White Paper presented by

Analytic Data Solutions, LLC

Title: Informatics in Research – Improving the Quality of Care

Author: By John Majarwitz, Kishore Nair and Dr. Bernd-Uwe Sevin

Date Written: 2/2/2011

Date Posted (to website): 12/19/2011

Brief Description:

This document will take a deeper dive into the "right hand side" of the reference architecture and examine the Decision Support and Analytics functions and the organization of the data to support these functions. It will also address how to approach the critical concerns for securing private health care data within the architecture. Finally, we will discuss how an information based architecture such as this can be leveraged to drive informatics for improving the quality of health care through evidence based research, patient education and identification of issues before they become problems.

META ANALYTIX

4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410

A Comprehensive Informatics Platform for Healthcare　　　www.metaanalytix.com

Informatics in Research – Improving the Quality of Care

By John Majarwitz, Kishore Nair and Dr. Bernd-Uwe Sevin

Overview

Previously, we published an article, "*Information Based Design – Next Generation Data Warehouses for Health Care Providers*", which identified a reference architecture for creating an enterprise data warehouse environment that delivers early value and builds upon itself, based on the strategy of planning broadly and executing modularly. Within that article, a reference architecture identified the major functions and data flow for the data warehouse environment. This document will take a deeper dive into the "right hand side" of the reference architecture and examine the Decision Support and Analytics functions and the organization of the data to support these functions. It will also address how to approach the critical concerns for securing private health care data within the architecture. Finally, we will discuss how an information based architecture such as this can be leveraged to drive informatics for improving the quality of health care through evidence based research, patient education and identification of issues before they become problems.
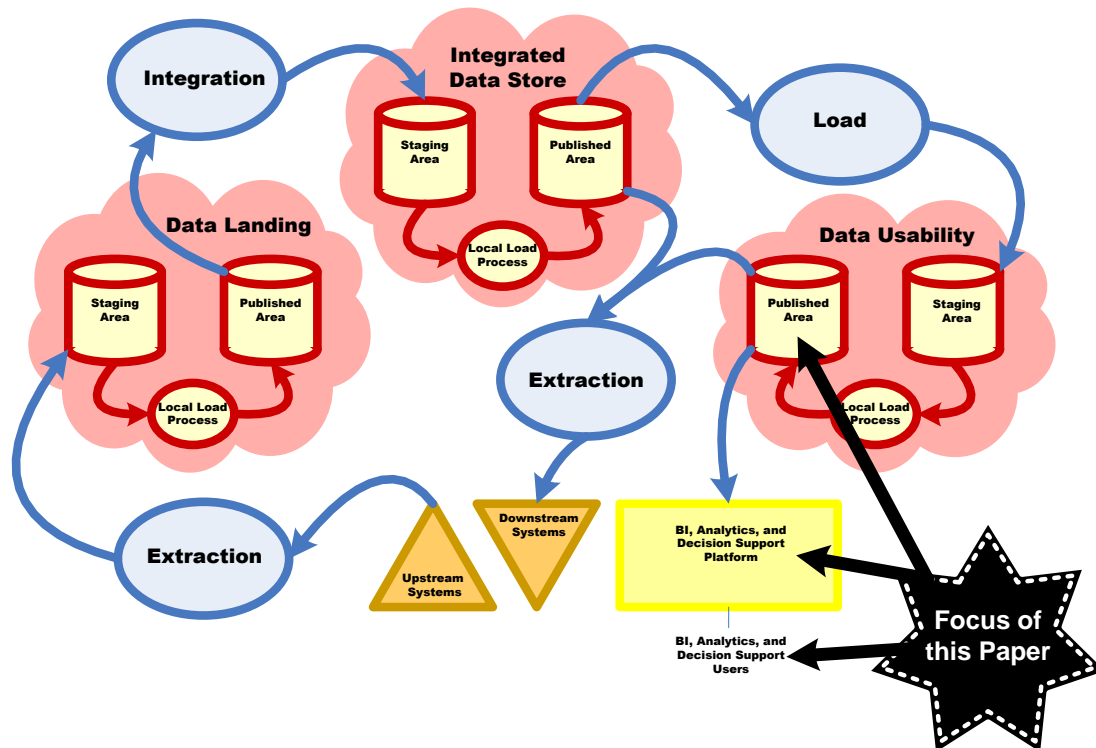

Figure 1: Area of Focus Within the Health Care Data Warehouse Reference Architecture

4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410

META ANALYTIX

A Comprehensive Informatics Platform for Healthcare   www.metaanalytix.com

This paper does not address how data is extracted, transformed, and loaded from upstream systems to make its way into the data warehouse environment (i.e. the "left hand side" of the diagram). That topic will be covered in a follow on paper. For purposes of this paper, let's assume that the means to acquire source data and load the warehouse is present and that all available data resides within the data warehouse environment. This will allow us to focus on how to organize and use the data for decision support and analytics.

Data Usability

If we take the point of view that we want to use data to create information that can lead to insight from which action can be taken, then the Data Usability Zone is the place to start this discussion. Usable data means that it must be: complete for its purpose; of high quality; and organized such that it is easy to understand. It also means that it must be easy to move from broad based analysis to more narrow and specialized analysis and vice-versa. Doing such requires data structures that work together and build upon each other. This type of integration is the reason for the existence of the data usability zone. A comprehensive, high quality, and integrated analytical warehouse (database) is the product of the Data Usability Zone.

Dimensional models are the most prevalent way that data structures are organized in the analytical warehouse. This type of model consists of metrics or measures that are quantified in a central table called a fact table. Surrounding the fact table is a set of dimensions, each containing descriptive attributes that are used to filter and group the measures in the fact table. Each dimension is independent of other dimensions and, as multiple dimensions are used together, they "slice and dice" the measures into smaller pieces. Business Intelligence tools and platforms can interface to these data structures and provide easy to use interfaces for end users. Specialized custom applications can also be created to leverage these data structures. The key design principle here is to de-couple the data from the tools and applications that process the data into reports and analytics views. This allows the data to be collected, integrated, and published once for the enterprise. If the data is coupled with the tool then there will be an individual database for each tool for which the data needs to be collected, integrated, and loaded. This creates considerable duplication of effort, creates opportunity for inconsistency between reports and views of different tools, and inhibits the growth and development of new tools.

Each dimensional model describes a distinct subject area. As such, each model is useful in isolation for the particular information it can provide. But there is also additional power that arises when dimensional models are combined. This is akin to thinking of a fruit stand. Each fruit, such as apples, bananas, oranges, etc. can be consumed alone and provide great satisfaction. But fruits can also be combined into delicious fruit salads which provide more complex flavors and textures. When dimensional models are combined, measures from each of the fact tables are related to one another to support more sophisticated analysis. Just as there are an almost endless number of ways to combine

	4800 Spring Park Road
META ANALYTIX	Box #18
	Jacksonville, FL 32207
	Office: 866.611.8595
	Fax: 877.893.1410
A Comprehensive Informatics Platform for Healthcare	www.metaanalytix.com

fruits, there are lots of ways to group and combine data together from individual dimensional models. But, unlike fruit salad, the combinations can also be combined and this means that results at one level of analysis can serve as input to an even higher level of analysis.

As shown in *Figure 2*, this creates an overall topology where dimensional models can be "stacked". At a foundational level there will be dimensional models that address the most fundamental measures used throughout the enterprise. This level is labeled as Decision Support Marts and provides the basic data at the precision that the sources provide. Examples are visits, admissions, membership, treatments, lab results, pharmacy, inventory, resources, supplies, and equipment.
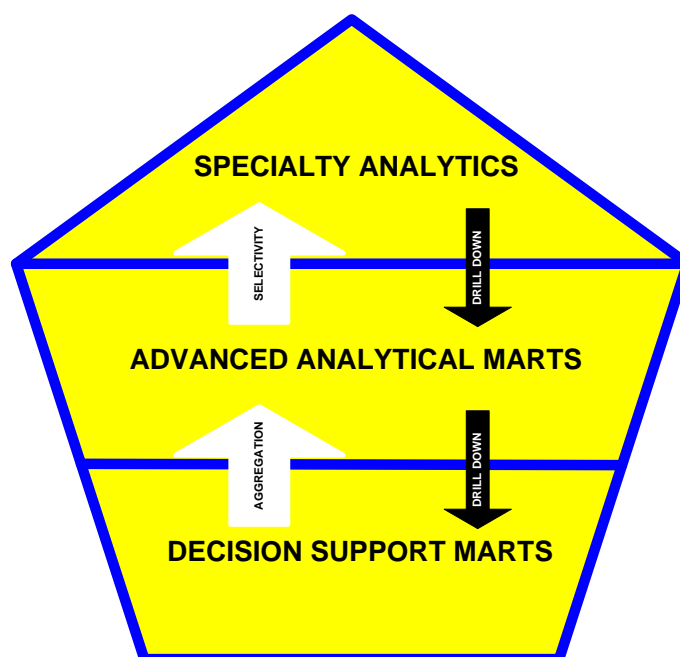
Figure 2: Logical Organization of the Analytical Warehouse

Building upon the Decision Support Marts are additional dimensional models that have a dependency on the base models. These are labeled as Advanced Analytical Marts in the diagram. The dependency to the lower level models means that data aggregation will occur from the lower level to create new information in the higher level. Summing, grouping,, and combining are some of the types of aggregation that can take place. Examples of dimensional models at the Advanced Analytical level are episodes (a series of visits and/or admissions tied to a medical event), outcomes (the eventual result from a course of treatments), occupancy (time view of admissions and length of stay) and, building further, dimensional models that address areas such as care management, utilization management ,productivity, and effectiveness studies can be created, leveraging some of the insight and data from the lower level dimensional models.

4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410

META ANALYTIX

A Comprehensive Informatics Platform for Healthcare   www.metaanalytix.com

Additionally, special studies and targeted research can also be undertaken by combining specific data sets from underlying dimensional models to create specific models that support specialized analytics. This is shown as Specialty Analytics in the diagram.   In these cases, selective sets of the underlying data are typically qualified by certain criteria and a special purpose dataset is generated. Comparative analyses and "what if" analyses are generally performed. The topics of interest include two broad categories: a) Improving Quality of Care and Outcomes and, b) Cost Management and Productivity. Examples of analytics that extend into this area are discussed in the final section of this paper. In all cases, a key capability is having a sufficiently robust source of data to draw from since the questions users are looking to answer may be unknown ahead of time. Thus it is important to create an analytical warehouse with the widest range of data coming from all sources to provide the flexibility and robustness to answer as broad a set of questions as possible with as much precision as possible.

In summary, the ability to "move up the chain" and perform advanced analytics is fueled by taking a data management approach which conforms and aligns data from multiple sources and produces an integrated data store which is used to populate the dimensional models. This removes the burden of data processing from the end user so time consuming tasks such as validating data for completeness and accuracy, translating codes from different systems, and refreshing/updating data are taken care of which allows end users to focus on "consuming" the data by performing decision support and analytics activities. As shown in the diagram, the ability to drill down is also provided since the data is connected (integrated). This allows for investigation into further detail based on higher level results.

Figure 3 below redraws the model for the analytical warehouse with the example dimensional models included to depict how the models are organized.

4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410

A Comprehensive Informatics Platform for Healthcare   www.metaanalytix.com
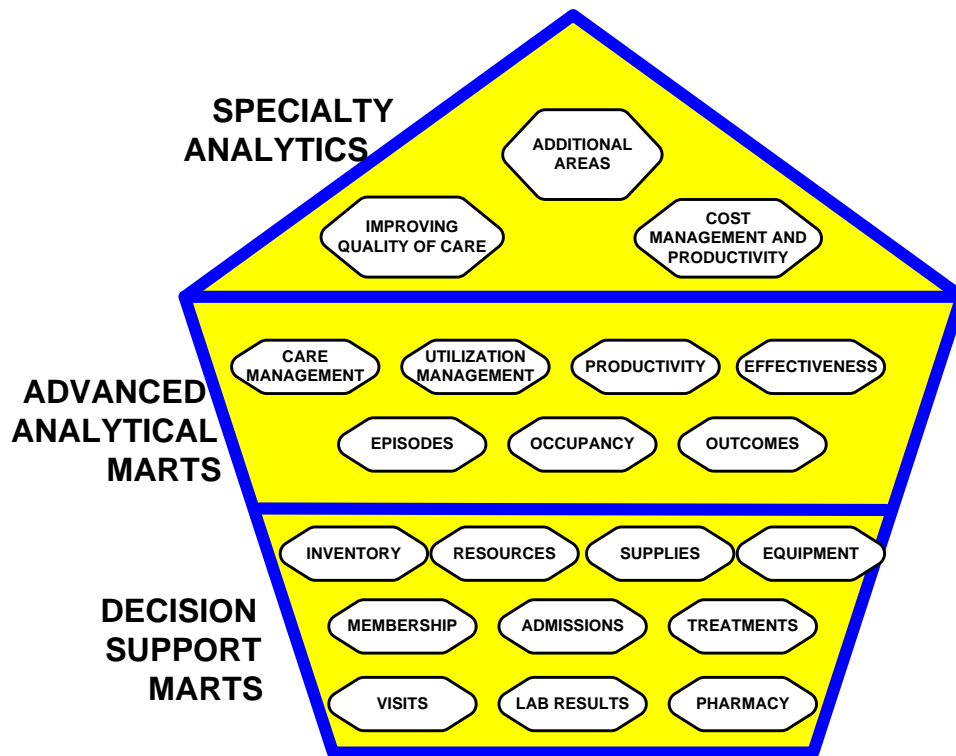
Figure 3: Example Dimensional Models Within the Analytical Warehouse

Privacy

There is an inherent tension between the need to ensure privacy and the desire to want to make data available more broadly for decision support and analytics. Protected Health Information (PHI) consists of a set of 16 data elements that must be kept confidential by providers. Many approaches for healthcare analytics address this by eliminating all private data elements. Yet several of these data elements are useful for decision support and analytics. By viewing privacy as a "trade-off" to having more robust analytical data, opportunity is lost. Instead, a security framework that ensures privacy, while preserving important data elements within the analytical warehouse, can be achieved without increasing risk.

The key concept for maintaining privacy is de-identification of individuals. The role of security and access control is to establish and enforce the policies that will ensure privacy is maintained. Because of the diversity and depth of healthcare data, in addition to traditional security and access control measures that are used to secure data for most enterprises, there are additional measures that need to be put in place because of the potential information that can be inferred when specific data elements

4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410

META ANALYTIX

A Comprehensive Informatics Platform for Healthcare   www.metaanalytix.com

are used together in combination. Altogether, the privacy framework needs to employ a number of security and access control approaches. An outline of these approaches is discussed below.

The easiest way to guarantee data privacy is not to retain it within the analytic environment in the first place. While this is not a realistic approach for all of the PHI data elements, it is applicable to many. To support the analytics described in this paper, none of the PHI data elements listed below need to be retained:

- Names
- Telephone numbers
- Fax numbers
- Electronic mail addresses
- Social security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/License numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Device identifiers and serial numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images and any comparable images

Thus, the vast majority of PHI data elements do not need to be retained within the analytical warehouse. Some of these elements (e.g. natural identifiers such as social security number, name, and account numbers) do need to be used when data is extracted and integrated in order to be able to relate data from different sources, but, once these relationships are established, the PHI elements are no longer needed. This is because the relationships can be preserved using a well established technique in data warehousing, called surrogate keys, in which the real values of identifying data elements are replaced with meaningless numerical values. These meaningless numbers maintain the uniqueness of relationships in the warehouse but they do not have any meaning and, therefore, this replaces the PHI data elements and establishes privacy. (On a technical note, it is important that the algorithm for generating surrogate keys be well designed such that the meaningless values cannot be reverse translated back to original values.) In summary, 13 out of the 16 PHI data elements do not need to be retained in the analytical warehouse.

What about the remaining three PHI data elements? They are:

- All geographic subdivisions smaller than a State

4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410

A Comprehensive Informatics Platform for Healthcare   www.metaanalytix.com

- All elements of dates (except year) for dates directly related to an individual including birth date, admission date, discharge date, date of death
- Any other unique identifying number, characteristic, code, or combination that allows identification of an individual.

These elements provide useful information for decision support and analytics. Let's examine each element to see why it is important and how it can be provided while maintaining privacy.

Geographic subdivisions – Being able to group data demographically is fundamental to decision support and analytics. Making State be the lowest geographic level would prevent being able to perform comparative analytics for regions within a State. For small, homogenous States this would probably be ok, but for larger, diverse states this would inhibit the ability to gain analytical insight. Thus, the lowest geography grain needed by the analytical warehouse would need to be county or zip code level. This means that county and/or zip code are PHI attributes that need to be securely maintained in the analytical warehouse to protect privacy. This can be achieved by not allowing access by tools or users to individual records within the analytical environment. This would prevent anyone from seeing the county or zip code for a de-identified individual. The role of county or zip code would be restricted to serving as a grouping criteria and the number of records grouped would need to be more than one in order to implement this security policy. A privacy filter between the analytical warehouse and the point of access by users and tools can achieve this security requirement.

Dates – Being able to group data by time is also fundamental to decision support and analytics. Making Year the lowest precision would prevent users from being able to perform more frequent comparative analytics, significantly limiting the potential value of the analytical warehouse. Instead, a precision of Day is needed for dates related to visits, admissions, lab results, and other transactional events. This allows for durations to be calculated (e.g., average length of stay, average duration of episodes, average treatment intervals, etc.).   A precision of Year, however, is sufficient for date of birth information which is important to preserving privacy. The exception would be for infants, where a precision of Month, up to a certain age, is probably necessary. As with State above, for dates stored with Day level precision, security would also need to be in place to prevent access to individual transactional records even though they are de-identified. The role of Day would be restricted to serving as a grouping parameter with more than one individual's transactions being grouped. As stated earlier, a privacy filter can achieve this security requirement.

Other Identifying Combinations of Attributes – This is a "safety valve" PHI requirement because it is open ended. It is not possible to prove that combinations of attributes, each of which is not restricted by PHI and each of which is harmless in its own right, cannot be combined in ways that are revealing. It is possible, however, to prevent attributes from being combined if it is known this can be revealing. So the important security requirement is to be able to dynamically restrict combinations of attributes from being used, as part of configuring security, so that, if revealing combinations are discovered, they can be

4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410

A Comprehensive Informatics Platform for Healthcare   www.metaanalytix.com

addressed. Additionally, as stated above, a privacy filter that ensures more than one individual's transactions will be grouped on any attribute for de-identified individual transactions will serve as a strong privacy mechanism.

As stated above, there can be combinations of data elements that reveal identity in certain contexts even though the same elements may be perfectly meaningless to identity in most situations. For example, the combination of occupation and procedure code could be identifying in cases of uncommon occupations and procedures. Or, a diagnostic code can also be identifying when combined with publicly available information – for example, suppose a woman who gives birth to triplets at the age of 40 has an article written about her in the newspaper. She has elected to publicly identify this aspect of her life but, knowing this, a seemingly de-identified record that is associated to a diagnosis code for triplets now becomes identified which means that all other associated diagnostic codes are now revealed to be part of the same identity which is information that the woman did not elect to reveal.

There are an almost endless number of scenarios that can occur where combinations of data elements can become identifying. Yet the solution is not to try to remove or mask these data elements. Doing so would eliminate vital data which can be used in many more ways that isn't identifying. Instead, the solution to maintaining privacy lies is applying privacy filters when access to the underlying data occurs. Not allowing the selection of distinct individual records would help ensure privacy in the examples cited above while also allowing the discrete data elements to be included in counts, sums, and groupings where their contribution is necessary to provide accurate decision support and analytics. As mentioned above, ensuring that groupings contain more than one individual's data is also an important security feature.

When traditional security and access control measures such as user authentication, role based access, password rotation, firewall protection and physical server security are utilized in combination with a privacy filtering solution then a robust solution is in place to assure privacy while also not compromising the availability of data for decision support and analytic functions. *Figure 4* below shows the positioning of the privacy filter in the reference architecture.
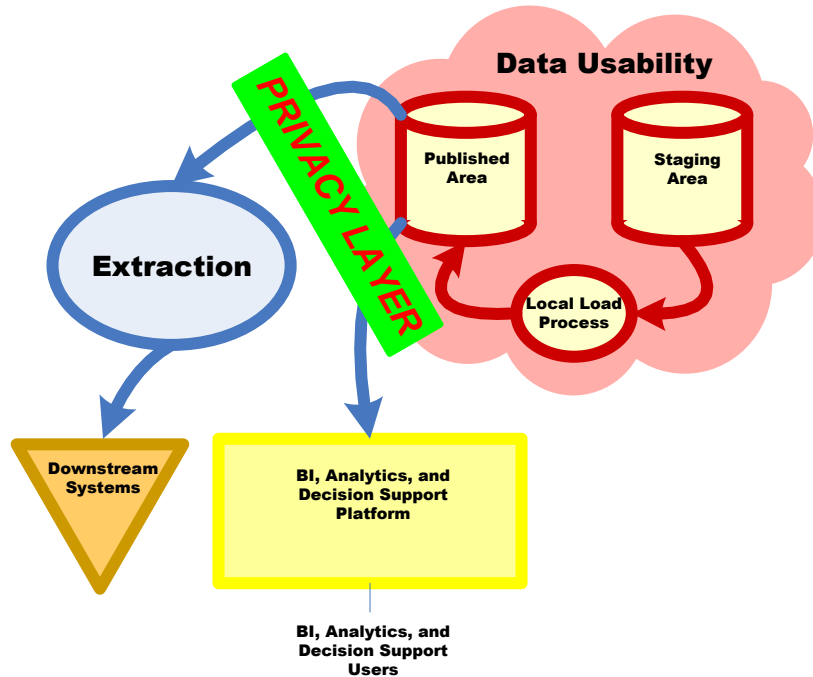
4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410
www.metaanalytix.com

META ANALYTIX

A Comprehensive Informatics Platform for Healthcare

Figure 4: Privacy Layer Governing Access to the Analytical Warehouse

Informatics Scenarios

Earlier in this paper we discussed how the Specialty Analytics function can be valuable for comparative and "what if" analyses relating to Improving Quality of Care and Outcomes, or Cost Management and Productivity. We will provide examples to illustrate how informatics can be applied:

Are We Practicing According To Expected Standards Of Care? – Under this scenario, the analytical goal is to evaluate quality of care as it relates to adherence to established pathways for diagnosis and treatment. This would be a situation where actual care can be compared to established expectations (defined pathways) to determine if there is a variance. If so, a deeper analysis can be performed to isolate the factors that may be at the root cause for these differences. For example, are the variances related to different physicians, certain sites, demographic factors within the patient population, and so forth. In order to perform this type of analysis, the available data must be comprehensive, covering the breadth of available sources over a long time horizon at as detailed a level as is available. Patient demographic data (e.g., gender, age, location), patient health data (e.g., weight, blood pressure, smoker), related patient medical issues data (e.g., co-morbidities such as hypertension, diabetes) are all needed in addition to the laboratory, pharmacy, radiology, specialist, and primary care data for the medical event being treated.

4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410

META ANALYTIX

A Comprehensive Informatics Platform for Healthcare   www.metaanalytix.com

To What Degree Are Patients Complying With Recommended Care? – Under this scenario, the analytical goal is to evaluate quality of care in the context of patient compliance. The analysis would center on determining the gap between the prescribed treatment and the actual patient behavior. Sensitivity analysis can be performed to determine the most important compliance factors that contribute to quality outcomes. As examples, topics such as the effect of preventative care or the effect of following medication schedules can be evaluated. Where non-compliance is observed, deeper analysis can be performed to determine root cause. In cases where complexity of treatment is found to inhibit compliance (for example, different schedules for medication or difficulty of access to treatment facilities) analysis can be performed to determine the extent of these conditions and insight can be gained to help determine mitigation approaches. A related topic is to evaluate what tradeoffs in compliance patients are making in cases when affordability of care is an issue.

Are Resources Being Used Productively? – This is a broad scenario geared toward cost effectiveness and efficient use of scarce resources. Equipment, facilities, supplies, scheduling, and people time are examples of resources that can be analyzed to determine if better utilization can be achieved. For example, analysis can be performed to reduce slack and align dependent resources to yield greater productivity with no additional investment. Insight can be gained into how to better align processes to eliminate redundancy and waste. Or, by examining trends and accounting for other influences, future projections can be created to identify when existing resource capacities will become overloaded. This can lead to proactive acquisition of additional resources before problems are observed.

Are There Factors In Treatment That Tend To Produce Higher Quality Outcomes? – Under this scenario, "what if" analysis can be performed to determine if a different course of treatment would have likely produced a better outcome. This could be useful when there are viable alternatives that have distinct differences for a course of treatment. By having richer and deeper data available, better corroboration or repudiation of treatment alternatives can take place over time to help best practices in care to emerge more rapidly.

Are Complex Procedures Being Performed Properly? – Under this scenario, the goal is to determine if the proper steps are being performed for complex procedures. Although current operational practices may not collect precise enough data at the point of service to record all the steps in a procedure, the trend is toward better information capture, so, over time, the ability to perform this type of analysis should improve.

These scenarios are just a few examples of how the integrated data in the analytical warehouse can be leveraged. There are many more possibilities. The opportunity to marshal healthcare data to improve quality of care and productivity is upon us. Seizing this opportunity by creating an analytical warehouse that drives decision support and analytics is a crucial step in the information age. Designing a solution

	4800 Spring Park Road
	Box #18
	Jacksonville, FL 32207
	Office: 866.611.8595
	Fax: 877.893.1410

META ANALYTIX

A Comprehensive Informatics Platform for Healthcare   www.metaanalytix.com

that is comprehensive, protected, and secure, will provide a crucial competitive advantage for healthcare providers, now, and for the future.

4800 Spring Park Road
Box #18
Jacksonville, FL 32207
Office: 866.611.8595
Fax: 877.893.1410

META ANALYTIX

A Comprehensive Informatics Platform for Healthcare   www.metaanalytix.com

About the Authors:

John Majarwitz has worked professionally in information systems and technology areas for over 25 years, including as a Chief Architect at AT&T Bell Laboratories, and, currently, as Vice President of Solutions and Technology at Meridian Technologies, Inc, an information technology consulting firm. John also operates End To End Information Solutions, which is dedicated to technical publication and information dissemination, and holds B.S. and M.S. degrees in engineering from Yale and Carnegie Mellon Universities.

Kishore Nair is an award winning technologist and entrepreneur. Kishore has a wealth of knowledge in technology, business intelligence, data management and entrepreneurship. In the past, he has worked at Fortune 100 banking and healthcare companies, leading several multi-million dollar technology initiatives and informatics initiatives. He is currently the CIO at Meta Analytix. He has a Master's Degree from Birla Institute of Technology and Sciences, Pilani, India.

Dr. Bernd-Uwe Sevin **M.D, Ph.D, MBA** is a recently retired Gynecology Oncologist from the Mayo Clinic in Jacksonville, FL. In the past, he has held several positions with this esteemed institution such as Emeritus Consultant, Professor of Obstetrics-Gynecology and is Board Certified in Obstetrics & Gynecology and Gynecologic Oncology from American Board of Obstetrics & Gynecology. He has also served on the Informatics committee for the Society of Gynecologic Oncologists. Most recently, he has served on the Department of Applied Informatics committee for the Mayo Clinic in an advisory capacity. For his work, he has received numerous awards and honors and has published over 100 peer-reviewed papers and research articles and over 170 non-peer reviewed articles and book chapters relating to women's health and Gynecologic Oncology. He received his M.D and Ph.D from Albert-Ludwigs University in Germany and did his Graduate School and Residency at Stanford University. He received his Fellowship in Gynecologic Oncology and Executive MBA program in Healthcare Administration from University of Miami.

AboutMeta Analytix:

Meta Analytix is a comprehensive informatics solutions and consulting firm with a focus on Healthcare providers. Our solutions, IATROMATIX, INTEGRA (the first BI appliance for Healthcare) and Knowledge On Demand (SaaS) provide affordable business intelligence solutions to organizations of all sizes. To learn more, please visit us at http://www.metaanalytix.com