MASTER PATIENT INDEX/RECORD LOCATOR SERVICE

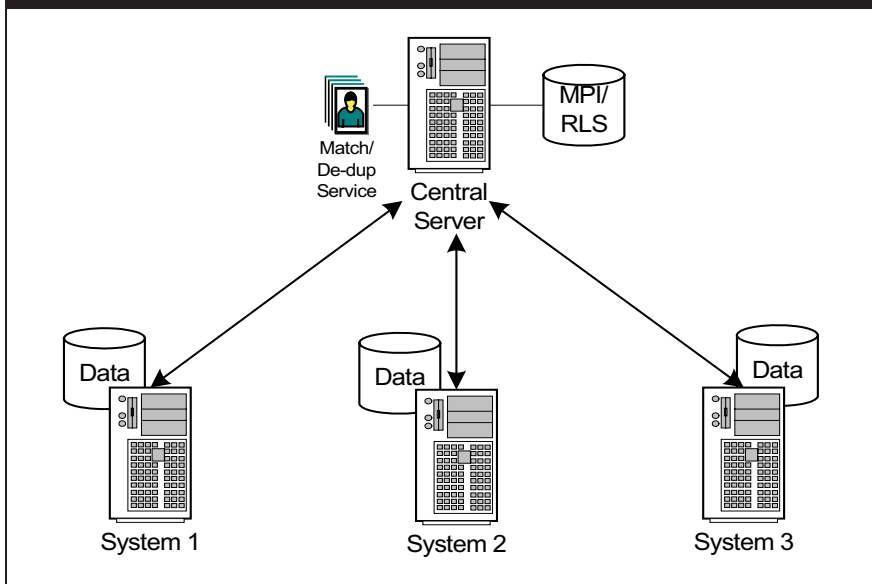
Like Model 2 above, a central organization authority provides a network service to perform matching and de-duplication services to participating systems. But in this case, participating systems register their person records with a central Master Person Index (MPI) that provides a de-duplicated set of person records available for matching to local system databases (see Figure 4). The MPI serves as a record locator service (RLS) in that it knows, for a given person record, where that record is found in participating systems from across the organization. Participating systems can use this RLS to find data related to persons in their systems and can integrate data across systems if desired. The central server in this model now maintains a persistent database of people in the MPI, but its database does not contain system-specific health data. It only contains a collection of person-identifying data to facilitate matching and person record location.

This model leverages knowledge and software investments fully. It promotes data integration which increases the organization's ability to relate person-centric data from different departments together. The model promotes application integration by making it easier for systems to more reliably share data about the same people. Batch and interactive real-time matching is possible (depending on product capabilities), and software acquisition cost should be lower through shared licensing. This model promotes development of specialized knowledge on matching/de-duplication issues organization-wide and may better enable interoperability with health information exchanges

Model 2: Centralized Services



Model 3: Master Patient Index/Record Locator Service



(HIEs)/regional health information organizations (RHIOs).

On the other hand, systems development within departments is less independent than with other options. This model creates a systems dependency on the central MPI/RLS that may constrain or affect program operations. Current systems within departments will need to be modified to conform to this organization-wide service offering and it may require significant effort to bring a new system into MPI participation. This model entails significant technical support costs at the center of the organization as well as additional central management beyond strictly technical services as central services become more sophisticated. As with Model 2, each participating system may need to compromise on tuning matching/de-duplication service for its particular set of person records.

MODEL 4: DATA WAREHOUSE OR INTEGRATED SYSTEM

In this model, the MPI/RLS from the previous model is supplemented by a full data warehouse containing consolidated records from all participating systems which include program-specific data (see Figure 5). An alternative is the deployment of a fully-integrated system that replaces the distributed systems entirely. A hybrid model is possible where only

some participating systems contribute data to the central repository (or allow themselves to be integrated into a central system) while others continue to receive MPI/RLS services.

This model shares almost all of the benefits and limitations of the MPI/RLS model above, but goes even further in achieving data integration through its data warehousing features. Projects that once may have been independent now may become

even more dependent on central services than before. There is likely also significant additional cost involved in creation and management of data warehouse.

WHERE DO WE GO FROM HERE

The culture and system strategy of an organization go a long way in determining which approach to de-duplication/record matching services represents a good fit. Organizations whose systems are more distributed among programs or departments have more to gain from thinking about these different strategies. Organizations with a more centralized approach to system deployment likely already have struggled with these issues in the course of their system integration planning. Interestingly enough, these same models can be applied to health information exchange/regional health information organization planning with many of the same considerations. Depending on how loosely-coupled or tightly-coupled an HIE project wants to be can help determine what strategy they employ for record matching and de-duplication. **JHIM**

Noam H. Arzt, PhD, is president and founder of HLN Consulting, LLC, San Diego, and does consulting in healthcare systems integration, especially in public health. He can be reached at arzt@hln.com.

Model 4: Data Warehouse or Integrated System

