MASTER DATA MANAGEMENT WITHIN HIE INFRASTRUCTURES:

A FOCUS ON MASTER PATIENT INDEXING APPROACHES

September 30, 2012
Prepared for the Office of the National Coordinator for Health Information Technology by:

Ben Purkis, Director, MDM Practice
Genevieve Morris, Senior Associate
Scott Afzal, Principal
Mrinal Bhasker, Principal
David Finney, Principal

Table of Contents

Disclaimer

This report was created by Audacious Inquiry, LLC under a contract with the Office of the National Coordinator for Health Information Technology (ONC). The content, views, and opinions do not necessarily reflect those of the Department of Health and Human Services or ONC.

Executive Summary

Having the right patient data, at the right place, at the right time is the goal of health information exchange (HIE). This starts with accurately capturing and coordinating a patient's identity across multiple disparate organizations. If the information presented at the point of care is matched with the wrong patient, it is not only unusable, it is also dangerous for the patient. Delivering the right patient information is crucial to realizing the benefits of HIE. In the absence of a unique national identification number or some other unified way of identifying people and organizations, master data management (MDM), much science, and a bit of art, makes this important work possible.

MDM and master patient indexing (MPI) have developed over the last twenty years to offer organizations from banks to large retailers to health organizations a more consistent understanding of their customers' identities and activities across diffuse networks and disparate systems. MDM solutions, which are integrated with other mission-critical systems, typically utilize two approaches to link peoples' identities across multiple silos of data. Deterministic matching approaches attempt to line up different pieces of demographic information, such as last names or Social Security numbers, across source systems to look for exact matches. Probabilistic matching approaches, which are more sophisticated, attempt to deal in a more nuanced way with the inevitably error-filled, unstable nature of identifying information in source systems. Hybrid approaches may also be used.

With a well-implemented MDM toolset, health information organizations (HIOs) can maintain a relatively high degree of confidence that patient identity information is consistent, disambiguated, and de-duplicated, even across a large number of source systems or as the health data itself remains federated. While directed exchange—the first phase of many HIOs' implementation plans—does not necessarily require a sophisticated MPI, as HIOs develop to more advanced services, an MPI will be necessary. Query-based exchange relies on an MPI to work in coordination with a record locator service to pull patient records from various organizations and return the results to a provider querying the HIO. Without the MPI that can resolve identities across these organizations, the query functionality will not work.

Moving past query exchange, advanced services such as provider notifications and hospital readmission reports will be supported by an MPI that can attribute a patient to a provider. Additionally, analytics for programs like accountable care organizations (ACOs), patient centered medical homes (PCMH), and other value-based purchasing models will require that patient identities are accurately maintained as they move across the continuum of care. If HIOs plan to support these types of initiatives, they will need an MPI and master data management processes to maintain patient identities.

It is critical that HIO leaders possess a strong understanding of MDM and MPI as they develop long-term plans and identify services that solidify their position of value in a service area. Looking beyond directed and even query-based exchange, becoming the trusted arbiter of patient, provider and healthcare organization identity information for a state or region is a powerful role for an HIO. The purpose of this report is to offer these HIO leaders a primer on the key issues related to MDM. In addition, Appendix C includes vendor supplied descriptions of their MPI products.

Core Concepts in Master Patient Indexing

Matching Approaches

There are a variety of different approaches that can be used in a master patient index (MPI) to address matching the identities of individual patients that are scattered across many disparate care settings. These approaches to patient identity management can rely on the use of a unique patient identifier, a voluntary patient identifier, patient biometrics, or an algorithmic matching approach. Each of these approaches has pros and cons; however, consumer rights concerns, financial requirements, politics, and other influencing factors have driven the U.S. healthcare system and data exchange initiatives towards an algorithmic-based set of solutions for cross-system and inter-facility patient identity management.

The algorithmic matching approach employs patients' personally identifiable traits such as name, address, phone number, social security number (SSN), gender, etc., in order to match records together. Within the algorithmic approach, there are two methods of matching records together. Matching methodologies can fall either under a deterministic model, a probabilistic model or a hybrid of the two. These methodologies are explained later in the report.

Deterministic and Probabilistic Models

As discussed, the challenge of record matching can be addressed by one of two standard approaches: deterministic matching (sometimes called exact match logic) or probabilistic matching.[1] The theory of probabilistic matching, pioneered by statistical decision theorists Fellegi and Sunter in the 1960's, recognizes that each field-by-field comparison is subject to error.[2] This approach considers both the probability of a mismatch between data values in two records that represent the same entity, and the probability of a coincidental match between two records representing distinct entities. When calculating the likelihood ratio that the records refer to the same entity as compared to the hypothesis that they refer to different entities—while also allowing for incomplete values and/or error conditions within the records—the process is said to be probabilistic. A probabilistic matching algorithm determines, with some predetermined acceptable level of certainty, that two records likely refer to the same entity and therefore link them. This is done by assigning a score to indicate the likelihood that two records are a match. The higher the score, the greater the likelihood there is a match between records.

Deterministic matching examines a subset of attributes and marks two records as referring to the same patient if they have an exact match based on this subset of data. A simple example would be to link two records if they agreed on last name, first name, and phone number (many real-world examples have complicated rules which deal with missing attribute values and other anomalies). The two main drawbacks to this approach are that it often misses matches because of variations in data values (e.g. "ROBERT" versus "BOB", or errors in entering a phone number), and that this technique does not scale well to large datasets because it does not take into account attribute frequency; that is, a match on the last name "SMITH" does not mean as much as a match on the last name "EINSTEIN."

[1] A third approach that relies on matching through shared identifiers is sometimes used, especially within a single health system. Matching through shared identifiers only works when there is a reliable identifier (such as a medical record number (MRN)) that is completely and consistently populated in all data sources and is absolutely free from recording error. While an HIO most likely utilizes an MRN for identifying patients in its MPI, each hospital and provider that sends data to the HIO will utilize its own unique MRN. Consequently, utilizing a shared identifier for matching patients is not realistic within an HIE's MPI.

[2] A Theory for Record Linkage. Ivan P. Fellegi and Alan B. Sunter. Journal of the American Statistical Association, Vol. 64, No. 328 (Dec., 1969), pp. 1183-1210. http://www.jstor.org/stable/2286061.

Probabilistic matching avoids some of these drawbacks by recognizing the variability and volatility in attribute values or attribute significance (phone number vs. gender) and incorporating that knowledge into the decision whether to match two records or not. Among all the approaches to record matching, probabilistic matching allows the greatest flexibility and provides the highest accuracy when properly configured. Neither the technique of shared identifiers nor the deterministic matching method is able to match records under conditions of high variation in the data which are likely present within a hospital or health system and always present in a cross-facility HIE effort. Only probabilistic matching mimics the human ability to recognize that two slightly dissimilar records do, in fact, represent the same identity. This is done through the use of matching techniques that rely on applications of attribute weights (phone number receives a higher weight than gender), Enhanced Soundex (names with similar phonetic sounds receive a higher score), frequency indexing (common names receive lower scores, uncommon names receive higher scores), nickname tables (tables that equate formal and informal names), and edit distance calculations (the number of changes needed for two values to be equivalent, the lower the number of changes the more likely the records are a match).

Relying on a completely automated probabilistic record matching and linking approach, requires an extremely high threshold for accuracy, or the correct linking of two identities. The fundamental challenge in driving towards that goal is limiting false-positive identity correlations (falsely linked two different patients), while limiting the false negative correlations as much as possible (not linking two records for the same patient). Mitigating these risks is possible and is a cornerstone of effective patient identity management.

False Positives and False Negatives

When implementing and configuring an MPI solution and algorithm, the goal is to make as many correct linkages as possible and to minimize errors. When measuring the accuracy of an algorithm, a highly functional system has few false negatives and fewer false positives.

- False Negatives – Failure to match two records that represent the same entity.
- False Positives – Creating a link between two records that do not represent the same entity.

Cause of False Positives and False Negatives

The root causes of false positive or false negative linkages are numerous; however, the most common issues are related to data quality and matching thresholds. Data quality is a critical concept in master data management. The phrase "garbage in, garbage out" evokes the basic underpinning of the consequence of data quality shortcomings. Data quality generally refers to the completeness, validity, and accuracy of data flowing into the MPI. For example, if a record does not contain a full set of possible data elements and is missing a phone number, then the ability to match is diminished. Further, if invalid data is populated, such as a fake address for a trauma patient, again, matching is diminished. Lastly, inaccurate data capture or entry (for example during a patient registration process in an emergency room) causes matching challenges.

Matching thresholds can also cause false positives or false negatives. A threshold is the level at which records are automatically linked, rather than manually linked (manual linkage implying human intervention and disambiguation – sometimes referred to as potential linkage resolution). When comparing data elements in two records using an algorithm, a score is assigned. The threshold is correlated to the score at which records are automatically linked, not linked, or fall into a queue for a data steward to review. A higher threshold implies stricter matching requirements, i.e. a higher score is necessary for the records to automatically link. The set threshold will determine the number of false positive or false negative matching outcomes that will be encountered. When dealing with a father/son scenario where both individuals live in the same household with similar names, Jr. vs. Sr., under certain matching approaches the two associated records and data sets have a potential of linking if key differentiating data elements such as birthdate are missing. Similarly when addressing a twin scenario, the name and supporting demographic data suggests that would be the same individual; however they are in reality two separate distinct individuals. In an opposing scenario, the lack of data can also present a problem and cause records which should be linked, to not be linked.

During an implementation of a configurable solution, matching threshold must be viewed through the lens of false negative and false positive tolerance. In treatment scenarios, there may be a near zero tolerance for falsely linking patients. However, in a non-treatment use case seeking to identify readmissions across hospitals, that tolerance may change. With a higher threshold for auto-linking records, any data quality issues have a greater impact on the ability to match and therefore result in a lower match rate because of false negative outcomes. However, in matching scenarios that have a patient safety impact (i.e. linking clinical data), a false negative is a preferred outcome when compared to a false positive. Choosing an appropriate matching threshold in conjunction with an appropriately tuned algorithm will minimize both false positives and false negatives for an optimal MPI system. Beyond the initial configuration of a probabilistic MPI, a “tuning” effort can be undertaken to optimize the performance of the algorithm given the data set that the solution is operating against. For example, after a period of MPI operations, the administrators can review the linking and overall output of the solution and determine if matching thresholds should be modified (e.g. lowering of the auto-link threshold to capture more linkages) or if the tools should be accounting for specific regional variances in names.

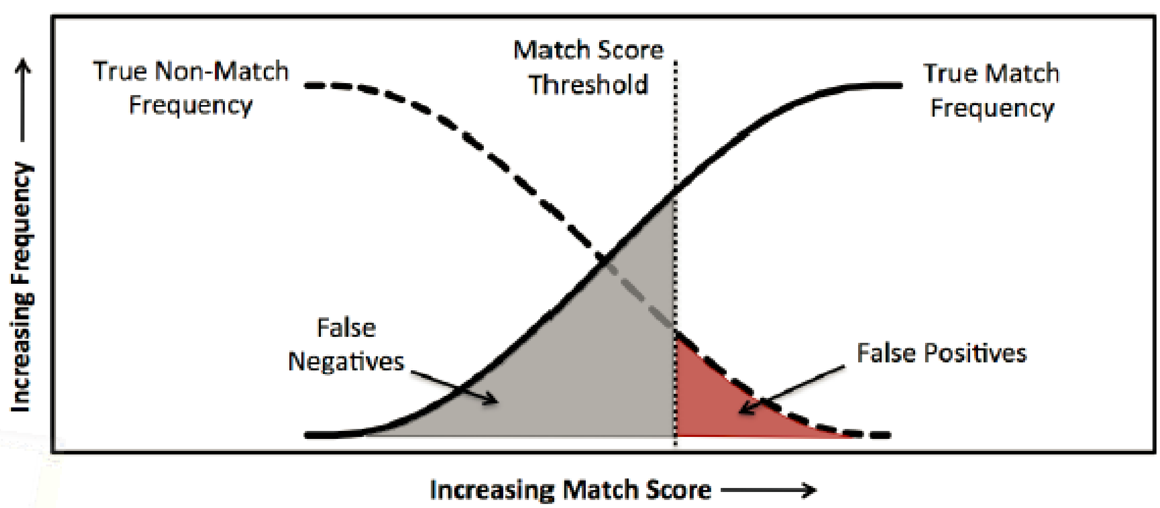

Figure 1: Illustration of the relationship between false positives and false negatives[3]

Balancing False Positives and False Negatives

When deploying an MPI solution, it is important to determine if an aggressive or conservative linking strategy will be pursued. Aggressive in this context, refers to erring on the side of linking records, as

[3] Privacy and Security Solutions for Interoperable Health Information Exchange: Perspectives on Patient Matching: Approaches, Findings, and Challenges. Alison Banger, Shaun Grannis, and David Harris. RTI International: June 30, 2009.

opposed to working to avoid false positives as an imperative (conservative). It is possible to tune an algorithm to minimize the number of false positives and false negatives; however, there is an important balance that will need to be addressed in regards to performance of the system and the amount of human intervention that will need to take place. Depending on the type of data that is being rendered, the tolerance for incorrect or missing matches will determine how finely tuned the algorithm will need to be in order to address the issue of false positives and false negatives. For instance, when dealing with healthcare data, it is imperative that records are linked appropriately with a very low rate of false positives, while false negatives are more acceptable. This is due to the higher level of patient danger if two records are incorrectly linked together. For example, if two records are incorrectly linked, and one patient has a medication allergy, the other patient's allergy list will not update, and she could receive a medication that is life threatening. If two records are not linked together that should be, there will be incomplete information, but providers are aware that information is rarely complete and tend to operate accordingly, which explains why false negatives are not as unacceptable as false positives.

Challenges of Relying on Demographics for Matching

It is critical for patient demographic data to be accurate and sufficiently populated in order to be an effective element in matching. Many demographic attributes will also change over time, such as a last name (due to marriage) or a home address (after a move). In addition to these demographic changes, data-entry errors, such as misspellings or transformation of data, inevitably cause variability in the records, described above as data quality issues. The matching approach must account for all of this variability to serve the needs of an HIO. Implementers and managers of the MDM solution must, where possible, take a hard look at the underlying workflows and policies that result in the demographic data being collected, and work with owners of source data to ensure that it is as accurate as reasonably feasible to support matching. From an HIO perspective, this can be a significant challenge. Data quality problems are frequently deeply rooted in the business processes and personnel practices of data source organizations. The ability of an HIO to drive changes at that level may not be feasible, or at least not feasible in a medium-term timeframe. The configurations and tuning of matching algorithms must account for the current data quality situation and may be updated over time if and when progress is made on improvements at the data-source level.

Transactional versus Batch Matching

Regardless of the matching approach utilized, there are two modes for processing patient identities for matching: transactional (real-time) or batch mode.

Transactional

In transactional matching updates from source systems are sent and received by the MPI in real time. The data is added or updated, and the matching of the records is done on the fly. Demographic changes or changes in entity composition can be sent to downstream sources so that all participating systems have the most up-to-date information. This method provides great benefit, since at the time of the addition or update, all of the systems will have the most up-to-date information. In the transactional model, human intervention can take place in order to allow for data remediation when a record link is in question.

Batch

With the batch mode, additions or updates are stored and processed at a determined time. Down-stream source systems are updated once the data has been processed. In this model, there is a period of time in which the source system, the MPI, and downstream systems receiving the data will be out-of-sync. This method of matching is most often utilized in research situations to create reports. For example, if an organization were to process a large batch of claims data in order to show unique identities across health plans over time, the claims data could be processed as a large batch file. Additionally, this method could be used if a persistent connection to a source system were not possible. In that scenario, messages may build over a given timeframe then be transferred via a secure File Transfer Protocol (FTP) to the HIO for processing through the MPI.

Data Governance

A data governance and data stewardship plan lays out how an organization prioritizes data issues, when it manages issues, and who should manage issues.

What is Data Governance?

Data governance encompasses the management and ownership of data within an organization. The processes associated with the ownership and management of data can be described as:

- The overall management of the availability, usability, integrity, and security of an organization's data;
- A system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who, what information, when, under what circumstances, and using what methods can action be taken;
- The people, processes, and information technology required to create a consistent and proper handling of data across an organization; and
- The activities that ensure data-related work is performed according to policies and practices as established through governance.

When putting together a data governance plan, there will be various short and long term goals. Success in implementing the strategy will be measured incrementally as change throughout the organization will take time, and ultimately effective data governance and stewardship is a long-term strategy that will require commitment from the top level of the organization. Data governance is not only a function of the information management department, but it should, in parallel, involve future IT project plans, in order to incorporate developed standards across the organization.

Data Remediation

Data remediation, a core consideration of a sophisticated matching plan, is the manual intervention that occurs when the matching algorithm does not have sufficient information to make a definitive linking decision. Records will be flagged as having a task associated with them and be identified with a particular issue. There are generally four different task types that a system customarily identifies.

1) Potential duplicate: Two records from the same source that, when compared, score between the clerical review and the auto-link threshold. To perform remediation, the user can make a resolution within the MPI; however, the user should also merge the two records within the source systems, if the records are determined to be the same. Figure 3 provides an example of

a duplicate record. The MRN is different because the first name has not been entered exactly the same; however, the remaining information verifies that it is the same record.

MRN	Name	Sex	DOB	Address	Phone	SSN
JHH 31243	Matthew Taylor	M	8/13/75	805 Hilltop Lane Baltimore MD 21022	410-312-7638	312-67-5342
JHH 76812	Matt Taylor	M	8/13/75	805 Hilltop Lane Baltimore MD 21022	410-312-7638	

Figure 2: Example of a potential duplicate record.

2) Potential linkage: Two records from different sources that, when compared, score between the clerical review and the auto-link threshold. To perform remediation the user can make a resolution within the MPI. Figure 4 provides an example of two records from different organizations that should be linked together.

MRN	Name	Sex	DOB	Address	Phone	SSN
JHH 31243	Matthew Taylor	M	8/13/75	805 Hilltop Lane Baltimore MD 21022	410-312-7638	312-67-5342
UMH 87125	Matthew Taylor	M	8/13/75	805 Hilltop Lane Baltimore MD 21022	410-312-7638	312-67-5342

Figure 3: Example of potential linkage between records from different sources.

3) Potential overlay: This scenario exists when an existing record is updated and the changes to the record are significant enough that existing matches are now in question. For instance, if an existing record is updated from John Smith to Jane Jones, there's a reasonable question about whether the record still matches John Smiths from other source systems. To perform remediation, the user must determine which attributes are correct or determine if there was a process flow issue. Figure 5 provides an example of two records with the same MRN; however the remaining information shows an overlay of records that should not be matched.

MRN	Name	Sex	DOB	Address	Phone	SSN
JHH 31243	Matthew Taylor	M	8/13/75	805 Hilltop Lane Baltimore MD 21022	410-312-7638	312-67-5342
JHH 31243	Greg Smith	M	8/13/75	817 Hilltop Lane Baltimore MD 21022		

Figure 4: Example of potential overlay of records.

4) Review identifier: Two records from the same source that do not score high enough to be a potential duplicate; however, the two records share the same unique identifier such as a SSN. Remediation steps must determine whether the records are the same or not; if they are the same, the records should be merged within the source system. If they are not the same, a data steward should indicate such in the data remediation tool.

Identity Volatility within a Master Patient Index

Within an MPI, an individual person will be represented as a composite—the sum of her collected data from all of the source systems that contain shreds of information about her. MPI technologies, as described above, link these identities to form a composite view of an individual. The composite is identified by a master identifier linking all other identifiers; this is sometimes referred to as en enterprise identifier, or EID. For an HIO, this EID can be thought of as an umbrella ID, linking any local IDs across data sources within the network. Perhaps surprisingly, however, an EID is different from a national ID number (such as an Social Security Number) in one very important way—it is not a static value, and as the identity composition changes over time, occasionally so can the EID and the local identifiers associated with it.

Factors Influencing Volatility

There are many factors that influence the volatility of an identifier. During an MDM implementation, it is desirable to mitigate risk factors so as to ensure that EID changes remain as infrequent as possible.

When an existing record within a given source system in a network is updated (for example if a patient has a second encounter and her phone number or address is updated), the closing or widening of the data gap with records from other source systems can incur an EID change. For instance, imagine two records that are not currently linked, each with its own respective EID. One of the records is then changed in the source system at a hospital, firing a message to the HIO, now making it similar enough to the other that the scoring allows for them to automatically link. One record will change its EID to match the others. A similar scenario can occur when two entities that are initially linked become unlinked due to a data change. The EID changes can be minimized in this case by allowing historical values to be used in the matching algorithm, i.e. previous addresses or names are used to determine if the records should remain linked. Similar discarding of an EID can occur after manual intervention, when two records do not meet the automatic threshold for linking or delinking, but are possible candidates. An effectively tuned algorithm can minimize this scenario.

Impact of Volatility on Forms of Data Exchange

One of the goals when implementing an MPI is to design and configure the toolset in order to identify as many entities automatically, either at the start of the implementation or as the entities are created. This should help to minimize EID changes as much as possible. Since in many implementations the EID is utilized in downstream systems, when an EID does change it impacts the environments in a variety of ways. Whenever the EID changes, additional processing time will be incurred (though minimal on the MPI side); however, as the EID change is communicated to downstream systems, they must consume and process the change. This can mean corrections, merges, and unmerges at the source-system level. Additionally, and perhaps more importantly, the clinical documents that are linked based on the medical record number (MRN) to EID correlations must now change to reflect the updated information. If new information is sent and a new link is created, not only must the identities in the MPI link, but the downstream systems storing the clinical documents associated with each MRN must be accordingly modified to show the clinical documents as one entity within the HIE infrastructure.

MPI as Foundational Infrastructure

The ability to accurately identify and link patient records across healthcare entities is a foundational capability in health information exchange. Indeed, accurately linking identities from disparate sources is the first step to enable searching of data through an HIE infrastructure. In order for an HIO to provide services that facilitate, and perhaps drive, care coordination, cross-facility patient reconciliation and matching is required. Previous Audacious Inquiry reports described how simple data types and message formats (e.g. HL7 ADT messages) shared through an HIO from healthcare facilities, can create significant opportunities for new care coordination capabilities and drive novel insights into utilization and potential interventions. In order to enable these use cases, an MPI is required.

A Note on Federated MPI Models

Other models beyond a centralized MPI have been developed and are being deployed. In a scenario where regional HIOs seek to exchange data among each other, their infrastructure solutions must be able to support cross-community queries into each other's MPIs and respond when a patient match occurs. The concept of a federated MPI solution exists but has inherent design challenges related to the federation of matching policies and processes. Specifically, in a federated MPI model a requestor of data must broadcast the patient query out to the MPI solutions serving the other communities. This patient query carries demographic data pertaining to that patient and must be processed by the receiving MPI solutions to evaluate if that system has a corresponding match. In a federated MPI model the requestor is not presented with an opportunity to disambiguate potential matches from each responding MPI (as they may be presented in a single MPI domain), due to the large number of identities this may return and the manual work that would be required to ensure a correct match. While either deterministic or probabilistic matching approaches may be used, either solution must respond to query within a single patient identified, as the workflows to disambiguate multiple MPI responses are not defined and are likely prohibitive. The concept of a "broadcast" patient query exists in the cross-community access IHE profile. If a user within a network launches a query out to a broader network of networks, using specific patient demographic data, the receiving MPI systems must be sure that the responses to that query are accurate and unambiguous. The deterministic matching approach (when the request includes sufficient patient demographic data) is the most effective way to accomplish this end result.

Master Data Management, Care Coordination, and Healthcare Reform

Master Patient Indexing Beyond Core HIE

Healthcare reform and new payment paradigms are putting increasing pressure on provider organizations to effectively and efficiently coordinate care. Whether it is in the context of a patient centered medical home (PCMH) program, an accountable care organization (ACO), or a hospital readmission-reduction program, those charged with a patient's care need to have full situational awareness of the healthcare utilization of that patient across the entire system (whether the system is defined as an organization, an ACO, a provider network, form of coverage, state, etc.). The realities of disparate systems and the fluid nature of patient demographic information mean that some form of patient matching and indexing technology is required to accomplish effective care coordination at scale.

As a clearing house of real-time ADT data and an organization with touch points with a wide range of healthcare organizations across a medical trading area, an HIO and its MPI are well positioned to serve a central role in care coordination. This is a great example of the “utility model” of health information exchange, where the HIO serves as a unique enabler of a business-driven use case of its participants. HIOs across the country are developing a number of services that utilize the MPI and facilitate care coordination.

As an example, at least two relatively mature HIOs are delivering ADT information about a patient’s medical services encounter, for instance at the time of hospitalization, to a permitted recipient with an existing relationship with a patient, such as a primary care provider or payer. Until now, participants were not likely to know when one of their patients has been admitted to a hospital, or alternatively they may find out well after the admission and/or have incomplete data. This use case hinges primarily on three events:

1) Patient attribution (matching a patient and provider),
2) Alerting rules, and
3) Alert notification output

Such systems are designed to provide real-time notifications for care coordination and quality improvement purposes when patients are admitted, discharged, or transferred within a hospital. These alerts can serve to initiate a process for coordinating care and/or providing follow up care.

In addition to real-time notifications, organizations charged with coordinating care have a need to see and understand in aggregate how patients are moving across care settings. Often, claims-based data cannot be collected and delivered to such organizations in a timely enough manner to make impactful decisions. HIOs’ MPIs can be used to deliver actionable reporting on patient movement; for instance, in Maryland, the HIO has developed at least two types of readmission reports. The first type of readmission report is a 30-day all cause basic readmission report which simply assigns the readmission to the previous hospital visited by the patient, regardless of cause. The second type of readmission report is a 30-day all cause readmission report utilizing the state’s regulated readmission logic. Hospitals may also request custom one-time or recurring reports using either the existing HIO data or the HIO data enhanced with the requesting hospital’s input. Once a report is provided, the participant may then use the information only as defined in the permitted purposes of the HIO’s policies and procedures, namely for quality assessment and improvement activities, including care coordination.

Master Provider Indexing

HIOs are implementing provider directories as part of their rollout of Direct secure messaging. Most MDM solutions oriented towards patient information management also offer provider-specific MDM toolsets. The more robust provider indexing tools may not be necessary simply for enabling Direct, and so each HIO must conduct a cost-benefit analysis with regards to investing in master provider indexing; however, those HIOs with strong provider-specific capabilities available to them may be able to play a role managing provider information more efficiently among a range of other stakeholders. For example, a unified approach to provider information management, say at a state level, could to act as a mediator of the “collect-once, reuse-many times” provider data service for verified licensure, sanction, and provider demographic / geographic information. Organizations pursuing such services have estimated that the savings of such an approach is quite significant. The solution can support the costly primary source verification processes in addition to maintaining consumer-relevant provider information and communications relevant electronic address information (Direct and/or EHR endpoint data).

Analytics

Entities such as state Medicaid agencies, payers, ACOs, hospital systems, practices, and even health benefit exchanges have unique needs, data sets, analysis responsibilities, and capabilities. Coordinating data analysis efforts can enable significant advancements in a state's health data analytics strategy by combining data sets to create new insights into cost, performance, and outcomes. This coordination will become even more necessary in the context of episode-based care coordination and payment. As measuring utilization, cost, and outcomes becomes increasingly centered on cross-organizational episodes of care rather than each individual case, or unit of care, technologies that enable consistently accurate consolidation of disparate data sources becomes important, both for the provision of care, subsequent analysis, and potentially predictive modeling. This capability is a core component of many HIOs' technology platform and is a pre-requisite for managing large scale data sets and to developing insights into that data. It will be important to state health policymakers and to HBEs in understanding the migration of individuals among plans and in ensuring coordinated care and improvements in the patient experience are provided as individuals move between different provider networks associated with each payer. It is also important to recognize that state and federal policymakers stand to benefit from the creation of population level health measures. Over the long term, a unified planning approach to a state's use of claims based and provider-originated health data for population health reporting will drive efficiency and create a platform for future innovation.

Appendix A: Glossary

API: An application programming interface (API) is an interface between software components.

Blocking attribute: Internal-use attributes with values constructed from normalized attribute values. The intent is for an MPI to define its blocking attributes so that identities with the same value for a blocking attribute are much more likely to find matches for each other than identities chosen randomly from the entire data set.

Data quality: Data quality is a perception or an assessment of data's fitness to serve its purpose in a given context.

Data validity: Data validity indicates the percentage of times a particular attribute is populated. Furthermore, it will also indicate the percentage of the data set which contains valid, non-anonymous value values.

Deterministic matching: Records are determined to refer to the same patient if they have an exact match based on a subset of data, i.e. name, DOB, SSN.

Duplicate: Two separate records from the same source that are in fact the same patient and should be one single record.

Edit distance calculations: The number of changes needed for two values to be equivalent. The lower the number of changes, the more likely the records are a match. For example, an SSN of 123-45-6789 vs. 213-45-6789 has an edit distance of one because only one swap between the one and two must occur to make a match.

Enhanced Soundex: Names with similar phonetic sounds receive a higher score.

False negatives: Failure to match two records that represent the same entity.

False positives: Creating a link between two records that do not represent the same entity.

False positive filter (FPF): Applies deterministic logic to specific false positive matches, and uses the result to apply a penalty score.

Fellegi-Sunter theory: A theory of probabilistic matching pioneered in the 1960s that recognizes that each field-by-field comparison is subject to error.

Frequency indexing: Common names receive lower scores, uncommon names receive higher scores.

GUI: A graphical user interface (GUI) is used to interact with a database using graphics, rather than text commands.

Historical values in matching: The use of previous addresses or names (maiden names) as part of the matching technology that allows for a stronger link between records and consequently a larger number of matches.

Identity Volatility: The vulnerability ratio for the identity itself to change, in other words, when the identity is added to the system how often will its enterprise identifier change based upon data changes for that particular identity or other identities in the MPI.

Linkage: Two separate records from different sources that are the same patient and should be linked together in the MPI.

Matching thresholds: The level at which records are automatically linked, rather than manually linked.

Medical record number (MRN): The unique identifier assigned to a patient's record within a specific EHR system or organization.

Normalized attribute: Copies of incoming attribute values that have been "cleaned up" (i.e., standardized, or normalized) in some way, to make the matching process easier. For example, a normalized last name attribute would be a copy of the incoming last name attribute that has been cleaned up, perhaps by removing leading and trailing whitespace characters, removing all non-letters, and mapping to all upper-case.

Nickname tables: Tables that equate formal and informal names.

Overlay: An overlay occurs when one identity's information overrides the information of another, different record. For instance, if an existing record is updated from John Smith to Jane Jones, and the demographic information is changed, this is an overlay.

Probabilistic matching: The process of using statistical analysis to determine the overall likelihood that two records match.

Review identifier: Two records from the same source that do not score high enough to be a potential duplicate; however, the two records share the same unique identifier such as a SSN.

Appendix B: Third Party MPI Products versus Integrated Solutions

In today's HIE product ecosystem, broadly speaking there are two flavors of MPI solutions—the built-in MPI and the standalone, full-feature MPI. Most HIE product vendors have an MPI solution that is built into their HIE product. This type of built-in MPI is generally purpose-built to serve specific HIE use cases and does not have all the features of a standalone MPI solution. This is not surprising, given that for HIE companies MPI is only one component of their overall offerings while for standalone MPI companies the level of product investment and differentiation is likely to be much greater. The following table compares and contrasts the general differences in the two types of MPI solutions in the market.

Feature/ Characteristic[4]	Standalone MPI	Built-in MPI
Algorithm	Probabilistic, with full support for statistical tuning based on particular demographic data sets.	Mostly deterministic with support for soundex, transpositions, etc.
	Configurable and customizable matching rules. The HIO can add its own data fields that are used in matching.	Fixed matching rules. Focus on generic rules that work for most populations.
	Less dependence on existence of identifiers such SSN.	A little more dependent on identifiers.
	Generally requires more work out of the box to configure and tune.	Very little or no algorithm configuration required.
	Designed for accuracy. The algorithm can unlink an existing patient link based on changes in data.	Generally, these MPIs are designed for consistency. Once two records are linked together, they can only be unlinked manually.
Data Remediation / MPI Management Tools	User interface to support data remediation.	Very little or no support for this.
	Support for task assignments and workflows related to data remediation.	Very little or no support for this.
	Tools to examine false positive and false negative candidates.	Very little or no support for this.
	Built for complete identity management and flexibility.	Built to maximize auto-linking and minimize manual identity management.
Integration and Interoperability	Great support for interoperability standards (HL7, IHE PIX/PDQ)	Inconsistent and spotty support for external applications.
	More integration work required to make things work. Despite the interoperability standards, making any HIE product work seamlessly with a third-party MPI is generally a major challenge.	By definition, these are very well integrated into their respective broader HIE products.
Cost	High	Typically free with the HIE product

[4] The table attempts to paint a general picture. Specific products may have different capabilities. A few HIO vendors provide their MPI solutions as a standalone product as well.

Appendix C: HIE and MPI Vendors

Audacious Inquiry reached out to a number of HIE vendors that offer an MPI solution, as well as vendors that offer a standalone MPI solution, to provide information about their product offerings. The vendors were sent 18 questions about various aspects of their systems. All content contained in Appendix C was provided and reviewed by the vendors.

IBM

General Overview

The Standard Edition of IBM InfoSphere MDM software is designed to support complex data management needs around person/patient identification, matching, linking and de-duplication of person data, as well as integrating to clinical systems and EHRs to support clinical care and HIE activities. For effective management of patient records, InfoSphere MDM supports the following core capabilities:

- Creation and maintenance of a common index of all personal data and encounter information to allow on-demand access to patient data across the continuum of care.
- The ability to correlate and cross-reference different system identifiers to a single person and a single unique identifier.
- Utilization of sophisticated probabilistic matching algorithms to ensure highly accurate results.
- Broad compatibility and integration experience with existing systems, platforms, and data sources – including legacy systems.
- Provides a platform that can expand to support the ability to identify and manage patient relationships with other persons, families, entities, payers, contracts, guarantors, physicians, and resources.

MPI Capabilities

To derive the accuracy and benefits of a probabilistic approach to patient data matching, InfoSphere MDM:

- Uses a probabilistic algorithm for patient matching.
- Provides this algorithm for patient searching, which may require different levels of accuracy.
- Produces customized weight tables that are generated based upon the actual frequency of data in the participating systems.
- Provides an algorithm that supports an unlimited number of data attributes.
- Stores and matches on an unlimited number of historical versions of the patient attributes.
- Provides separate algorithms for identifying patients, providers, organizations, and households.
- Provides an algorithm that awards both positive and negative scoring per each attribute.

The IBM InfoSphere MDM Patient Hub comes with a pre-populated data model configured with the most popular patient attributes, which can be extended and tailored according to a client's business needs. A flexible and robust data model allows for post-production changes to be easily made, for instance: adding sources, attributes, or new data entities; and refining or expanding the matching algorithm and thresholds.

The solution provides flexible views of patient data and can easily provide the ability to synchronize data between upstream and downstream systems.

To integrate, InfoSphere MDM supports HL7v3, HL7v2, delimited, fixed length, and XML based messaging. InfoSphere MDM also supports the IHE standards and Web-services, Java, and C++ API as points of integration. InfoSphere MDM authenticates users either through an internal data sources or integration with an external LDAP v3-compliant directory server.

Identifying and Linking Patients

IBM InfoSphere MDM uses a probabilistic algorithm to match records. Specifically designed to consolidate fragmented patient identities into a linkage set, the patient data hub provides a complete patient view by accurately identifying and instantly linking the records for each patient. As a result, it is possible to provide the clean, complete personal profiles required for more cost-effective and responsible interactions with customers, patients, or partners. Achieving the highest levels of accuracy requires a methodical and thorough process. IBM InfoSphere has customized the following process to match records.

The candidate selection process selects records from the database that are likely to match. For each record in the database, search keys are created. This enables the record to associate with other records that likely match it. Search keys are designed to be robust against common errors such as false positives or false negatives and create multiple keys for each record. These multiple, robust search keys insure that all records with a reasonable chance of matching are considered, a process known as casting a wide net.

Comparison functions operate at the attribute level and determine the degree to which the attributes match. This step pulls together disparate data and is the foundation for building a complete record and an accurate view of each patient. These can be simple binary functions, either the attributes agree exactly or not, or complicated hierarchical comparisons involving phonetic coding and edit distance functions. These functions include: social security numbers or other license number comparison which comprehend typographical errors, name comparisons which utilize phonetic matching, nickname matching, exact matching, frequency based comparison, edit distance comparisons, address and phone comparison which is capable of standardizing the address, and intelligently comparing the address. Scoring provides the mechanism for combining the individual attribute comparison results into a meaningful value – indicating the likelihood two records represent the same patient or not.

MDM InfoSphere Initiate Inspector is an application that alerts data stewards to potential data quality issues and gives them tools to inspect and resolve these issues using simple editing methods and drag-and-drop paradigms. Inspector assists in the administration of routing rules, definitions and assignment of issues within the team, and distribution of the workload to the most appropriate resources.

Deployment Options and Licensing/Pricing Models

IBM InfoSphere MDM technologies are deployed on premises and can support multiple registry styles.

Medicity

General Overview

Medicity's MediTrust Identity Management Services (IMS), including Community Master Patient Index (CMPI) and Enterprise Master Patient Index (EMPI), match patient transactions accurately across information systems and care locations (acute care and ambulatory) within the HIO, based on community algorithms which are typically very different from the connected enterprise EMPI algorithms. In addition, the MPI technology goes beyond patient matching, ensuring that provider and health plan information is also consistent throughout hospital records. Medicity's CMPI supports a wide variety of provider, payer, patient, and other identifiers. The CMPI abstracts external identifiers from their associated data using internal identifiers that enable any number of external identifiers to link to a logical entity like a payer or a provider. Medicity deploys an EMPI in a passive mode for every connected organization with federated data storage to prepare identity and data for community publication. The value of the two-tier model is that each enterprise can manage identity according to their local environment, and still have identity match at the community level based on another set of criteria.

The community and enterprise MPI indexes update in real time with demographic and other feeds from physician and hospital information systems. The CMPI receives new records through system interfaces, adjudicates them, and matches them with an existing patient or enters them as a new patient into the index, with appropriate notation in the record written to Medicity's clinical data repository.

MPI Capabilities

Medicity's products support all applicable IHE Profiles in the ITI Technical Framework Volume 2, including ITI Profiles such as CDA R2 Clinical Document XDSb and XCA Transactions, PIX, PDQ, XCPD, and ATNA, among others. User authentication can take place against LDAP, Active Directory, or Medicity's internal user directory. Authentication to the system is used to then dynamically determine associated access controls to various types of clinical data. LDAP authentication can be problematic in a community setting with multiple LDAP systems; Medicity recommends the use of O-Auth or SAML as an alternative. In addition, Medicity offers both native APIs based on Microsoft .NET Windows Communication Foundation (.NET WCF) and supports the IHE PIX v3 Manager transactions via a SOAP based Web Service for searches. The following patient searches are supported:

- Demographics
- Local identifier, i.e. MRN
- Unique identifiers, i.e. SSN or driver's license number
- MPI identifier

Medicity supports a Lucene Index for patient searches as well as common data attribute level algorithms for specific attributes such as: Soundex for names and alternate name/nick name matching on first name attribute to enable matching of Bob, Rob, and Robert, transposition errors including first name / last name swaps, keyboard errors and restrictions around common invalid data values, all ones, all twos, etc. Unlike centralized MPI methodologies, where algorithm adjustments that benefit one organization often adversely impact other organizations in the system, Medicity's matching implementation enables the matching algorithms to be tuned for each participant that contributes data to the network. Medicity provides a user interface for manual identity management. In cases where a match is not certain (as defined by the policy makers), the system creates two separate patients and flags them for manual resolution. At the enterprise level, the Enterprise Patient Merge tool ("Patient Merge/Link" tool) allows authorized users to perform patient merge. Only the owners of the data are able to access and modify this patient data. Personnel usually perform this merge activity at the individual data contributor level (e.g. hospitals, ref labs, etc.). At the community HIO level, the "Patient Move/Link" tool enables authorized users to link or move patients between patient search catalogs. This can be performed by an administrator at the HIO level. Finally, the Medicity MPI allows for patient / provider correlations. These relationships can be extracted from the HL7 ADT messages and/or entered manually by the user when attempting to access a patient's record.

Identifying and Linking Patients

Medicity's CMPI uses deterministic logic and weighted rule sets to adjudicate patients. Medicity can customize its algorithms to meet both the requirements of the data sources and the organization's tolerance for error. The CMPI toolkit provides a mechanism for viewing and resolving potential duplicate records as well as logs detailing mismatched identities. The CMPI includes a utility to help administrators identify and merge or link duplicate items in the index. The system automatically generates a log of duplicates and potential duplicates. An administrator can then:

- Merge two patient records, including identifiers, into one record at the organizational (EMPI) level; or
- Link two patient records, leaving both records in the community index at the community index level. A common use case for such linkage involves maiden and married names. The patient with the maiden name remains in the database but is linked to the patient with the married name. The system recognizes both patients as one person.

To balance false negatives and false positives, IMS uses probabilistic weighting. The weighting value for each factor is defined via statistical analysis, in coordination with clients, with the objective of reducing false positives to a value, asymptotically approaching zero (reducing overlays at the expense of duplicates). Medicity's implementation methodology integrates this statistical analysis with a variety of other techniques, such as false negative reduction, deterministic matching, and run-time performance optimization across all assigning authorities in the community.

A minimum set of patient attributes can be defined between Medicity and the client according to their tolerance for false positives. A typical minimum set of data would be as follows:

- First Name
- Last Name
- Middle Initial
- Suffix
- Date of Birth
- Social Security Number
- Gender
- Home Phone
- Address
- Zip Code

Deployment Options and Licensing/Pricing Models

Medicity offers self-hosting, third-party hosting, and Medicity hosting based upon the requirements of the client. Medicity's solutions can be acquired via a traditional licensing model or based on a hosted Software-as-a-Service (SaaS) model.

Mirth

General Overview

Mirth Match is engineered to be an enterprise class Identity Cross-Reference Service built on open standards and leveraging the strengths of existing open source components to provide a powerful, scalable and comprehensive solution capable of handling common healthcare MDM patterns such as EMPI, master provider index, facility index, or simple member list de-duplication. Built as a reference implementation of the [HSSP Identity Cross-Reference Service (IXS) specification](), Mirth Match correlates or cross-references entities (e.g. patients, providers, etc.) located within disparate source systems by their identifiers. The standard IXS web service interfaces provide a documented standard mechanism by which entity data can be registered, indexed, searched, and retrieved. By maintaining a unique community identifier and providing standardized search mechanisms, healthcare applications and healthcare enterprises can find, exchange and reference entity data while maintaining the data's context and associations.

MPI Capabilities

Mirth Match supports both the IHE PIX and PDQ integration profiles and is able to integrate with LDAP v3-compliant directory server for authentication. The Mirth Match administrative user interface allows users to manually register, modify, activate, deactivate, link, unlink, merge and unmerge entities (e.g., patients). The MPI can maintain patient-provider correlations via traits or attributes, such as a provider's NPI as an attribute on a patient entity. The MPI can be searched via a user programmable API or an administrative user interface. The following types of patient searches are supported:

- Demographics
- Local identifier, i.e. MRN
- Unique identifiers, i.e. SSN or driver's license number
- MPI identifier

Identifying and Linking Patients

Mirth Match supports both probabilistic and deterministic scoring algorithms, as well as hybrid probabilistic/deterministic matching. Support for gathering and looking up attribute value frequency information is built into Mirth Match. This information can be used to adjust the calculated similarity scores. Mirth Match supports a three-way balancing between false negatives, false positives, and the workload on support staff, through two configurable similarity score thresholds: the potential link threshold and the auto-link threshold.

When the similarity score calculated by Mirth Match for a pair of entities (patients) is below the potential link threshold, no link will be created between the entities. When the similarity score is above the auto-link threshold, a link will be created. When the similarity score is between the two thresholds, a work item will be added to the support staff work queue requesting that a human make the final decision as to whether or not a link should exist between the two entities. To control the balance between false negatives, false positives, and workload, the administrator adjusts the values of these thresholds. If the auto link score threshold is increased, the number of false positives is reduced and the number of work items that must be handled by the support staff is increased. If the potential link score threshold is increased, the number of false negatives is increased, and the number of work items that must be handled by the support staff is reduced.

When a new entity (a patient, for instance) is registered, either during initial backload or later during the active life of an installation, Mirth Match executes a multi-step process:

- Compute values for normalized attributes from the incoming source attribute values. The exact logic varies by attribute and installation, but usually includes aspects such as mapping text to all

upper-case, stripping leading and trailing whitespace from name components, discarding non-numeric characters from SSNs, discarding known invalid values or out of range values, etc.
- Compute values for blocking attributes from the normalized attributes. The exact details vary by installation, but speaking generally, blocking attributes are constructed so that entities (patients) with the same value for a blocking trait are more likely to be linked together than entities at random would be.
- For each blocking attribute, find all entities with the same value of the blocking attribute. We refer to these entities as the candidates.
- Compare each candidate entity to the entity being registered and compute a similarity (degree of match) score between the two. The set of attributes compared, the method used to compare their values (exact match, Levenshtein,[5] etc.), and the weight given to each attribute are all configurable.
- For each candidate, if the computed similarity score is above the configured auto-link score threshold, create a link between the new entity and the candidate. If the computed similarity score is between the configured potential link and auto-link score threshold, create a work item requesting that a support staff person review the possible link and make a decision.

Mirth does not have any mandatory attributes for patient matching. For domains based on Mirth's Fellegi-Sunter[6] matching module, there are no intrinsically mandatory traits, but using scripting, an installation can configure which attributes are mandatory. Mirth has the following recommendations for patient matching:

- Last Name: required
- First Name: required
- Middle Initial: optional
- Gender: optional but preferred
- DOB: required
- City/Locality: optional
- Postal Code/Zip: optional
- SSN: optional but preferred
- Phone: optional
- Mother's Maiden Name: optional
- City of Birth: optional
- Birth Order: optional (for multi-birth)

Deployment Options and Licensing/Pricing Models

Mirth supports three deployment options:

[5] The number of changes necessary to make two records match. The lower the number of necessary changes, the more likely the records are a match.

[6] A theory of probabilistic matching that recognizes that each field-by-field comparison is subject to error.

- Mirth supplied physical appliances (preconfigured rack mountable computers) positioned in client data centers,
- Mirth supplied virtual appliances deployed in client virtual machine infrastructure, and
- Mirth hosted and managed virtual appliances, located in a SOC2 / SOC3 datacenter.

Mirth Match is licensed either perpetually or with an annual subscription. Pricing is based on the net entity count for the instance (e.g. net patient population).

Orion Health

General Overview

Orion Health uses the NextGate EMPI, an EMPI software application that links and tracks related patient data from disparate applications and facilitates its use for an enterprise master index of patients. It uses probabilistic and deterministic algorithms to evaluate patient records and assigns an EID to facilitate data quality management and interoperability. The EMPI is designed to identify and resolve duplicate patient records, improve the accuracy of patient identification, enable a patient-centric view of information, and simplify the connectivity with EHR systems, HIE solutions, and the NwHIN.

MPI Capabilities

With NextGate's EMPI, the identification of duplicate records is performed continuously on every record added or updated in the EMPI. Data such as addresses are standardized to facilitate comparison, and probabilistic comparison techniques are used to match records where fields do not match exactly. The matching algorithm assigns detailed weights to each record field being analyzed to arrive at a composite match score. This composite match score is compared to a set of thresholds that help determine whether 1) an automatic match should occur, 2) the records should be listed as potential duplicates/matches, or 3) the records are instead, distinct entities.

Data Quality Manager (DQM), a web-based, user-friendly stewardship graphical user interface, is designed to provide streamlined workflow to manage the data residing in the EMPI. Detailed reports are created to indicate, among other items, which records require attention (sorted by facility, if necessary), which records have been merged or unmerged, and which end user is responsible for creating a potential duplicate.

NextGate represents that its EMPI is compliant with IHE profiles and standards, and satisfies the MPI portion of the NwHIN CONNECT architecture. It uses the IHE PIX and PDQ profiles to allow healthcare providers and other persons involved in the delivery of care to locate the patient's EHR information across multiple systems. The EMPI generates standard PIX/PDQ HL7 v2 and v3 messages to notify external systems of merge activity to allow those systems to perform their own internal merge operations, if desired.

The metadata model within NextGate EMPI is designed to be highly configurable to accommodate the storage of additional data elements associated with patient data. The data model saves names, nicknames/aliases, general demographics (e.g., gender, date of birth, ethnicity, marital status, religion, etc.), identifier numbers, addresses, phone numbers, credit card numbers, and much more.

Identifying and Linking Patients

The core functionality of the NextGate EMPI is to use a variety of criteria and sophisticated matching algorithms tuned for the local population to evaluate patient records and accurately identify patients. If a patient record matches an existing entry in the EMPI, site-specific rules are then invoked to link the system identifiers for the patient. There is theoretically no limit to the number of patient records that can be linked together, and data pertaining to each identifier is maintained. Records can also be unlinked and separated into distinct records with their original data.

The identification of duplicate records is performed continuously on every record that is added or updated in the EMPI. When an assumed match occurs, a survivorship calculator constructs an entity called the Single Best Record (SBR) from information in each of the linked underlying records to form a golden record.

Patient record merges can occur either automatically or with operator review. The “automatic merge” function is triggered when a new record matches an existing record above a certain configurable threshold. Once this threshold is achieved, the EMPI will automatically merge the two records into the SBR, and store the original records unchanged. Records that are automatically merged are noted in the EMPI log files for future review. This can trigger an outbound message to one or more target systems notifying them of the merge event. This automatic merge event triggered within the EMPI can generate a merge message that can be consumed by downstream systems, automating the merge process across the enterprise. An automated merge can be reversed at any point in time through the GUI or via an API call.

The criteria used to perform matching are fully configurable based on the fields present in the records from the source systems. The analysis process is performed as part of the initial load and is used to determine the best fields to use, the algorithms to be applied, and the weight to be given to each field.

The EMPI operator also has the ability to manually review multiple unique records side by side and determine if a merge is required. Using the Data Quality Manager (DQM – the data stewardship GUI tool), an operator initiates the merge operation and the EMPI merges the two unique IDs and all other associated system identifiers so that they effectively become the same person. The “survivor” record can inherit data from either record, such as address, alias, and contact information.

The default Orion Health HIE is designed to allow searching via demographics, local identifiers and the MPI identifier. Specific unique identifiers can be captured in the MPI and configured for searching against at a customer’s request. The NextGate EMPI supports searches where variations between search attributes occur. The EMPI uses probabilistic matching algorithms (rather than deterministic algorithms) to accommodate variances in search attributes and other data elements. Commonly referred to as “fuzzy logic,” these algorithms are combined with other methods to facilitate searches against data with varying levels of quality. The matching algorithm can identify potential transposition of attribute values, identify the potential transposition of characters within alphabetical attribute values, identify the potential transposition of characters within numeric attribute values, and incorporates the double metaphone algorithm.

The Orion Health Clinical Portal has been integrated with several third-party Identity and Access Management (IAM) solutions, ranging from Integrated Windows Authentication, LDAP (including Microsoft Active Directory), SAML for SSO, and user account provisioning in the portal’s user repository via SPML. Orion uses industry standard protocols such as LDAP and/or SAML in its standard integration with third-party IAM solutions.

Deployment Options and Licensing/Pricing Models

Orion Health offers both SaaS and traditional client-hosted deployments.

- The Orion Health SaaS offering is comprised of three core components: Subscription Software, Application Management Services, and Infrastructure Hosting Services. In addition to the Orion Health SaaS solution, Orion Health provides implementation and integration activities. Orion Health provides a full-service, 24x7 support and maintenance SaaS Team.
- Traditional on-premise licensing options are available on either a perpetual or subscription license model. This licensing flexibility allows organizations to meet their solution needs within the requirements of their specific operating and/or capital budget cycles.

QuadraMed

General Overview

QuadraMed offers a number enabling technologies around its core EMPI solution. Smart Identity eXchange (SmartIX) – Quadramed's Enterprise Master Patient Index Solution with advanced probabilistic record matching and linking abilities enables clients to identify and link all instances of a patient within and across its data domains. SmartIX resolves duplicate patient records and improves accuracy of patient identification across multiple facilities and health organizations to facilitate accurate electronic health information exchange. SmartIX includes SmartMerge; QuadraMed's advanced record management application for use in review and resolution of potential duplicate and overlapping records, and establishment of enterprise record linkages.

SmartMerge improves the task of merging duplicate patient identifiers in an MPI by displaying and categorizing all duplicate and multiple person identifiers. This powerful module enables sophisticated duplicate record management, flagging, workflow, and reporting. With a simple and easy-to-use interface, individual errors or groups of errors can be quickly reviewed, managed, and addressed. SmartID, another enabling technology, is an enterprise-wide patient search and selection tool which has been proven to increase the accurate identification of current patient records, facilitate the sharing of patient demographic information, and to reduce duplicate record creation. Finally, SmartSwipe, an add-on module to SmartID, minimizes keyboard data entry errors and locates patient records quickly and accurately by performing patient searches by card swiping state issued driver's license, credit cards, or Department of Defense ID cards and reading the data contained on the card's magnetic stripe.

MPI Capabilities

QuadraMed supports the IHE PIX and PDQ integration profiles and can integrate with LDAP v3 compliant directory servers. SmartIX supports a variety of searches through APIs, including web services queries, HL7 query responses, PIX, and PDQ. When searching, the product is able to accommodate aliases, data transposition, data proximity, digit rotation, misspelling, and other typographical errors via QuadraMed's proprietary, probabilistic algorithm. SmartIX provides a user interface for manual identity resolution, and is able to maintain patient and provider correlations within the MPI. The SmartIX product supports the following types of patient searches:

- Demographics
- Local identifier, i.e. MRN
- Unique identifier, i.e. SSN or driver's license number

- MPI identifier

Identifying and Linking Patients

SmartIX employs a probabilistic scoring algorithm. Scoring takes into account the relative frequency of values among many sophisticated matching capabilities, that significantly outperform deterministic and manually configured or "rules based" matching approaches. When implementing an MPI, QuadraMed utilizes SmartScan to perform an analysis of the MPI data to identify duplicate and overlapping medical records and prepare data for loading into SmartIX. The SmartScan utilizes the proprietary algorithm to categorize and report potential duplicate records. The SmartScan reports create a listing of candidate duplicates and overlaps used to load into SmartIX. The workflow automation tools streamline the processes necessary to cleanse an MPI and maintain a superior level of data integrity. QuadraMed uses client-specific business rules (i.e. Auto Linking Rules) to programmatically link records from disparate source systems in order to facilitate the creation of the MPI and minimize manual review and intervention. Matches that do not meet the criteria for programmatic linking are output to the SmartMerge work queue for manual review and linking.

QuadraMed does not have a maximum number of data elements required for the SmartIX MPI. The SmartIX data model supports the creation of an unlimited number of user-defined fields that can contain numbers or text. The user-defined fields can be defined to contain a single or multiple occurrences. The number and type of elements that are required for programmatic linking depends on the business rules that the customer defines. QuadraMed's probabilistic algorithm uses the following demographic elements for matching; the more elements that are available, the higher the statistical probability of the match.

- Last Name
- First name
- Middle Name or Initial
- Gender
- Date of birth
- Address (City, State, Zip)
- SSN
- Mother's Maiden Name

Deployment Options and Licensing/Pricing Models

QuadraMed does not share this information publicly.