Building an Enterprise Master Person Index

Save to myBoK

This practice brief has been updated. See the latest version here. This version is made available for historical purposes only.

An enterprise master person/patient index (EMPI) is a software application that identifies persons in an integrated delivery network (IDN) across disparate registration, scheduling, financial, and clinical systems. While EMPIs have been used in healthcare since the 1990s, interest in them has markedly increased due to the shift to a more customer-centric focus in healthcare operations, consolidation of healthcare organizations, implementation of electronic health records, and a need to define the population being served. Hence, many vendors and healthcare organizations have switched to a person index as opposed to a patient index.

Financial, marketing, and customer satisfaction initiatives rely heavily on the basic premise that an organization can accurately identify customers being served, determine frequency of customer use of specific services, and determine profitability of services. Further, EMPIs may assist organizations in addressing HIPAA patient identification and tracking requirements.

Types of EMPIs

EMPIs are generally categorized as (1) vendor-neutral or "best of breed" or (2) embedded into a core vendor solution. The first classification implies that the EMPI can be integrated readily with any other vendor application. A core vendor EMPI is generally sold as an inherent or add-on module to other vendor applications, and the EMPI does not readily integrate multiple disparate systems.

EMPIs are commonly deployed in either an active or passive mode using existing Health Level Seven (HL7) messages, with additional data requirements being defined during the vendor selection process and implementation. Many enterprises choose to initially launch an EMPI in a passive mode and then migrate to active mode. Initial business goals, timelines, and budget will determine the deployment method for any given organization.

Active

An active deployment method implies that the EMPI is at the front end of the registration or scheduling process. Thus patient identification is undertaken using the EMPI software, which requires integration of the EMPI and the legacy systems. The user will identify the patient from an enterprise or corporate level, and at a select point in the identification pathway, the user drops to the facility level registration or scheduling system. This process is generally transparent to the user.

Passive

A passive deployment method does not directly impact the registration or scheduling pathway. Rather, identification is undertaken behind the scenes or on the back end of the registration function. Generally, thresholds are established whereby a person is automatically linked with or merged to existing data. If the threshold is not met, the registration data is held in a work queue for later resolution.

Both methods should have the ability to identify persons at a corporate and local level, as initial deployment would involve loading all databases to facilitate initial corporate and local level identification and duplicate identification.

Terminology

- Duplicate—more than one entry or file for the same person in a single facility level MPI
- Overlap—more than one MPI entry or file for the same person in two or more facilities within an enterprise
- Overlay—one MPI entry or file for more than one person (i.e., two people are erroneously sharing the same identifier)

Algorithms

The matching algorithm used for identifying potential duplicates or overlays and for linking various identifiers across the enterprise is a critical component of a successful MPI solution. The algorithm must be sophisticated, powerful, flexible, and accurate. Without a powerful algorithm to support accurate patient identification, the healthcare organization will not meet the objectives identified for the EMPI. The organization will continue to create errors and will be forced to expend considerable time and money on significant manual maintenance efforts. There are three types of matching algorithms available in the industry today: deterministic, rules-based (sometimes known as ad hoc weighting), and probabilistic.

Most hospital legacy information systems use deterministic, or "exact match," algorithms. They require exact matches on a combination of data elements such as name, birth date, gender, and social security number. Deterministic algorithms are considered to be 20 to 40 percent accurate in identification of potential duplicates and often result in a high volume of false matches. Therefore, one could expect that less than half the duplicates are identified by a deterministic method. Deterministic algorithms are particularly weak in identifying individuals when there is transposition of numbers or letters, name changes, limited data, or large databases.

A more sophisticated technique for record matching uses rules-based algorithms. Rules-based algorithms are sometimes referred to as "fuzzy logic," or even mistakenly called probabilistic. A rules-based algorithm allows an organization to assign weights, or significance values, to particular data elements and later use these weights in the comparison of one record to another. This type of algorithm requires the facility to estimate weights in advance and then apply those weights to the data analysis process. Usually, several iterations of trial and error analysis are required until acceptable results are obtained.

Some rules-based algorithms use phonetic searching. The accuracy of a rules-based algorithm varies widely as a result of placing the scoring in the hands of the customer, with ranges of 50 to 80 percent of the potential duplicate record population being identified.

Probabilistic matching is considered to be the most sophisticated technique available, with an accuracy rate of 90 percent or higher. Probabilistic algorithms are based on complex mathematical formulas that actually analyze the facility-specific MPI data to determine the precise match weight probabilities for attribute values of various data elements.

For example, consider an MPI file where the name "Jones" appears much more frequently than "Wheatley." A match on the name "Jones" has lower significance (less likely to be the same person) than a match on the name "Wheatley" (higher probability that a match represents the same person). Supporting digit transpositions and rotations, alternate name cross-referencing, distance editing, and enhanced phonetic searching can further enhance probabilistic algorithms.

Data Integrity

Data has integrity if it is complete, accurate, and consistent. A clean MPI contains only one record, or a unique identifier, for each person. A review of the identified duplicates and overlays often reveals procedural problems that contribute to the creation of errors. The following all affect the quality of MPI data:

- Decentralized registration
- Converted data
- Lack of standards
- Lack of staff training
- Difficulties associated with registering laboratory specimens
- Accepting data from physician offices without verification procedures

A review of the data elements associated with errors may reveal the need to collect new data elements or enforce the importance of accurately collecting existing data elements. For example, missing data from legacy system conversions or incomplete data collection during registration compromises the registrar's ability to select the correct patient. An important aspect of achieving and maintaining MPI data integrity is evaluating these procedural causes of duplicates and other issues affecting the MPI.

An organization should develop standard definitions of MPI data elements (data dictionary), standards for capturing and recording patient demographic data (naming conventions), and performance standards that hold staff accountable for accuracy.

Staff training and review of employee quality and productivity should focus on data quality.

Monitoring of new duplicates is a critical process, and tracking reports should be created and implemented. The HIM, registration, and ancillary departments should establish communications to identify, report, and correct new duplicates. Routine corrections of all identified duplicates should be one of an organization's core MPI maintenance functions.

Methods and Thresholds

The duplicate error rate describes the quality of the EMPI data. The error rate is calculated by the total number of duplicate records divided by the total number of records, multiplied by 100. An error rate is assigned to the EMPI based on the file size and the number of duplicate records identified.

A threshold measure is used to interpret comparison scores. The score is the result of the comparison between two records. Typically, a dual threshold model is used. When scores fall below the lower threshold, the records are assumed to represent different individuals, and the associated medical record numbers or enterprise identifiers are assumed to be accurate.

When scores are above the upper threshold, the records are assumed to represent the same person and an enterprise identifier is automatically assigned. If the score falls between the two thresholds, the record is flagged for manual review. Records with a comparison score in this range create potential duplicate tasks that are placed in a work queue for review and resolution. The threshold reflects the trade-off between potential incorrect linkages within the EMPI and possible duplication.

EMPI Data Overview

Data Elements

Data elements included in the EMPI should:

- Accurately match persons with their single EMPI record
- Facilitate access to longitudinal (lifetime) patient records
- Facilitate linkage with clinical data repositories, pharmacies, and laboratories
- Improve access to patient information resulting in significant benefits for patients and healthcare providers

To achieve these goals, AHIMA recommends a set of core data elements to be included in EMPIs. See "Recommended Data Elements," below. Some enterprises have begun to capture basic clinical data such as allergies, living will information, and privacy notice acknowledgement at the EMPI level.

Data Ownership

The issue of data ownership is a potentially difficult one that organizations must address early in their planning process, particularly if the corporation that is purchasing the EMPI software license does not own all the participating facilities or sites. This challenge is further complicated by the implementation of HIPAA, as organizations must consider the relationship between covered entities, organized healthcare arrangements, business associates, and the obligations for disclosure of information to patients.

Participating for-profit entities of an IDN must consider the Gramm Leach Bliley Act limitations for data sharing.[1] Furthermore, as different facilities are contributing data to the EMPI, the organization should develop a comprehensive strategy to address demographic changes and duplicate resolution. Specifically, it should be determined who will have the authority to change what level of data and how data changes will be communicated.

Enterprise and Corporate Identifiers

Inherent in the deployment of an EMPI is the assignment of the enterprise identifier as discussed in the "Methods and Threshold" section above. While enterprise identifiers can be used for patient care, they are not commonly used by any downstream systems. Instead, they serve as a behind-the-scenes identifier to link and identify persona at a corporate level, with the existing identifiers such as medical record number or account number still providing identification at the local or facility

level. However, with the push for more linking of clinical data to facilitate care across an IDN, organizations and technology may embrace the propagation of the corporate identifier.

Maintaining the EMPI

AHIMA recommends that the responsibility for EMPI maintenance be centralized under the direction of HIM professionals. Employees responsible for EMPI maintenance should be carefully trained, have adequate tools and procedures, and be supervised to ensure their consistent compliance with established guidelines.

A comprehensive maintenance program should include:

- Ongoing process to identify and address existing errors
- Advanced person search capabilities for minimizing the creation of new errors
- Mechanism for efficiently detecting, reviewing, and resolving potential errors
- Ability to reliably link different medical record numbers and other identifiers for the same person to create an enterprise view of the person
- Consideration of the types of physical merges (files, film, etc.) and the interfaces and correction routines to other electronic systems that are populated or updated by the EMPI

Staffing Resources

Adequate staffing is needed to maintain and ensure the quality of the EMPI. Staff members should have the authority to resolve duplicates, investigate demographic overlays, and link persons across the enterprise. A working knowledge of EMPI, facility-level MPI, registration procedures, and duplicate correction procedures is recommended.

Various staffing configurations can be implemented. Consideration should be given to corporate-level oversight of the EMPI and how this may affect the resolution of duplicates at the local level. This should include communication and coordination of efforts across and within facilities. Consistent policies and procedures across the enterprise will greatly aid in reaching EMPI goals, including ensuring the accurate identification of patients and consistent resolution of ambiguous linkages or duplicates.

Education and Training of EMPI Staff

The personnel performing duplicate resolution activities require a foundation in the registration process to facilitate process improvement and ongoing communication. This includes an overview of work flow, obstacles encountered, and department expectations. Additionally, training should include the knowledge of downstream systems that are affected by duplicate records and the resolution steps required in those systems. This training should be performed before staff begins the resolution activity.

The staff that will perform the identification, research, and resolution of duplicates should receive in-depth training. This process should be supported by detailed policies and procedures. Each staff member should be able to demonstrate competency in the areas of duplicate identification and resolution. Ongoing education should also include feedback from a well-defined quality monitoring program. Updates to the training program should be performed periodically and should be based on a review of the initial training to incorporate system modifications and upgrades and internal process improvements.

New patient identification policies and procedures must be formulated prior to launching an EMPI, with careful consideration given to how the registrars will search for a patient if the EMPI is deployed in an active mode, how to interpret scores and weights, and how to select the correct patient. Policies and procedures should be regularly reviewed and updated as systems and processes change.

Recommended Data Elements

Data Element	Definition	Data Type (with HL7 abbreviation)
Enterprise identification number	Primary identifier used by the enterprise to identify the patient	

	across facilities (e.g., the enterprise number or corporate number	
Facility identifier	Primary identifier used by the enterprise to identify the facility contributing data to the EMPI (e.g., the facility code)	
Internal patient identification	Primary identifier used by the facility to identify the patient at admission (e.g., the medical record number)	Extended composite ID with check digit (CX)
Person name	Legal name of patient or person, including surname, given name, middle name or initial, name suffixes (e.g., Junior, IV), prefixes (e.g., Father, Doctor)	Extended person name (XPN)
Date of birth	Patient or person's date of birth. Year, month, and day of birth are entered (e.g., YYYYMMDD). It is essential that the year of birth be recorded as four numbers, not just the last two numbers	Time stamp (TS)
Gender	Gender of patient (e.g., male, female, unknown/not stated)	Coded value in user-defined table (IS)
Race	Race of patient. Race is a concept used to differentiate population groups largely on the basis of physical characteristics transmitted by descent. Races currently used by the federal government for statistical purposes are American Indian/Eskimo/Aleut, Asian or Pacific Islander, black, white, other, and unknown/not stated	Coded element (CE)
Ethnicity	Ethnicity of the patient. Ethnicity is a concept used to differentiate population groups on the basis of shared cultural characteristics or geographic origins. Ethnic designations currently used by the federal government for statistical purposes are Hispanic origin, not of Hispanic origin, and unknown	Coded element (CE)
Residence	Address or location of patient’s usual residence. Components include the street address, other designation (e.g., apartment number), city, state or province, zip or postal code, country, type of address (e.g., permanent, mailing)	Extended address (XAD)

Alias/previous/maiden name	Any names by which the patient has been known other than the current legal name, including nicknames, maiden name, previous name that was legally changed, etc.	Extended person name (XPN)
Social Security number	Personal identification number assigned by the US Social Security Administration	String data (ST)
Telephone number	Telephone number at which the patient can be contacted. This may be a home or business telephone number or the telephone number of a friend, neighbor, or relative	Extended telecommunications number data type (XTN)

Note

1. The Financial Modernization Act of 1999, also known as the Gramm Leach Bliley Act, requires financial institutions to provide customers with a notice of privacy policies and procedures and to satisfy various disclosure and consumer opt-out requirements.

References

KLAS Enterprises, Orem, UT. “EMPI In Depth Analysis.” 2001, 2002.

Hewitt, Joseph B. and Michele O’Connor. “Connecting Care through EMPIs.” *Journal of AHIMA* 73, no. #10 (2002): 32–38.

Hieb, Barry R. “The EMPI Magic Quadrant for 2001: A Maturing Market.” Market Analysis, Gartner Consulting, June 14, 2001. Available at www4.gartner.com/Init.

Wheatley, Victoria. “Unique Identifiers: Preparing for HIPAA.” AHIMA National Convention Proceedings, San Francisco, CA, September 2002.

Prepared by
The AHIMA MPI Task Force:
Lorraine Fernandes, RHIA, Chair
Mary Brandt, MBA, RHIA, CHE, CHP
Donna Fletcher, MPA, RHIA
Karen Grant, RHIA, CHP
Leanna Hatton, RHIT
Susan Postal, MBA, RHIA
Vicki Wheatley, MS, RHIA
Terry Winter, MEd, RHIA, CHE

Acknowledgments
Victoria Barcena, RHIA
Jill Burrington-Brown, MS, RHIA

Article citation:
AHIMA MPI Task Force. "Building an Enterprise Master Person Index." (AHIMA Practice

Brief) *Journal of AHIMA* 75, no.1 (January 2004): 56A-D.