DATASKILL

# The Practical Application of IBM Content and Predictive Analytics in Health Care

Prepared for DATASKILL, Inc. by Robert Cass, RN PHN

March 2013

# Forward by Dr. William Bria – HiMSS13

**Woven into the fabric**

By William F. Bria, MD
President, AMDIS

The 2013 Healthcare Information Management and Information Society (HIMSS) annual convention was a busy, informative and insightful time in New Orleans. During the Physicians' IT Symposium, cohosted by AMDIS, I was particularly struck by the realization that we are at the point where the vision of the pioneers in this field has begun to be realized.

Lawrence Weed, MD, gave the last talk of the day during the symposium. When I first heard him speak some 20 years ago about knowledge couplers, I found his ideas offensive. He said back then that being a doctor is not being this walking encyclopedia where you know everything and have the answer to everything. You are a sensor. According to him, doctors are the human linkage with the patient that translates their pains, sufferings and concerns who know how to use resources to get the questions answered.

Twenty years ago, I thought that after eight years of schooling all of that knowledge was part of my training. I also thought that we, as doctors, were independent and stood alone. No, we are not. I have learned that physicians will never have the full compendium of knowledge on any topic. Rather, it is our responsibility to act on behalf of our patients and use the available tools and information to deliver the best care.

Today, I see that Larry was exactly right and experienced a very satisfying feeling of closure. Decades ago, he saw the future of what we do as physicians that is now in effect and being utilized in every setting in which patients are cared for. Being with someone that is truly a visionary is like standing on the shoulders of a giant.

Another thing that struck me during HIMSS13 was the tremendous number of vendors of analytics and next-generation tools. It was satisfying to see this evolution in and maturity of health IT. First, we got people using computers, then we got them to start looking at data. Now, we're starting to benchmark that data to learn and continue to learn to provide better care.

I visited several telecommunications companies and was impressed to see how vendors are making those devices consumer friendly, intuitive and approachable, so routine office visits can give way to routine telemedicine visits. Telemedicine isn't just happening in remote locations anymore. We're realizing that there is no reason to pay to wait for hours in a physician's office when we can accomplish the same things from the comfort of home at a convenient time.

We have arrived as an element of U.S. healthcare. Health IT is no longer the sideshow. With Meaningful Use and other programs that have gotten the full and direct attention of the healthcare community, we are now woven into the fabric of the U.S. healthcare system.

## Executive Introduction

Healthcare information technology and big data have never been as prevalent they are today. Unstructured data, however, persists throughout organizations. The next great challenge is how to cost effectively combine and extract meaning from these disparate data sets so that patient care and hospital operations can be evaluated and acted upon in as close to real-time as possible. [1]

## Introduction

A report delivered to the U.S. Congress in August 2012 defines big data as "a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information." [2]

Premier IBM partner Dataskill, a California based company, has been working hand-in-hand with IBM to build scalable, interoperable, and highly usable solutions leveraging big data through the use of IBM's Content and Predictive Analytics (ICA) big data solution. This white paper explores the challenge as well as Dataskill's proposed solution for a population-based ICA solution for a major academic medical center's cardiac electrophysiology program.

Historically, aggregate healthcare data has been difficult to access due to fragmented data repositories and the unstructured nature of healthcare information. While large healthcare organizations have pursued large, single solution electronic health record implementations in the hope that these solutions would provide a complete data picture of care and operations, structured, semi-structured, and unstructured data remain prevalent in a surplus of applications and data repositories across an organization. These disparate data sources are extremely difficult, if not impossible, to combine and extract meaning from without the momentous effort of data normalization and concept mapping, making this method impractical and extremely costly. In addition, this method does not address the problem of unstructured, document-based physician notes where care provision is primarily documented. [1]

The lack of vision into these documents obscures population-based treatment successes, failures, and outcomes. It is very difficult for physicians and administrators to glean important patient and population-level data for future use. Abstracting data from charts or electronic documents for subsequent evaluation and use is not cost effective or feasible to obtain in a timely manner, even with the use of elaborate data normalization and concept mapping structures.

As such, it is important to evaluate new tools that simplify, streamline, and cost-effectively unite disparate data sets in the healthcare setting. The benefits to patients, caregivers, and hospital operations are too great to ignore any longer.

## The Problem and Opportunity

Data can be found in every nook and cranny of a health organization's wards, units, specialties, operational, and business verticals. Due to the specialized nature of today's healthcare, generalized electronic healthcare record (EHR) vendors cannot keep up with the ever-changing pace of modern medicine with respect to data collection needs. This perpetuates an explosion of specialty data systems, spreadsheets, custom databases, and paper-based data collection tools that do not communicate with the EHR. Trying to develop, test, and maintain integration points of all these systems into an EHR are virtually impossible. There is one place in the record where most of the clinical data is summarized, however. That is the physician notes. Procedure and visit notes describe the specific healthcare encounter in terms of the who, what, where, when, and how of the encounter. They also can provide the opportunity to link specific surgical procedural encounter notes with the initial clinic and subsequent follow-up notes, providing a longitudinal view of a patient's experience across the surgical/procedural episode.

Unfortunately, very few physician procedure and visit notes are created in a structured manner where results, techniques, implants, treatments, and products can be evaluated beyond a single note at a time. Therefore, a method for parsing the encounter notes via ICA and then combining the data from the notes with biomedical and operational data can provide substantial population-based insights into the care and operations provided. ICA makes it feasible to do this in a cost-effective, in real or near-real-time, and consistent way in order to integrate key data into the hospital's and clinician's daily practice, with the endpoint to improve patient care and hospital operations. Research has shown that with data mining techniques, healthcare big data can be leveraged to save as much as $300 billion annually. [1,3]

One example is that of Seton Healthcare, where they are using ICA to identify high-risk congestive heart failure (CHF) patients for interventive care to avoid preventable readmissions[4]. ICA analyzes both structured (lab results) and unstructured data (physician notes, discharge summaries), and Seton can identify and take action on trends and patterns in patient care and outcomes.

Dataskill has been working on a similar solution targeting data related to a Cardiac Electrophysiology Program of a major academic medical center.

## The Solution

Dataskill is currently proposing a system to a major academic medical center (the "University") that will combine PDF-based procedure notes from a specialty system with outpatient visit notes from an EHR. Billing data from the hospital's financial system will also be included, making the system even more valuable. The objective is to take both structured and unstructured data from three separate systems of record, run the ICA NLP engine against the non-structured physician notes set to normalize, categorize, and index the data set. This will allow not only the data to be aggregated and queried, but will reveal

correlations between the data, via ICA's visual tools, that may have been previously unknown. Once the system is in place, the University will be able see which doctors are achieving the best patient outcomes by procedure, how they are doing it, and how much it costs. In addition, hospital coding will be evaluated to find how well the hospital stacks up to standard billing metrics. This allows the University to scan for under or over-billing that can prove very costly via lost revenue or increased audit risks.

Cardiac Electrophysiology, the specialty unit that is the focus of collaboration, is a sub-specialty of interventional cardiology that focuses on the modification of errant electrical activity within the heart. Pacemaker and defibrillator implants, and cardiac ablations represent the majority of the procedures performed in this specialty. Potential sequelae that need post-procedure monitoring include post-op bleeding, infection, stroke, anticoagulant blood level monitoring, and recurrence of arrhythmia, among others. The ability to predict which patients are at high risk for developing post-procedure complications will result in better patient outcomes and reduced cost to both patient and the University.

**Case example #1:** Afib is an abnormal heart rhythm characterized by rapid, irregular contractions of the upper chambers of the heart and can lead to fainting, blood clots, and even death. [5] Afib patients are routinely prescribed anti-coagulants to prevent strokes caused by blood clots. It will be highly valuable to the University to have a system to identify which patients were at high risk of treatment failure due poor treatment adherence, and which social and/or medical factors influence outcomes. Knowing these factors would provide the University with the opportunity to customize treatment compliance programs based on the patient's individual risk profile. The ICA tools accomplish this by reviewing thousands of visit notes and identifying those with stroke as an outcome, identifying and categorizing treatment adherence risks associated with poor outcomes, and then using this information to make recommendations to the care provider for possible inclusion of the patient into a high-risk treatment adherence protocol. Even better - the tool is self-learning, continues to evaluate treatment outcomes, and will continue to make expert recommendations based on an ever-expanding pool of data.

**Case example #2:** Currently, hospital billers interpret the physician's visit note and billing is based on that interpretation of what transpired during the encounter. Interpretations may vary from individual biller to biller, even for the same document. Additionally, what is reported by the physician can vary from physician to physician, even for the same procedure. Coders are trained to bill only for what is documented. [6] These discrepancies can lead to under and over billing which can produce great losses or increase the risk of audits. By using ICA to evaluate large numbers of procedure and visit notes, categorizing their terms and concepts, then linking them to the billing data, Dataskill will be able to produce population-based patient profiles for particular diagnoses and treatments that have been most clinically and cost-effective. Running these templates against new procedure/visit notes will flag the document if it is outside of the norm, which will help physicians and billers to more accurately document and code. Automated coding is also a possibility, which would result in huge financial savings for the University.

## Conclusion

The health care enterprise is replete with siloed, disparate clinical information systems that contain structured, semi-structured, and unstructured data. The need to capture this data so the health care enterprise can leverage the information to practice evidence-based medicine is real and compelling. The need for natural language processing, retro and prospective analytics, and correlation mapping in the healthcare setting has never been greater. Dataskill has developed an IBM Content Predictive Analytics solution that provides a scalable, interoperable and user-friendly mechanism for extracting meaning and actionable intelligence from disparate data sources and non-structured data.

To find out more about how Dataskill can help your organization with its analytics program, contact Dataskill at (800) 481-3282 or visit the Dataskill website at www.Dataskill.com.

## Works Cited


1) Transforming Health Care Through Big Data: Strategies for leveraging big data in the health care industry. (2013). Institute for Health Technology Transformation. http://www.ihealthtran.com/big_data_in_healthcare.html

2) Hartzband, D.D. (2011) Using Ultra-Large Data Sets in Health Care. 2011. Session (p. 3) e-health policy.org. http://www.e-healthpolicy.org/docs/2011_Sessions/20111005_HARTZBAND_BigData.pdf

3) McKinsey Global Institute. (2011). Big data: The next frontier for innovation, competition and productivity. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

4) "Seton Healthcare uses IBM Content and Predictive Analytics to improve care & lower CHF readmissions", 2 Mar. 2012. IBM Enterprise Content Management. Web. 2 March 2012. http://www.youtube.com/watch?v=bLe1GBs7S3M

5) National Heart, Lung and Blood Institute. What is atrial fibrillation? http://www.nhlbi.nih.gov/health/health-topics/topics/af/

6) Courseware. Hospital billing and coding process: Patient accounts and data flow in the hospital. http://www.coursewareobjects.com/objects/evolve/E2/book_pages/ferenc/pdfs/ch04.pdf