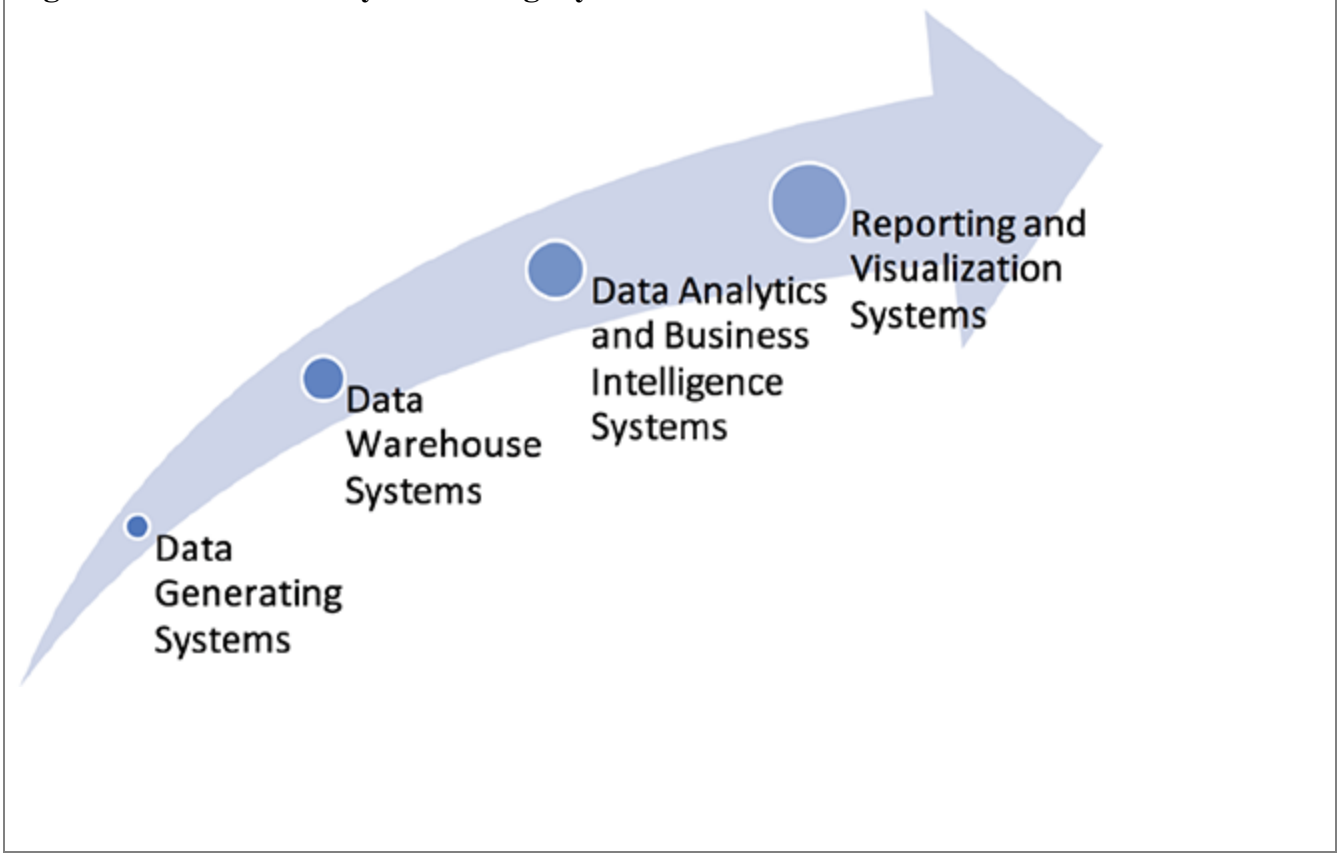
Tracking the Journey of Electronic Health Data through Health Information Systems

Save to myBoK

By Bunyamin Ozaydin, PhD; Shannon H. Houser, PhD, MPH, RHIA, FAHIMA; and Ferhat D. Zengul, PhD

A health information system is any computer system that supports decision-making management in every aspect of the healthcare industry by collecting, processing, storing, and distributing information. Health informatics and information management (HIIM) professionals are on the front lines of dealing with health information systems, and play very important roles in their proper operation. This article discusses the journey of electronic health data in health information systems, the opportunities and challenges throughout this journey, and its implications for HIIM practice.

Figure 1: Information System Category Functions



Information systems consist of a wide range of content areas and characteristics. This article illustrates health information systems from the health data lifecycle perspective. Information systems are categorized based on how each interacts with the health data moving through the system, from the data's inception as a single data point to its maturity as actionable information or knowledge. At the highest level, information systems can be categorized into four major types based on what they do with data. Figure 1 above offers a visualization of the different categories. These four categories are systems that:

- Generate health data
- Aggregate these data into data warehouses
- Perform analytics on the data
- Visualize health data

Defining Data-Generating Systems in Healthcare

Data-generating systems are where data is born. Almost all of these systems are the operational systems designed to automate and document business workflows, and in the meantime capture, process, and display data. The focus of these systems is to manage the day-to-day operations of a healthcare organization. Data generation is a byproduct of these operations. Historically, the databases of these operational systems have not been designed for analytics purposes. Therefore, the analytical quality of the data generated by these systems has been relatively low.

Figure 2: Data-Generating Systems in Healthcare

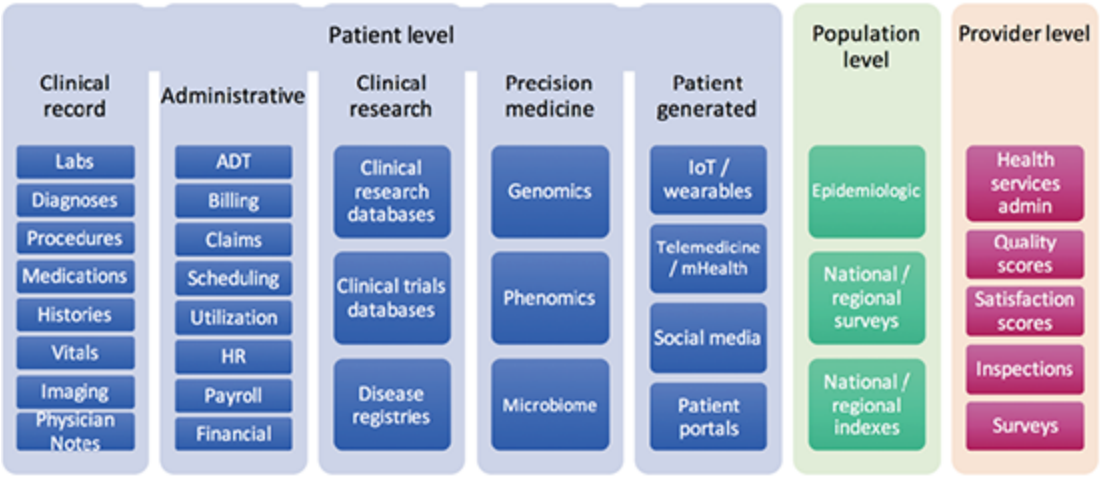


Figure 2 above summarizes the systems that generate health data. From a data perspective, these systems are grouped into three categories:

- Systems that generate data at the patient level
- Systems that generate data at the population level
- Systems that generate data at the provider level

At the patient level, the clinical record is sometimes perceived as the only data source for analytics purposes. The administrative systems data and data collected for clinical research purposes have also been traditionally utilized for analytics. More recently, patient-level data in the categories of precision medicine and patient-generated are also being used for analytics.

At the population level, the data is generated for the national or regional epidemiology data repositories—like the ones maintained by the Centers for Disease Control and Prevention's National Center for Health Statistics—and surveys and indexes created for specific populations or diseases.[1] The provider level encompasses data that is mainly used by health services administration practitioners and researchers. This data is generated for repositories such as those maintained by both government and non-governmental organizations like the Centers for Medicare and Medicaid Services and the American Hospital Association.[2,3]

Data Warehouse Systems Important for Efficient Analytics

Once created in a data-generating system, a copy of the health data continues the lifecycle in a data warehouse. At this stage, the patient-level data generated in the operational systems of an organization are moved into an organization-wide data warehouse system, also known as an enterprise data warehouse. Since the systems that are categorized under population and provider levels in Figure 2 are outside of a single organization, data from those systems are usually aggregated in separate data warehouse systems as indicated in Figure 3 below.

Figure 3: Health Information System Categories

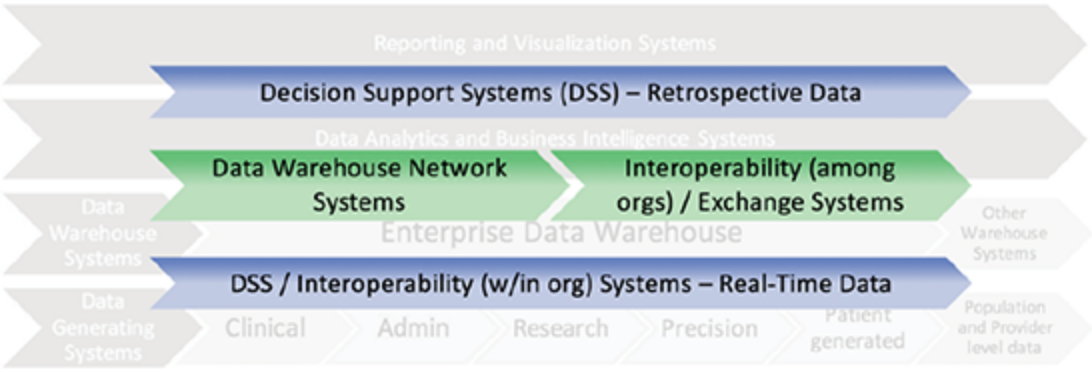
Data warehouses are needed for efficient analytics. First, databases of operational systems are highly normalized to optimize frequent data modifications. However, for analytics purposes, data modifications are much less frequent and database structure needs to be optimized to run queries faster. Second, performing analytics on the operational systems jeopardizes their performance of primary tasks by putting extra burden on computing resources. As the data moves from the operational systems into a data warehouse system, the data coming from multiple systems get integrated via the establishment of relationships among the various bodies of data, thus enabling analysis of integrated data from disparate systems. During the transfer from operational systems to a data warehouse, data goes through a set of processes called Extract-Transform-Load (ETL). In addition to integrating disparate datasets, the ETL processes also perform data cleansing tasks such as data imputation for missing values based on pre-determined criteria.

Health data then get analyzed in data analytics and business intelligence (DA/BI) systems, which rely on integrated and cleansed data provided by data warehouse systems. Furthermore, health data get reported in the reporting and visualization systems that rely on the results produced by the DA/BI systems. Although DA/BI and reporting and visualization systems are often an integrated part of a data warehouse system, depending on the system architecture, they may also run alongside of one.

Other Systems Using Health Data

Alongside the systems described so far, other systems also use health data of different formats. Some of these additional systems are summarized in Figure 4 below.

Figure 4: Other Systems Using Health Data



Organizations with relatively mature enterprise data warehouse implementations are starting to come together to establish data warehouse networks. Such networks are more prevalent with academic medical centers, where networks are beneficial to the institutions' research missions. Since individual enterprise data warehouses vary from one organization to another, there is a need to standardize the data elements to be included in a network. Usually, only a subset of data from an individual organization gets mapped and copied in a de-identified manner into a standard system such as Informatics for Integrating Biology and the Bedside (i2b2).[4] Then, these systems are connected together and establish a network, such as Shared Health Research Informatics Network, which is a network of i2b2 implementations.[5]

There are also systems providing interoperability among multiple data-generating systems within an organization, and others that provide interoperability among organizations that belong to a certain group or geographic location. The former are positioned in-between data-generating systems and data warehouses, because they would have to rely on real-time data for day-to-day operations. These are usually systems that are built in-house and are referred to as an organization's interface engine. Since these serve systems from different vendors, they have to transfer data in a standard manner, such as by using Fast Healthcare Interoperability Resources (FHIR).[6] Health information exchanges—interoperability systems that connect different organizations—are also based on a standardized manner of data transfer, such as a Continuity of Care Record or Continuity of Care Document.

Health data is also used by Decision Support Systems (DSS), whether they are clinical, administrative, or otherwise. Similar to interoperability systems, some DSS depend on real-time data and others require only retrospective data, but depend on the results of DA/BI systems. Therefore, the former have to interface with data-generating systems and the latter would perform above the DA/BI systems as shown in Figure 4.

Implications, Challenges of the Health Data Lifecycle

HIIM professionals have an important role in all stages of the health data lifecycle. Their impact on the data-generating systems is vital for data accuracy, precision, consistency, and timeliness. Therefore, there are several data management implications for HIIM professionals dealing with these information systems:

- **Comprehensive point of view about data.** Health data is not limited to the clinical record, but also includes administrative (billing, claims, etc.) and other types of data. There is a broad spectrum of systems and different types of data that goes beyond this limited perception. It is essential to have a comprehensive understanding of health data to harness its power and develop strategies to overcome business challenges.
- **Challenges with data.** Harnessing the power of data has its own challenges, given that data comes from disparate sources in different formats, such as radiology data as images, physician notes as unstructured text, precision medicine,

and patient-generated data in very large volumes. It is important to recognize that some of the challenges with data is not unique to healthcare. Therefore, healthcare can adopt data best practices from other industries.

- **Real-time feedback into operational systems.** As DA/BI and reporting and visualization systems mature, the next challenge becomes feeding the insight gained from higher-level systems back into the operational systems in real time, bringing new insights to end users. This requires the ability to develop applications that can natively work on any vendor's operational system—platforms like SMART on FHIR offer great promise.[7]

For all these layers of systems to work in an efficient manner, the quality and integrity of the initial data must be ensured. The role of HIIM professionals in data accuracy, accessibility, consistency, precision, and timeliness enables and supports all health information systems and ensures they are used to improve patient care.

Notes

1. Centers for Disease Control and Prevention. National Center for Health Statistics home page. www.cdc.gov/nchs/index.htm.
2. Centers for Medicare and Medicaid Services. Data.CMS.gov home page. https://data.cms.gov.
3. American Hospital Association. Home page. www.aha.org.
4. Partners Healthcare. Informatics for Integrating Biology and the Bedside home page. www.i2b2.org.
5. Partners Healthcare. "Data Sharing Network (SHRINE)." www.i2b2.org/work/shrine.html.
6. Health Level Seven. "Welcome to FHIR." FHIR Release 3 (STU). www.hl7.org/fhir.
7. SMART. "An App Platform for Healthcare." https://smarthealthit.org.

Bunyamin Ozaydin (bozaydin@uab.edu) is an assistant professor, Shannon H. Houser (shouser@uab.edu) is an associate professor, and Ferhat D. Zengul (ferhat@uab.edu) is an assistant professor at the University of Alabama at Birmingham, Department of Health Services Administration.

Article citation:
Ozaydin, Bunyamin; Houser, Shannon H.; Zengul, Ferhat. "Tracking the Journey of Electronic Health Data through Health Information Systems" *Journal of AHIMA* 89, no.1 (January 2018): 40-43.