Al-Mustaqbal University College
Intelligent medical systems department
Class: First
Subject: Introduction to Medical Informatics
Lecturer: Dr. Ali Kareem Abbas
Lecture: (2)

Healthcare Data, Information and Knowledge

Healthcare Data, Information and Knowledge: How to use healthcare data to fuel modern health care organizations.

Significance of Information Systems: how information technology enables patient care, how information technology is used by healthcare providers.

Learning Objectives

- Define Data, Information, and Knowledge

- Understand how vocabularies convert data to information

- Describe methods that convert information to knowledge

- Distinguish informatics from other computational disciplines, particularly computer science

- Describe the differences between data-centric and information-centric technology

Al-Mustaqbal University College
Intelligent medical systems department
Class: First
Subject: Introduction to Medical Informatics
Lecturer: Dr. Ali Kareem Abbas
Lecture: (2)

The amount of healthcare data being generated in today's information age and digital-first economy is unrivaled. The healthcare industry alone generates 30% of the world's data volume. That amount of data leads to incredible opportunities and daunting challenges.

On one hand, if stored, managed and accessed at a high level of sophistication, data can inform exceptional patient experiences.

On the other, that amount of data can create headaches for healthcare providers when the data exists in multiple disparate sources, is missing or outdated, is difficult to access, or is too complex to understand.

Al-Mustaqbal University College
Intelligent medical systems department
Class: First
Subject: Introduction to Medical Informatics
Lecturer: Dr. Ali Kareem Abbas
Lecture: (2)

Rapid digitization, heightened competition, a global pandemic and other disruptive forces have dramatically transformed patient expectations. Today's patients expect personalized, seamless and digitally- enabled experiences across the entire patient journey – from the moment care is needed, to post-operative care. Data must enable and empower every step of the journey. But where do you start?

Case Study: The Story of E-patient Dave

In January 2007, Dave deBronkart was diagnosed with a kidney cancer that had spread to both lungs, bone and muscles. His prognosis was grim. He was treated at Beth Israel Deaconess Medical Center in Boston with a combination of surgery and enrolled in a clinical trial of High Dosage Interleukin-2 (HDIL-2) therapy. That combination did the trick and by July 2007, it was clear that Dave had beaten the cancer. He is now a blogger and an advocate and activist for patient empowerment.

In March 2009, Dave decided to copy his medical record from the Beth Israel Deaconess EHR to Google Health, a personally-controlled health record or PHR. He was motivated by a desire to contribute to a collection of clinical data that could be used for research. Beth Israel Deaconess had worked with Google to create an interface (or conduit) between their medical record and Google Health. Thus, copying the data was automated. Dave clicked all of the options to copy his complete record and pushed the big red button. The data flowed smoothly between computers and the copy process completed in only few moments.

What happened next vividly illustrated the difference between data and information. Multiple urgent warnings immediately appeared (Figure 2.1). Dave was taking hydrochlorothiazide, a common blood pressure medication, but had not had a low potassium level since he had been hospitalized nearly two years earlier.

Figure 2.1: Urgent warning in e-patient Dave's record

Worse, the new record contained a long list of deadly diseases (Figure 2.2). Everything that Dave had ever had was transmitted, but with no dates attached. When the dates were attached, they were wrong. Worse, Dave had never had some of the conditions listed in the new record. He was understandably distressed to learn that he had an aortic aneurysm, a potentially deadly expansion of the aorta, the largest artery in the human body.

Why did this happen? In part, it was because the system transmitted billing codes, rather than doctors' diagnoses. Thus, if a doctor ordered a computed tomography (CT) scan, perhaps to track the size of a tumor, but did not put a reason for the test, a clerk may have added a billing code to ensure proper billing (e.g., rule out aortic aneurysm). This billing code became permanently associated with the record.
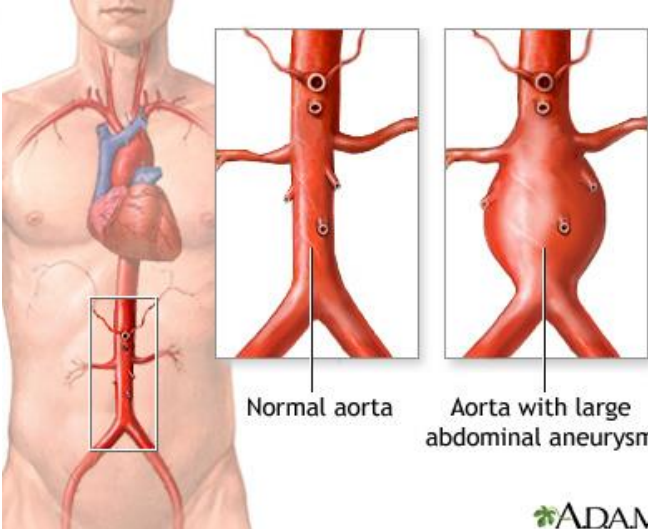
After Dave described what happened in his online blog[2] (http://epatientdave.com/), the story was picked up by a number of newspapers including the front page of the Boston Globe.[3] It also brought international attention to the problem of meaning. It became very clear that transmitting data from system to system is not enough to ensure a usable result. To be useful, systems must not mangle the meaning as they input, store, manipulate and transmit information. Unfortunately, as this story illustrates, even when standard codes are stored electronically, their meaning may not be clear.

Al-Mustaqbal University College
Intelligent medical systems department
Class: First
Subject: Introduction to Medical Informatics
Lecturer: Dr. Ali Kareem Abbas
Lecture: (2)

Figure 2.2: e-patient Dave's conditions as reflected in the newly-created personal health record (PHR)

Data, Information, Knowledge: Definitions and Concepts

- Data are observations reflecting differences in the world (e.g., "162.9"). Note that "data" is the plural of "datum". Thus, "data are" is grammatically correct; "data is" is not correct.

- Information is meaningful data or facts from which conclusions can be drawn (e.g., ICD-9-CM[1] code 162.9 = "Lung neoplasm, Not Otherwise Specified").

---

[1] International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)

Email: ali.kareem@uomus.edu.iq

Al-Mustaqbal University College
Intelligent medical systems department
Class: First
Subject: Introduction to Medical Informatics
Lecturer: Dr. Ali Kareem Abbas
Lecture: (2)

- Knowledge is information that is justifiably believed to be true (e.g., "Smokers are more likely to develop lung cancer").

Data

Computers do not represent meaning. They input, store, process and output zero (off) and one (on). Each zero or one is known as a bit. A series of eight bits is called a byte. Note that these bits and bytes have no intrinsic meaning. They can represent anything.

Bits within computers are aggregated into a variety of data types. Some of the most common data types are listed below:

Integers such as 32767, 15 and -20. And, floating point numbers (or floats) such as 3.14159, -12.014, and 14.01; the floating point refers to the decimal point

Characters "a," and "z" or (Character) Strings such as "hello" or "ball"

Note that these data types do not define meaning. It does not matter whether 3.14159 is a random number or the ratio of the circumference to the diameter of a circle (known as Pi or π).

Converting Data to Information to Knowledge

We live in the real world that contains physical objects (e.g., aspirin tablet), people (e.g., John Smith), things that can be done (e.g., John Smith took an aspirin tablet) and other concepts. In order to do useful computation, one has to segregate some part of the physical world and create a conceptual model.

The conceptual model is used to design and implement a computational model. The real world contains a person, John Smith. There are many other things in the

Al-Mustaqbal University College
Intelligent medical systems department
Class: First
Subject: Introduction to Medical Informatics
Lecturer: Dr. Ali Kareem Abbas
Lecture: (2)

real world including other people, physical objects, etc. There are many things that we can say about this person, they have a name, height, weight, parents, thoughts, feelings, etc. The conceptual model defines what is relevant; everything that is not in the conceptual model is therefore assumed to be not relevant.

1- Data to Information

The next step is to convert data into information. Consider the example, "162.9" is, in and of itself, meaningless (i.e., it is a data item or datum). However, ICD-9-CM gives us a way to interpret 162.9 as "Lung neoplasm, not otherwise specified."

However, associating ICD-9-CM 162.9 with a patient record labels the patient record (and thus the patient) as having "Lung neoplasm, not otherwise specified." Of course, one could design systems that turn data into information without using vocabularies.

2- Information to Knowledge

Multiple methods have been developed to extract knowledge from information. Note that it would not make sense to directly convert data (which by definition are not meaningful) to knowledge (justified, true belief).

Thus, information is required to produce knowledge. Transformation of information (meaningful data) into knowledge (justified, true belief) is a core goal of science.

Clinical data warehouses (CDWs) are described that are often the basis for attempts to turn clinical information into knowledge, as well as methods for transforming information into knowledge.

Al-Mustaqbal University College
Intelligent medical systems department
Class: First
Subject: Introduction to Medical Informatics
Lecturer: Dr. Ali Kareem Abbas
Lecture: (2)

Clinical Data Warehouses (CDWs)

Clinical data are collected via electronic health records (EHRs). Clinical records within EHRs are composed of both structured data and unstructured or (free text).

Structured data may include billing codes, lab results (e.g., Sodium = 140 mg/dl), problem lists (e.g., Problem #1 = ICD-9-CM 162.9 = "Lung Neoplasm, Not Otherwise Specified"), medication lists, etc.

In contrast, free text is simply human language such as English, called natural language processing (NLP). Clinical notes are often dictated and are represented in records as free text in both computer science and informatics.
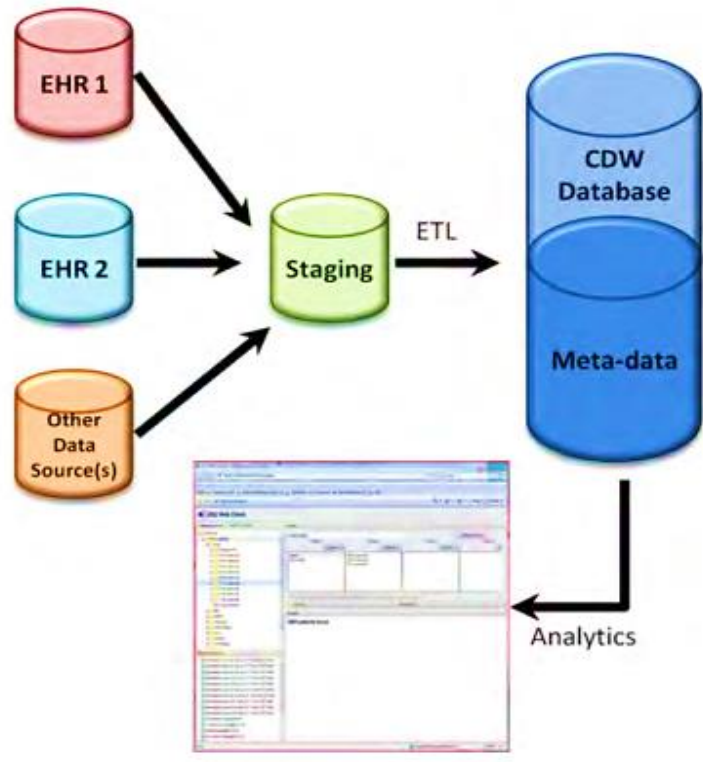
From an informatics perspective, structured data is much easier to manage – it is computationally tractable. Ideally, but not always, these data are encoded using a standards.

EHRs are designed to support real-time updating and retrieval of individual data.

Data from multiple sources including one or more EHRs are copied into a staging database, cleaned and loaded into a common database where they are associated with meta-data.

Once loaded into a CDW, a variety of analytics can be applied and the results presented to the user via a user interface.

Examples of simple analytics include summary statistics such as counts, means, medians and standard deviations. More sophisticated analytics include associations (e.g., does A co-occur with B) and similarity determinations (e.g., is A similar to B).

Al-Mustaqbal University College
Intelligent medical systems department
Class: First
Subject: Introduction to Medical Informatics
Lecturer: Dr. Ali Kareem Abbas
Lecture: (2)

What Makes Informatics Difficult?