IBM Software
White paper

Information Management

Using IBM Unified Data Model for Healthcare to Maximize the Value of Unstructured Data in a Population Healthcare Management Program

IBM

Medical Inflation

It is acknowledged that the costs of healthcare are increasing at a faster pace than those of other services. Together with factors such as growing populations and an increase in the age demographic this medical inflation puts further strain on health services whose funding cannot grow at a similar rate. This applies when the health services are privately or publicly funded.

New treatments are being discovered all the time for medical conditions that were previously thought to be untreatable or where previous treatments were less effective and had a poor chance of successful outcomes. These new treatments, however, come at a price. The research that leads to such discoveries can be costly and until the treatments become mainstream they have higher operational costs associated with them.

New drug therapies can improve the life expectancy and quality of life for patients with chronic conditions but again at increased cost. Development of new drugs, clinical trials, licensing and having them adopted by national treatment programs takes much investment of time and money. The drug companies must recover this investment in order to make profits and fund further research. Often the initial use of these "high tech" drugs can involve enormous costs that might appear unjustifiable but a patient always wants to have the treatment that has the best chance of a positive outcome.

New medical equipment allows healthcare providers to offer treatment previously unavailable but this equipment can involve a considerable capital expenditure, ongoing maintenance costs and costs associated with training healthcare staff in its use.

These new developments will keep more people alive and in better health for longer after all that is what they are intended to do. An aging population requires more medication and other healthcare interventions. So the result of all this medical progress is that we treat more people with more expensive therapies for a longer time.

Making Healthcare Funding Go Further

Healthcare providers and payers understand that detecting disease early and starting treatment early not only improves the outcome for the patient but also reduces the total cost of care for the patient. For this reason, healthcare programs that focus on early detection and care management can reduce the cost per patient or per head of population for providing good health care.

Screening programs such as those for breast cancer, cervical cancer and prostate cancer have saved many lives and by identifying people with the conditions early, also reduced to cost of treating them.

Care management programs that help patients with a known condition to improve their quality of life while reducing their need for intrusive medical or surgical intervention also reduce the total cost of care. Healthcare management programs that include regular interaction between patient and practitioner can identify and mitigate risk of developing other conditions.

IBM Unified Data Model for Healthcare (UDMH) plays a significant role in addressing these issues by supporting the tracking of patients' medical data, their participation in care management programs, and the enrolment of patients with chronic conditions in the appropriate registries.

Patient Information

Key to all of this, is information about the patient. Traditionally, a patient had reasonably regular contact with their family doctor who maintained records of visits, tests, diagnoses and treatments. Similarly, their pharmacist maintained records of prescriptions dispensed. If the patient had the misfortune to need to be admitted to hospital, more records were maintained there. The problem with this is that the information is dispersed and not easily compiled to give a single comprehensive view of the patient. Even when an attempt is made to bring this information together, some can be missed because of a misspelt name, change of address, and so on.

Electronic health records (EHR) go some way to resolving this problem. If the practitioners caring for a patient have access to and regularly update the patient's electronic health record, then a complete view of the patient can be made available. Electronic health records, like any other formal patient record system tend to include information that is discretely and deliberately added by a healthcare practitioner. So the various practitioners must make a conscious decision to make an entry in the patient's EHR. These entries are made in a very structured format identifying diagnosis codes, procedure codes and medication codes for instance and identifying dates of visits and treatments. What the EHR does not have are the notes, observations and opinions that the practitioners document but do not enter in the EHR. This might be information that is held in the patient's paper medical chart. This unstructured information can contain further critical insights into a patient's condition and the risk of a deterioration in that condition.

In this information age, it is estimated that 80 percent of all data being generated is unstructured so imagine how limiting it is to only have access to 20 percent of the data when making a decision and in healthcare these can literally be life and death decisions.

UDMH supports the capture and retention of both structured clinical data and unstructured data, together with its interpretation.

Care for a Diabetic Population

Diabetes is a chronic condition that is affecting more people around the world. It accounts for a huge proportion of healthcare expenditure and is a condition that if not properly managed can destroy the patient's quality of life and ultimately be fatal. The American Diabetes Association recently reported that the cost in the US of diagnosed diabetes in 2012 was $245 billion (41 percent increase in 5 years) including $176 billion of direct medical costs. Managing the condition and reducing associated risks is obviously a high priority for those who suffer from it and for the government agencies, such as Medicare and Medicaid in the US, and private insurers who pay the costs of care.

A diabetes care program includes regular visits to the patients care team for assessment as well as self-testing and reporting. The patient's self-administered drugs can include oral medications and insulin injections. The healthcare provider or other organization managing a diabetes care program monitors key patient health metrics such as:

- Blood Glucose Levels – The most common way to assess a diabetic patient's blood sugars. Blood glucose levels are checked multiple times during the day by the patient or practitioner to check if their blood sugars are under control. These results are used to calculate how much oral medications or insulin they needs to take to reach/maintain the target blood glucose levels.
- HbA1c Levels measured via blood tests typically every 3 months. The patient's results classify them as having good control, adequate control or poor control depending on the HbA1c levels. Practitioners want to see a consistent or improving level of control.
- Cholesterol – Abnormal cholesterol levels can contribute to cardiovascular disease and diabetics are more prone to having unhealthy cholesterol levels. Regular cholesterol level checks for diabetic patients are important to reduce their chance of cardiovascular disease and premature death.

- Blood pressure – Having high blood pressure and diabetes increases a person's risk of developing heart disease and stroke. High blood pressure is linked to developing diabetic complications such as sight damage, kidney disease and nerve damage so maintaining acceptable blood pressure is another important element of the diabetic patient's care.

These are quantifiable measures and by taking and recording them regularly, the diabetes care team can identify trends and more easily spot where there has been an unusual development. Other regular tests include a foot exam and an eye exam as diabetes can lead to a deterioration of vision and nerve damage.

This information is gathered and recorded in the course of regular encounters between the diabetic patient and their healthcare practitioners. The regularity of these encounters is itself a measure of how well the patient is managing their condition. Missed appointments or tests are an indication of somebody who is not fully engaged with the program and who is at increased risk of deterioration. This, however, is still only a small selection of patient information and it is all very structured.

UDMH includes analytical structures that specifically support the management of diabetes and other conditions.

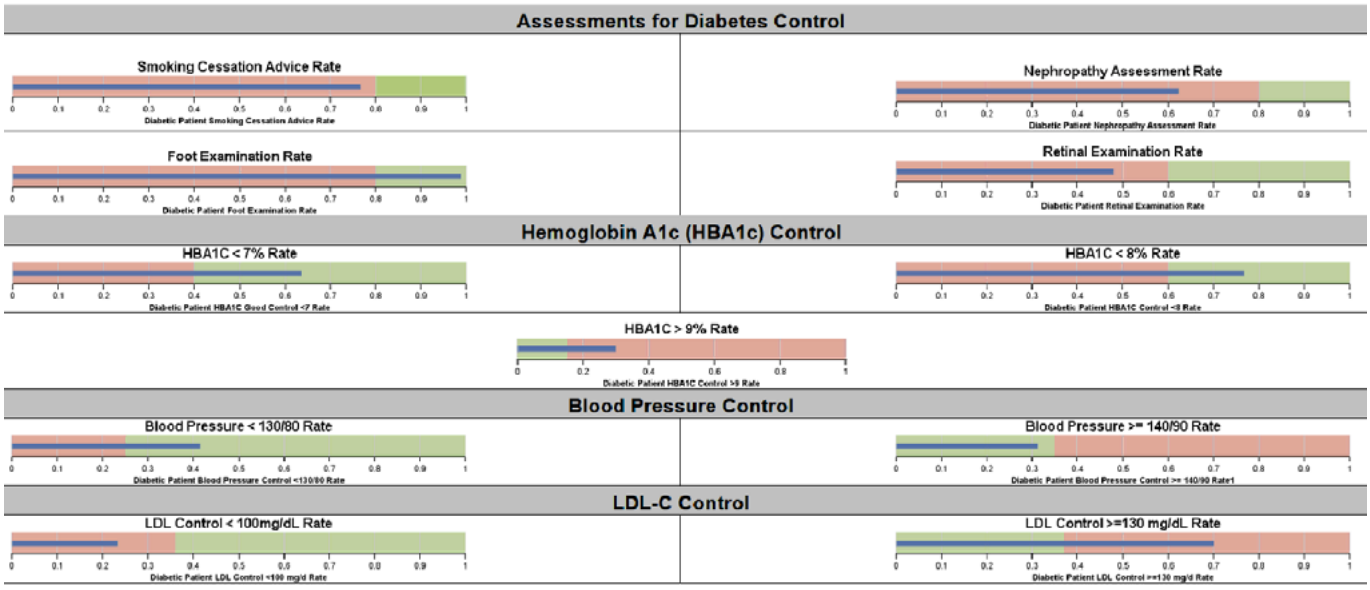

Figure 1. Sample report showing ssessments for diabetes control against targets, for patients with diabetes and who have had strokes

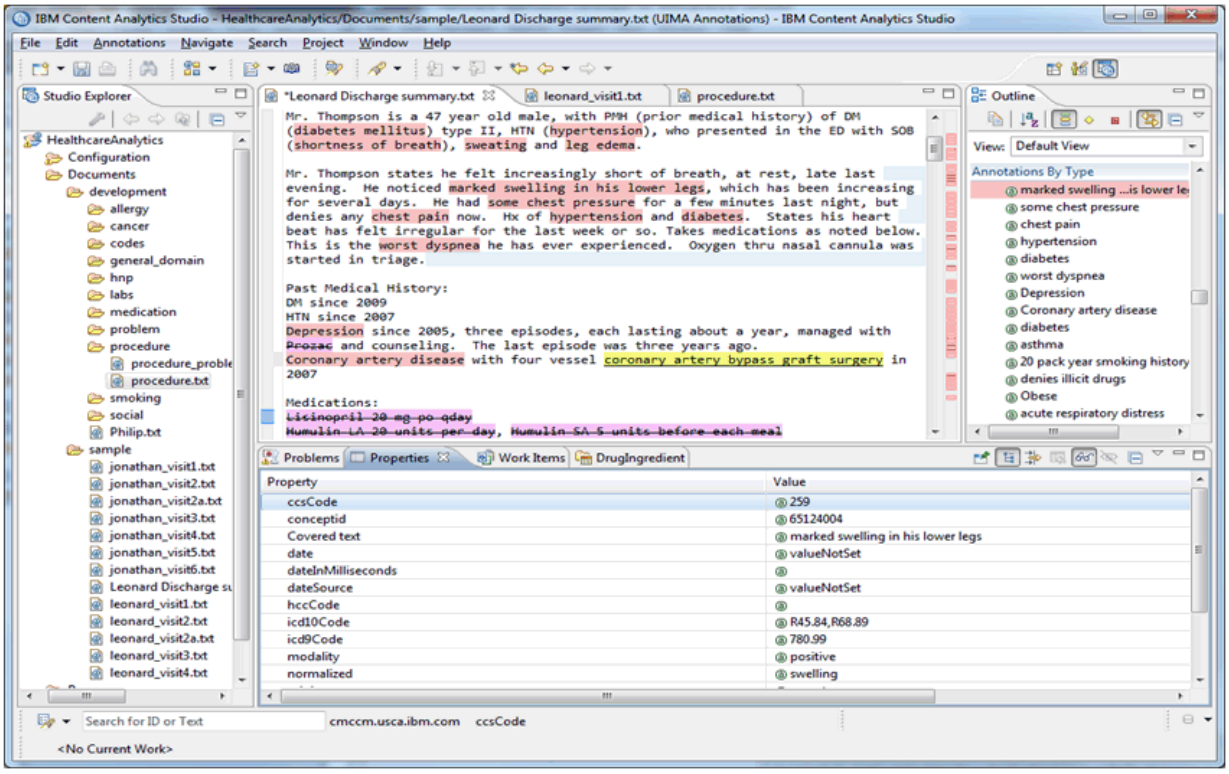

Figure 3. Sample showing the capture of terms from unstructured data

What else might be captured

As mentioned previously, diabetics can self-monitor their blood glucose levels on a daily basis using a small pin-prick test kit (glucometer). The results of these tests can be automatically recorded on their smartphones if it is an electronic test kit or entered via a smartphone app. These results can then be uploaded and sent to their diabetes care team regularly so that the team are not basing all of their decisions on the HbA1c tests that only take place every 3 months or the blood glucose levels that were noted during an practitioner/patient encounter. A similar process can update regular readings from a blood pressure monitor and other tests like urine dip sticks can provide results to be updated via the same channel.

So now the care management team has access to data gathered via the regular scheduled encounters that are part of the program supplemented by self-reported data from the patients. This is still structured data and is based on a set of preselected measures typically targeted for the treatment of diabetes.

Unstructured Data

Depending on the infrastructure within a healthcare facility, practitioners might have to capture their patient's information in an electronic health record and a paper health record. EHRs are known as structured data sources and paper health records are known as unstructured data sources. For example, a practitioner might be able to enter information about the patient's diagnosis and treatment only in the electronic health record and the remaining information might be captured using typed notes, which are added to the patient's paper chart. Paper health records might include clinical notes, which contain concerns raised by the patient such as "difficulty sleeping" or "loss of appetite", which the doctor uses in reaching a diagnosis. They might also record other negative and positive type clinical observations such as "shortness of breath" or "no shortness of breath", which can help to form the diagnosis.

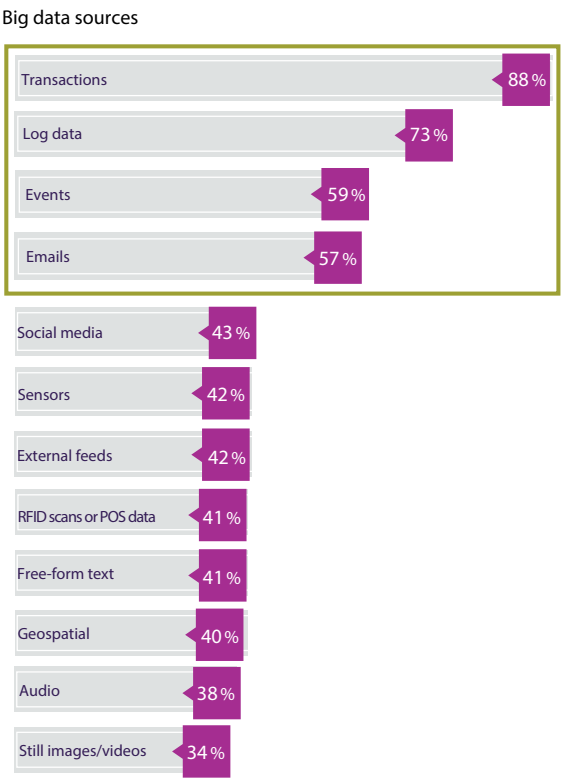

Figure 4. Organizations are mainly using internal sources for big data efforts [1]

The same might apply to discussions about family members' health or changes in diet or exercise. Unstructured data sources have known associated accessibility issues, the simple logistics of being able to physically access a paper medical record is just one example. Accessing the all important data from an unstructured source poses many challenges and this is where Natural Language Processing (NLP) tools can play a huge role in resolving some of these problems.

If the information that is embedded in these unstructured sources can be unlocked, it can greatly increase the healthcare practitioners' knowledge of the patient. To this end, healthcare organizations are now able to use NLP tools to intelligently browse free format text and even handwritten notes to extract key words and phrases in context that can indicate a problem or concern or a change in habit. Similar techniques can be applied to audio recordings of telephone consultations. Extracting meaningful data from unstructured sources in this manner allows a patient's EHR to be updated with the identified data and the source of the new information shown so that practitioners are aware of the distinction between EHR updates consciously entered by a health professional and those that have been inferred from an unstructured source.

But it is not just a case of extracting a few nuggets of wisdom from doctor's notes to increment the EHR content. The ability to store the original material, electronic documents, scanned images of paper documents, clinical images such as x-rays, and voice recordings adds even more patient knowledge. Now a practitioner reviewing a patient's details might notice an earlier reference to a patient concern that was picked up by natural language processing of clinical notes. They can then access the full document from which the information was extracted to better understand the circumstances of the particular

patient encounter on which the notes were based. This type of information can help healthcare professionals, not only treat the patient's primary condition better, but identify early indicators of other comorbid conditions or risks.

UDMH supports the capture of the data interpretations extracted from these sources as well as an assessment of the level of confidence in the interpretation. It also includes links to the source document image so that the original can be reviewed.

Developing a Roadmap to deal with unstructured data

The IBM IBV study "Analytics: the real world use of big data"[1] found that organizations are being practical about engaging with big data as they work to understand how it can be of value to their business. Most are educating themselves on the key use cases, defining a big data roadmap or are conducting pilot implementation activities.

This finding is echoed by IBM Industry Model customers who are looking to extend their existing investments in data architecture tools and processes to help harness the opportunities of big data.

A common theme is the use of big data to enhance and augment existing business intelligence solutions by increasing the volume and variety of data available for analysis.

UDMH provides a flexible and scalable data warehouse design capable of storing all of the comprehensive structured data already in use while enabling the inclusion of new unstructured data sources.

Diabetic Patients with a Comorbidity of Stroke

Diabetic patients are at an increased risk of developing other chronic conditions including heart disease, kidney disease and stroke. Let us assume that our diabetic population contains a patient cohort already diagnosed as having suffered from strokes. The care management for these patients already includes regular monitoring of blood pressure and cholesterol, which is relevant to the management of stroke patients but now needs to be enhanced to cover many other risk factors. A person with diabetes can lead a very normal life provided the correct medication diet and exercise programs are adhered to. A stroke, however, can be very debilitating and can radically change a person's quality of life as it might lead to partial or complete paralysis or severe cognitive impairment.

UDMH allows for the capture of all patient data including chronic conditions, comorbidities and assessments of functional status.

Preventing Further Strokes

A patient who has already had a stroke is at increased risk of having another stroke, which might lead to further damage. By gathering all of the risk factors for a patient that contribute to stroke, the healthcare practitioners can make an assessment of the likelihood of the patient suffering another stroke within a period of, for example, 5 years. Such assessments are based on a series of risk factors decided upon usually by the healthcare organization or responsible health agency and weighted to give an overall risk score. For stroke the risk factors can typically include smoking status, BMI, exercise, diet, blood pressure, and LDL-C as well as many others. Having assessed the risk of patients having another stroke, the healthcare practitioners can identify high risk individuals and ensure that their care management addresses the major issues. Some of these risk factors can be mitigated by medication but many require a change in lifestyle for the individual with smoking cessation often top of the list. This "rule based" risk assessment is a very useful tool in the management of patient health and by identifying and addressing issues for individuals at high risk it can help prevent further strokes, which would require admission to acute care, longer term rehabilitation with all the associated costs and even then, potential life changing damage to the patient.

Predictive Techniques

If we apply predictive analytic techniques to this same cohort of patients it is possible to generate a separate risk assessment based on a wide range of measures. The information inputs required to perform this analysis are wide ranging and includes both structured data measures such as patient vital signs, weight, blood pressure, LDL-C levels, and unstructured data that might be provided via clinical notes or surveys. By analyzing a large body of such information relating to many patients and correlating it with the incidence and timing of the occurrence of strokes, the predictive analytics tool can arrive at a risk assessment that is independent of the rules-based analysis. The more patients and the more data about these patients that is available, the more accurate the predictive risk score is.

By combining the predictive risk score with the rules-based risk score the healthcare practitioners can produce an overall risk score for a patient with a very high level of confidence. Investing a relatively small amount of resource in preventive measures for these patients can save major financial expenditure in the longer term as well as improving the quality of life for those concerned.

By enabling the storage of all patient information in a consistent and easily accessible format, UDMH provides the raw material necessary to support predictive analytics as well as a location to store the predictions and assessments that result.

Risk Assessment for the Wider Population

Having established a model for assessing risk of stroke recurrence in our diabetic with stroke cohort, why not apply the same techniques to assessing the risk of any diabetic patient suffering stroke. The key to this is having the patient data available, which again highlights the importance of having comprehensive patient information and not just the small number of structured measures that are typically tracked on a quarterly basis for diabetes patients. Indeed, the same approach can be adopted for other comorbidities such as heart disease and kidney disease.

Preventive healthcare does not apply only to a patient population with an existing chronic condition like diabetes. If the patient data is available, a risk assessment for major conditions can be processed for any individual. This would involve more wellness programs that aim to improve the health of the population and keep people fit and healthy rather than wait for the diagnosis of disease to start on a reactive healthcare program. Of course, the key to making all of this possible is the gathering and retaining of all patient data in all forms and making this data available as required to healthcare practitioners.

UDMH includes extensive support for population segmentation and also for the use of patient registries, to more easily identify individuals and groups who might benefit from early intervention or screening.

IBM Unified Data Model for Healthcare and the Big Data Platform

IBM Unified Data Model for Healthcare can be deployed to technologies and data appliances such as BigInsights BigSQl, DB2® with BLU Acceleration and IBM PureData™ powered by Netezza, which supports high performance for complex analytic workloads. BigSQL supports federation to many data sources, which allows users to send distributed requests to multiple data sources within a single SQL statement.

IBM InfoSphere® Streams enables continuous analysis of massive volumes of streaming data. The logical data warehouse model can be used to impose a consistent structure on the data as it used for real-time analysis and business decisions.

IBM InfoSphere BigInsights is an enterprise-ready Apache Hadoop-based solution for managing and analyzing massive volumes of structured and unstructured data. Databases can be implemented using IBM Unified Data Model for Healthcare and Hadoop Hive and the data can then be queried by data professionals using BigSQL or analyzed using BigSheets.

IBM DB2, IBM DB2 with BLU Acceleration, PureData/Netezza IBM DB2 database software offers industry leading performance, scale, and reliability on your choice of platform from Linux®, UNIX® and Windows® to z/OS®. IBM DB2 with BLU Acceleration speeds analytics and reporting using dynamic in-memory columnar technologies. IBM PureData System for Analytics (Netezza) is a purpose-built, standards based data warehouse appliance that integrates database, server, storage and advanced analytic capabilities into a single system.

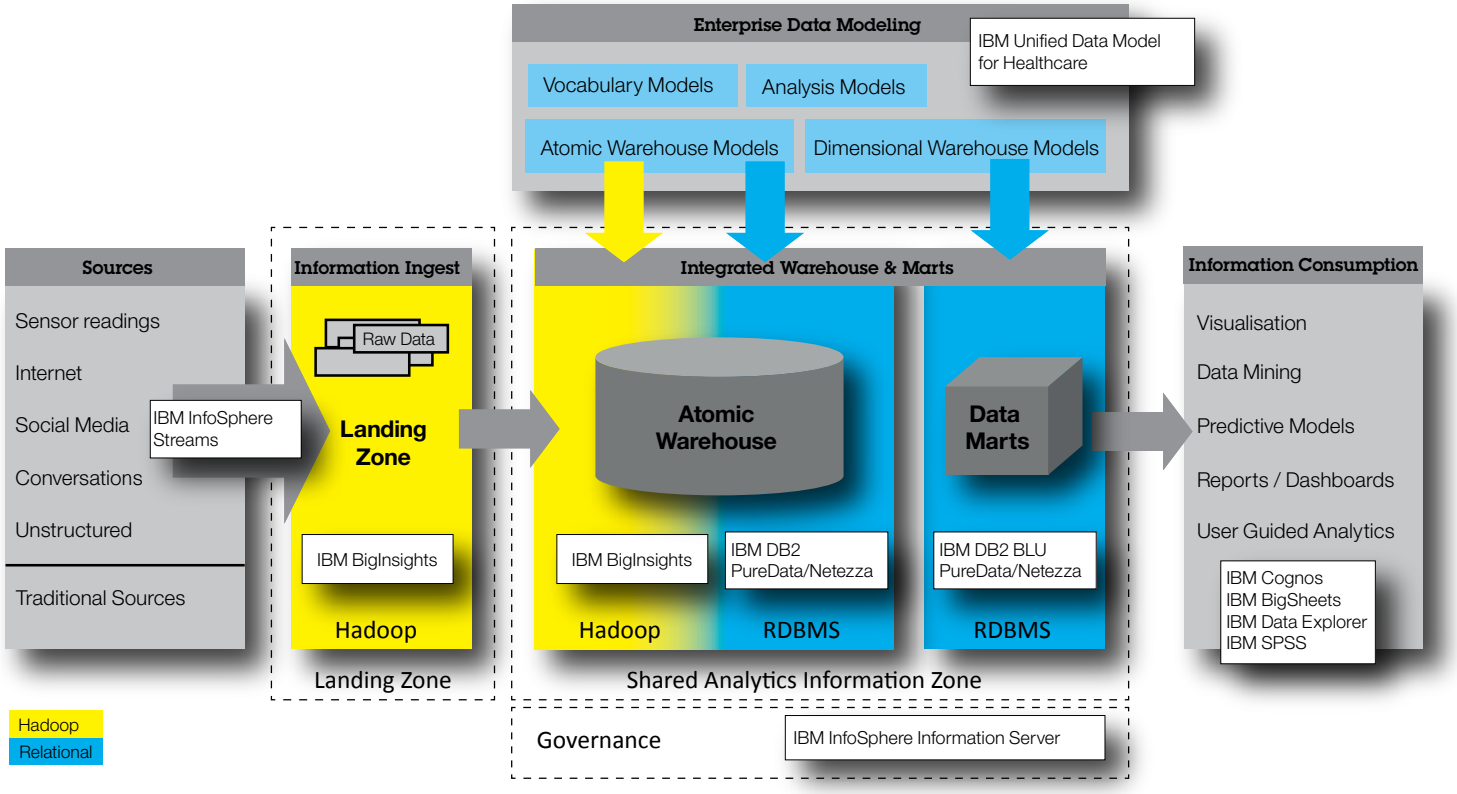

Figure 5. How data models are deployed using the big data platform

IBM InfoSphere Information Server
Business terms are deployed into metadata repository of where they are used to understand, govern and deliver trusted information to business users.

IBM Business Analytics has created a number of solutions that address the unique needs of healthcare providers. Whether it's streamlining operations, improving the customer experience or identifying new opportunities and markets - IBM has the analytic capabilities you need to drive better outcomes. For example InfoSphere Data Explorer discovery and data virtualization can be used for real-time access and fusion of data from across the logical data warehouse and enterprise applications and analysts can use modeling and statistical tools such as IBM SPSS® to create predictive models using data from both Hadoop and the RDBMS.

Conclusion

With healthcare funding under increasing pressure from medical inflation and a growing and aging population it is essential that we use all of the resources available to deliver efficient and effective care. Data is one of those resources and it is one that is largely untapped. Capturing and retaining all data related to a patient including unstructured data and its interpretation enables healthcare professionals to have insight into the person as a whole and in the context of their living arrangements. This leads to more informed and more accurate diagnoses and better health management at a lower total cost. The benefits are widespread:

The **Healthcare Provider** sees better utilization of resources and the ability to treat more patients. They can plan a greater portion of their treatment as proactive care rather than reacting to exacerbations of conditions.

The **Government Agency** paying for healthcare sees reduced costs per person in the programs allowing more of the population to avail of care without increasing public funding from taxpayers' contributions.

The **Private Healthplan** sees better care for its members and improved member satisfaction rates while reducing the cost of claims arising from acute care. This frees up more funding for wellness programs to further enhance the health and quality of life of its members.

The **Employer** sees better attendance from their staff, less unscheduled and disruptive medical absences and a happier, healthier, more productive workforce.

And of course the **Patient** sees an improvement in their health, better quality of life and an improved prognosis in circumstances where they might have seen family members suffer from similar conditions. They spend less time overall in healthcare settings because they invest in staying healthy rather than waiting for a problem to arise and reacting to it.

All of this is made possible by using this largely free resource, data, to its maximum.

IBM Unified Data Model for Healthcare makes this achievable by providing the data storage and analytical structures necessary to maximize the use of this resource.

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
November 2014

Please Recycle

IMW14812-USEN-00

[1] Schroeck, Michael; Rebecca Shockley, Dr. Janet Smart, Professor Dolores Romero-Morales and Professor Peter Tufano. "Analytics: The real-world use of Big Databig data. How innovative organizations are extracting value from uncertain data." IBM Institute for Business Value in collaborations with the Saïd Business School, University of Oxford, October 2012. http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html. ©2012 IBM.