Foundations in Decision Process Improvement

The Premise

As a thought leader and having a fondness for promoting innovation, I would throw down two principle and fundamental requirements that will challenge your status quo, to actually improve your decision making processes. The first is the requirement to build or provide for a **broad linked data infrastructure** in your enterprise. This vision is clearly the essence of a company called Zitgist, LLC. The second is the requirement to take a **new perspective on a financial** model that supports the notion of "Knowledge Capital"[1] as trademarked by Paul A. Strassmann.

Key Performance Indicators (KPI) most often associated with Corporate Performance Management (CPM) are the subject of many marketing and sales presentations by companies peddling proprietary solutions. Some firms have paid dearly for specialized Business Intelligence applications. So, why do companies not see better results? What are the KPIs not telling them about their business and why? I postulate it has to do with decision processes. Decisions are not just made, rather many things go into making decisions and it is these decision processes that could be vastly improved.

Foundations in Decision Process Improvement (DPI) is the conjugation of these concepts. The marriage between the CPM tools that provision KPI data (including information communication techniques i.e., early warning) and an enterprise that has complete integrated access to all data across the enterprise are crucial in order to achieve DPI on an unparalleled scale. Suffice it to say following along with or in conjunction with data integration is interoperability. The Institute of Electrical and Electronics Engineers (IEEE) defines interoperability as:

> "the ability of two or more systems or components to exchange information and to use the information that has been exchanged."[2]

This has been the Holy Grail of the Information Age. You say, "It has never been done before", should be your first reaction. The remainder of this white paper will focus on the foundation aspects to better DPI and the why and how these concepts come together.

Business Intelligence

KPIs come in many forms from simple to complex information formats to highly distilled dashboards and perhaps very stimulating visualizations. Business Intelligence (BI) systems have been the IT industry buzz word now for quite awhile. Far too often, the indicator provides a static value but does not carry with it a certain degree of knowledge that either communicates or infers a reason for the number. The number might represent good or bad news. Sales were up in Utah and flat in Georgia. In many cases the reported statistic while current is shown within a control chart with upper and lower limits. The chart gives a sense of "we must do something" if the number is trending toward a limit month after month. The consumer of the KPI may be required to make a decision but in many cases it is not clear why all of a sudden sales went up in Utah. So the decision is made to call a meeting. When decisions are made a consequence is a result. A factor regarding why the results are not better is directly related to the level of knowledge available when the decision maker encounters the KPI. Ideally, the decision should come from the end user executive charged with making such decisions in his/her area. Instead of calling a meeting where focus might end up on Georgia's sales figures, the decision process should have been robust enough to support another path. The executive should have alternatively decided to check that enough product is in the pipe line to supply the Utah sales territory. A savvy business man or woman may instinctively take the right decision in this case. However, be that as it may and the distractions that flare up in day to day operations such pathways are easily missed.

Decision Streams

Another facet about decisions and the impact the process has on results is the fact that no single decision in a typical day when made correctly or optimally necessarily will reflect on performance. Many good decisions whether small, medium or large over a quarter, season or campaign should and will have a positive exponential effect on

the management of corporate performance. Think of decisions in terms of good and bad streams on the order of a laminar flow water jet. When all the molecules are going in the same direction parallel to each other it is a beautiful thing. A series of bad decisions is the opposite, having negative consequences for CPM.

Are firms doing accounting on the countless decisions that are made each day? Are firms tracking decisions in an aggregated statistical manner over time within the context of categorical attributes such as criticality, impact/risk, relationships, Geo Proximity, valuation, outcomes, etc.? Perhaps only the US military does this in their training and real-time evaluations of the battle space.

Much depends on the type of decision, whether you are talking about programmed or non-programmed decision making activities. There are many good slide presentations on the Web for a general overview. No matter what the decision environment or scope, better information and knowledge brings about better decisions. In most any business today the Occupational Safety and Health Administration (OSHA) has special emphasis on a domain within CPM. In practical terms, a firm that has no policies and procedures in place and decides to implement a corporate wide program, is probably a good decision. It is a good decision because by statistically demonstrating and achieving high standards on an annual basis, liability insurance premiums will stay low, a KPI in CPM. So, the decision to go forward with initiatives causes a cascade of decisions in the firm from top to bottom. Decision accounting comes in the form of monitoring. For example, it might be decided that security and safety audits would be conducted in all areas. As a result of this decision, employees document or count all security issues and safety hazards. The information gathered would cause management to refine their decisions and modify the policy to require 55% of all the problems uncovered will be remediated within 55 days. What good does it do find a problem if you do not fix it? Thus, the statistics trickle back up the chain of command whereby the CPM group sets an ambitious goal for site safety of less than or equal to 10% total recordable incidents computed using the US Department of Labor's formula http://www.bls.gov/iif/osheval.htm for the year. It is no stretch of anyone's imagination to know that if you make your workplace a safer place, you would expect to have fewer employees injured on the job.

These decision streams seem very concrete and straight forward. Of course each year management wants to see this KPI go lower. They might decide to spend money training employees on safety in the areas they work. At some point the statistics may stall. In other words no progress is being made as the year goes on. Unfortunately, the OSHA statistics in the CPM dashboard yields no clues as to why the site is not meeting the more stringent goals. So the question goes out and down the chain of command speculating that somehow the increasing number of recordable incidents is related to the increasing number of first aids reported, more or less as a hunch. This is where the knowledge is missing. The number does not carry with it any notion of correlation. The consumer of this KPI is left undecided as to what to do. Some might say this level of information recorded on the production floor or any business unit for that matter is too granular. In reality, employees may be recording these events but the knowledge is locked up in weekly PowerPoint Presentations. We say it is locked up because it is in a digital form that is not structured data[A]. It is not machine to machine processable and to extract the kind of information that we need at the CPM level. To answer the hunch is manually labor intensive.

In this scenario, decision process improvement may be the mechanism that drives a firm toward zero recordable safety incidents. In order to do this though, the firm must embark upon building knowledge assets. In this case, data recorded on the number of first aids, their type, severity and frequency per type will be better stored as structured data. The asset grows over time from month to month as records are maintained for trending and review. The knowledge asset will be linked to other safety and security knowledge assets, such as near misses, recordable incidents and security.

The Asset

A knowledge asset as a term or concept may have a different frame of reference for those familiar with knowledge engineering. This is so because they are not thinking within the same context of the linked data infrastructure and semantic acculturation proffered by the community of people involved in the research and development of the semantic web. For example, a person may develop a knowledge asset by writing a paper that is a reference to a comprehensive body of information focused on a particular issue. This may include an overview of the topic,

A structured data - is data organized into semantic chunks (entities), similar entities are grouped together as relations or classes having properties for which descriptions of all entities in a group is a schema.

implications for policy, research results, landmark studies; and charts and graphs to clearly communicate the scope of the issue. The asset may contain links to other resources and leading researchers in the field.

In this paper, when we are talking about a knowledge asset, we are beyond data and information. As Tim Berners-Lee says,

> "It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data." [3]

It is literally a web, the analog to a spider web, most often thought of in terms of the World Wide Web, but within the scope of this discussion, derived from graph theory in mathematics or computer science. The theoretical behind the concepts is a mathematical structure used to model pairwise relations between objects from a collection. To be less abstract, the definition of Linked Data used internally by Zitgist follows:

> "Linked Data is a set of best practices for publishing and deploying instance and class data using the RDF data model, naming the data objects using uniform resource identifiers (URIs), and exposing the data for access via the Http protocol, while emphasizing data interconnections, interrelationships and context useful to both humans and machine agents."[4]

We may draw upon a comprehensive body of information and subject matter experts, but in practice linked data requires the Resource Description Framework (RDF)[B] that models first order predicate logic. As Mike Bergman describes in his AI3 blog "What Is Linked Data?":

> "RDF and First Order Logic (FOL) are powerful because of simplicity, ability to express complex schema and relationships, and suitability for modeling all extant data frameworks for unstructured, semi-structured and structured data."[5]

Semantic Knowledge Assets

Decision Process Improvement is necessarily dependent on the **Semantic Knowledge Asset**. Thus, we use this term to differentiate our transformation. At the CPM operational level, knowledge realization may be an extracted fact, a skilled resource or an inferential awareness that is the turn key for a decision that must be a transmutation in to goods and services a customer is willing to pay for, otherwise knowledge and its use in decision making would have little value. It is this very notion that may be the reason decision support systems are less than optimal both in terms of timeliness and precision.

As an initiative to commence the development of Semantic Knowledge Assets (SKA), business owners must accept the second requirement as stated in the 1st paragraph. The idea is that SKA(s) must truly become assets on the corporate balance sheet. Current accounting practices are capital and labor models. Presently, few firms methodically value knowledge assets even though companies create or generate knowledge. The natural aggregation of a valuated SKA(s) is knowledge capital if it is to be a recognized reportable line item in financial statements. In Strassmann's 1998 paper "The Value of Knowledge Capital"[6], he sites the example of Abbott Laboratories accumulation of Knowledge Capital. This exactly parallels with my thinking in terms of mind set to grow your knowledge capital. Knowledge today is paid for and expensed against current profits. Strassmann suggests that such expenses should be treated as the investment in knowledge capital.

To build the enterprise linked data infrastructures, initiatives are going to have to come from the top as a strategic objective. The lower echelons of personnel in the corporation may recognize the potential, but lack the resources and leadership to enable to get it to percolate to the top in most cases. At the department level, people are to entrenched in their duties and day to day operations. And neither will IS departments facilitate or project this into a strategic objective. Building the infrastructure does not require the skills necessarily of IT people. In fact, they may not be able to easily make the mental shift from the client/server or network centric database applications they find themselves supporting today, to the methodologies and tools to build SKA(s).

Consequently, it is not very difficult to understand that if you invest in the development of knowledge capital,

B http://www.w3c.org/RDF

business owners and shareholders expect a return or an increase in the value of their stock. Framing the strategic objective, the future role as it is an integral part of CPM for optimum performance through better precision in the decision process streams, the foundations for decision process improvement will be laid down with linked data infrastructures. The infrastructure development needs the financial model that makes knowledge capital possible. Optimum results at the CPM level in real numbers and better decisions that translate in to the goods and services for customers is the realized return on investment shareholders come to expect.

In review, it can be postulated that having the know how to rewire the enterprise with linked data infrastructures and cutting the clothe of the business schema's representative of ontologies is not enough to close the sale here. After all, most companies barely appreciate information resources management because managing information as a resource is very much immature. The linkage to make this happen must be when expenses to acquire information becomes the means for gaining Knowledge Capital. Managing information as most people accept it today, must shift to being an asset management perspective. As soon as the new financial perspective dawns, Semantic Knowledge Assets will follow as the strategic objective deploys.

Illustrative Cases

A short case in point will follow. Even though people like the view from the top, you have to go where the rubber meets the road. A linked data enterprise will probably take its seeds from existing operational systems. Assuming the virtual servers that host semantic technologies are in place, lets take a look at a decision process before and after a Semantic Knowledge Asset is developed. Two fictitious examples follow to illustrate Semantic Knowledge Assets in action. The first case is more exemplary of a sales and marketing environment. The second example may reside in a regulated manufacturing setting having to do with a filter product.

Commercial Operations - Sales & Marketing

In almost every large business today the enterprise will use one or several mega-software application such as Enterprise Resource Planning (ERP) or Materials Resource Planning (MRP). In Sales & Marketing the Customer Relationship Management or (CRM) is the standard application available from many large vendors as well as some absolutely amazing Open Source Software products like SugarCRM[7]. I have to believe that in the very near future, the face of these separate systems will change drastically. With a linked data infrastructure, think of this in terms of connecting your sales database to your financial systems with your operations database. Instead of giving you this in technical terminology, it is easier to imagine what type of knowledge can be mined from this combination. It is as simple as converting raw data from your CRM database with other applications as mentioned above, into a Resource Description Framework (RDF) known as a triple. Connecting to your business schemas or ontologies if you will, changes the entire competitive picture for the Marketing department. The representation of fundamental concepts such as product lines, market segments, business rules and even perhaps competitive intelligence will allow new insight for making business decisions effecting the revenue for the firm.

This type of Semantic Knowledge Asset is known as the Enterprise Semantic Web. It is internal use only even though some data resources may come from outside the firm. This is much different than data mining terabytes of information and reprocessing it. The result are solutions that let your associates change the business rules related to product classifications, product lines, and product hierarchies against market segments. A common problem that may be handled by spreadsheets and prone to errors is to show how sales results fall into different market segment buckets. With the SKA this is consummated through runtime inference engines because now that data is associated with meaning (i.e., linked). The automated reclassification and total visibility comes as knowledge realization when financial analysts have the ability to run and test different reporting scenarios and instantly see how it impacts the bottom line. This type of scenario was achieved in the past by custom programming, add-ons or retrofit software to many of present day enterprise applications.

SKAs have greater implications for supply chain logistics and exactly where you might best project your sales force. If your firm is marketing 2,500 different products, with the SKAs in place, it will be a much less difficult task to weed out the good and the bad revenue generators and analyze margins. Why you ask? Because presently the type of data most commonly manipulated is known as homogeneous data. Homogeneous means that the data maybe composed of similar or common elements or parts and has a certain uniformity. The type of data most often under utilized is heterogeneous. Heterogeneous means the data is composed of unlike or unrelated elements

complex in structure, consequently the tools in common use for homogeneous data will not work as well. A component part of SKA development is how to take advantage and visualize highly interconnected and typed data. An excellent example of this is a tool from MIT called EXHIBIT. These tools allow the user to develop the perspectives that make the most sense.

Decision Under The Old

In many modern production environments, things can and do go wrong. In regulated industries such as banking, food processing, nuclear energy or pharmaceutical operations, the things that go wrong are known as incidents or non-conformances. Typically, something that is non-conforming must be investigated for two reasons. The first reason is a regulating agency may require the firm to do so by law or license. The second reason is if you want to improve your process, product and quality, you must investigate and evaluate your non-conformances. Six Sigma initiatives to eliminate defects may even be applied as well as LEAN methodologies to maximize process velocity to this end.

There are several names that are congruent for describing the same thing. In Security Management Consulting, workplace violence, Intellectual Property Theft or counterfeit products to name a few are investigated as Incident Reports and usually handled by the Security Director in a firm. In a manufacturing setting defects might be recognized as deviations. In marketing departments, competitive intelligence reports and comparative analysis are used to direct and maintain market share. Whatever you wish to call it, the subject in question probably has something to do as a non-conformance issue.

The US Food and Drug Administration (FDA) is responsible for the Code of Federal Regulations called Current Good Manufacturing Practice[8] (cGMP). Manufacturers must comply with these regulations. If a drug manufacturing process fails to comply with any part of the regulation, it must be reported. A completed incident or deviation report is rich in knowledge. It consists of the investigation, evaluation, root cause, corrective action, preventative action and a product impact and risk analysis statement. In most cases, these are reviewed by many people before final closure. Incidents are expensive from start to finish. They may consist of such things as operator error, a procedural process error, equipment failures, environmental monitoring, out of specification and out of tolerance just to name a few broad categories. The same holds true for OSHA compliance to Federal Regulations within a firm. These are two example areas in a regulated environment. So to be generic, we will refer to all of these as non-conformance reports.

Non-conformance reports are like any other corporate record, they must be preserved with all the associated supporting documentation for many years in the case where auditors, compliance officials or government agencies need to go back and research product safety issues related to a particular product line or batch. One that comes to everyone's mind is the recall on children's toys manufactured overseas containing lead in the paint. A company's repository of non-conformance reports is a vast store house of knowledge locked away, literally. Software applications are used to track quality and the standard implementation is a User Interface front-end for editing and managing the work flow of reports supported by a relational database that is usually tightly controlled for security reasons. Consequently, these issues present a blockage which will require support from an authorized strategic objective to help build the knowledge asset.

The life cycle of a report does not lend itself to automation very well. It is intensely laborious work. The work is an excellent way to become the subject matter expert in what ever business you find yourself in. It brings together many subject matter experts. In processes that are complex, the investigator is required to pull data sometimes from dozens of repositories or silos such as databases, content management systems, including paper records and interviews with personnel.

After the non-conformance report write up is completed, the report usually goes through various comprehensive review processes for correctness, readability, completeness and consensus with a quality assurance compliance group or official that finally closes the report out. Needless to say, many decision streams are involved starting with the principle investigators skills in researching, investigating and composing the report. Reviews by middle and upper management as well as compliance officials must also decide to return it for more investigation, clarification or rewrite. Once again decisions are being made throughout the life cycle. More often than not, reviewers have questions that are pending before their decision to move the non-conformance along to the next level of review.

A very casual question from a reviewer might take hours to answer. For example, an evaluation states that potable water used in a piece of equipment uses a 0.354um filter medium. The reviewer seems to think that it should be 0.3um and inquires with the investigator to confirm the statement in the report write up. The investigator relied upon production technicians for the information because they are responsible for changing out the filters. However, in this case to satisfy the reviewer, the investigator has to locate the filter in inventory, collect the specifications from the supplier or manufacturer and retrieve an official equipment validation document in order to assure the reviewer that in fact it was a 0.354 um filter. Anyone in the LEAN world will tell you that if it takes more than 30 seconds to locate a paper document in your office with the information you are seeking, it is probably taking to long and time is wasted. This is one reason non-conformances are costly. The wasted time in locating information stalls the decision process which highlights a practical area for improvement.

Decisions Under The New

In light of the casual question asked above, lets look at the scenario from a Semantic Knowledge Asset perspective. One reason at least for this example, for building this out as a SKA is so that questions can be asked that are prerequisite to making a decision in that it does not take all day to get an answer. Developing the asset means that within the knowledge domain for which we are tracking and writing non-conformance reports, the linked data infrastructure would consist of an integrated platform perhaps utilizing the Virtuoso Universal Server by OpenLink Software[9]. This provisions the basic features of the Http protocol access to or exposure of data, universal resource identifiers (URIs) and the data interconnections, interrelationships and context useful to humans. From the domain expertise and subject matter experts in the department where the non-conformance occurred, the core knowledge asset is constructed.

The core asset utilizes the tools and integrated technologies of a semantic web. With the recent release of the Upper Mapping and Binding Exchange Layer [8] (UMBEL) and its linkages to existing ontologies, the day has finally arrived when integrated access to all data brings real value for improving the decision process. For example, data residing in spreadsheets can be converted to RDF. The legacy data that is bound up in your quality assurance application for managing non-conformances can be expressed as linked data. The unstructured validation document stuck in an electronic Document Management System can now be exposed as well as the manufacturers filter specification digitally bound in an Adobe[C] pdf file. Linked data can be applied to any formalism, source or schema. It is perfectly suited to integrating data between disparate systems whether that data is "Open", private, proprietary or subscription based.

In the best spirit of interoperability, linked data can be shared or exchanged between two or more enterprises, private parties or departments over a private network or intranet using Http. This is a great flexibility because an inquiry by a person reviewing the report may not have to go back to the investigator for a quick answer. And for an investigator, a good percentage of his time spent before writing the first sentence of a report is in locating the data associated with the component parts of the report. The boundaries where the data islands exist in the enterprise begin to fade with this new capability to access information unconstrained in a coherent and contextual sense and a certain level of confidence.

The non-conformance report, no matter what the category can be abstracted as having many subject concepts. The UMBEL is a repository of 21,000 subject concepts , a broad reference structure that can be accompanied by more specific domain conceptual ontologies to provide focused domain-specific context. In reading the report, the word "filter" may be a subject concept. The filter may not be particularly well identified because its use within the described system is implied. But essentially, the word filter as a Node in UMBEL would resolve to one or more domain specific ontologies and perhaps a selection of named entities. In this example, the named entity was a brand name product. This class of equipment uses a filter described as "Class P Grade Encapulation". We are not so much interested in the filter housing per se as we are in the actual class of the filter. The capsule as a class has specific properties. When these properties are structured data in our SKA taken from the manufacturers technical specifications, we now have the ability to find out (i.e., electronically search) the operating characteristics of the filter. The reliability in terms of accuracy comes from the comprehensive validation testing by the manufacturer to ensure consistent and reliable performance under a range of process conditions and the properties declared by the manufacturer are not ambiguous.

C Adobe is a Registered Trademark of Adobe Systems Incorporated

In Knowledge We Must Trust

In the event that we actually improve the decision process, the knowledge from inquiry must be trusted in terms of a level of confidence in its validity and accuracy or precision. The decision made may only be as good as the degree of confidence we have in the knowledge returned. Thus, when building ontologies, the larger concern within a knowledge domain is that of ambiguity. That is to say, a term may have more than one meaning or a slightly comparable dual definition depending on your world view of things. This is often a conflict between subject matter experts when developing ontologies. Building the SKA should go through a process of validation not dissimilar to methods used in the pharmaceutical industry to confirm that a system is going to work the way its specification and design says it is going to work. In pharmaceutical firms, buildings, equipment, processes, computer systems, etc., all go through validation before government agencies will license them for use in producing product. To qualify these systems, three testing methodologies are used IQ, OQ and PQ; Installation Qualification, Operational Qualification and Performance Qualification, respectively. While the strategy of validation will undoubtedly be different for SKA development and deployment, it should be no less rigorous regardless of whether or not the area of domain must be licensed if we expect our decision process improvement to yield confidence in the intelligence the SKA is able to provide.

A feature that an SKA can provide for is cross-reference to separate linked data endpoints. If the person reviewing the non-conformance report, queries about the capability of the filter, multiple sources may result.

1. The query may return the value from the qualification protocols that were run during validation. This tells us what filter we should be using in the water system equipment.
2. A query may return information from our materials inventory control system located in a materials management system (MRP) application with the commodity number ensuring that technicians are replacing the filter with like-for-like components.
3. It will very likely return the micron value as given in the manufacturer's specification.
4. And, if our SKA is really robust, the query might return information from the equipment preventive maintenance schedule showing when and what filter was replaced.

The knowledge coming out of the SKA confirms the use of the 0.354 micron filter. The confidence level is high enough whereby the reviewer decides not to ask the investigator to verify the statement in the report. In **Appendix I** is a graphic depicting a SKA in bioinformatics from the Quebec Genomics Center on the Semantic Web Atlas of postgenomic knowledge about the human and the mouse. It serves to illustrate in a bioinformatics domain the elements in an SKA and the distributed nature of knowledge from many resources.

Fewer Mistakes

People make mistakes or sometimes might change the truth about facts or statistics. Most of the time, non-conformance reports will have honest mistakes and if the firm is lucky, the mistakes will be recognized by the reviewers as the report goes through the review process. Mistakes are also costly in more ways than one. For instance, a formula used to support a product impact and risk analysis statement if computed incorrectly may lead to a wrong decision regarding the safety of the product resulting in the company discarding it, meaning it does not go to market. Or worse yet, a product was judged to be safe when in fact it was not and it was released by quality assurance for use in further manufacturing processes and ultimately making it to the marketplace. Granted, there is no fail safe system when it comes to human error, however relying upon our knowledge asset for verification purposes may help people decide when it is appropriate to question a set of facts (i.e., re-validate statements made in non-conformance reports) and join in consensus upon the truth of the report.

A decision made by way of a mistake is going to impact CPM. The fact that a single query sourced out four distinct references about a filter in the context of its use is not to supplant our own reasoning, rather its aggregation gives us some inference power for which the reviewer did not have when first encountering the question about the size of the filter in use. All improvements made to the decision process whether that includes the acquisition of better knowledge or reducing the number of decisions will impact CPM in a positive way.

Mining The Knowledge

While this discussion is not a specification for developing the Semantic Knowledge Assets, there are several ways to harvest the knowledge. The query language for linked data is called SPARQL (pronounce "Sparkle"). SPARQL is only applicable to an RDF data graph. Dataviewer applications, semantic web browsers and other tools facilitate the presentation layer. Writing non-conformance reports is not an automated task. However, templated SPARQL queries and other techniques can lead to very efficient and rapid deployment of various Web services and reports allowing investigators semi-automation in gathering information. For example, if the database application for managing non-conformance data was integrated into the SKA, trending across all past or closed deviations would be significantly faster and accurate. Binding instance data with display templates such as Fresnel with a SPARQL template would yield efficiencies as well. From the bioinformatics example in **Appendix I** where data converted to RDF was aggregated from available public resources, from the 65 million nodes graph you can ask a real question: "What characteristics of protein were assigned to genes involved in Paget disease?

Semantic Web Browsers allow for faceted browsing over a collection of structured data assets. This is unlike standard web browsers as these applications let you see information in a different light. Faceted means filtered by type or category such that you get to the information that matters much faster visually instead of by scripting. A faceted browsing example for a commercial real estate firm might be in aggregating real estate listings into a Semantic Bank and is useful for facilities management of real estate assets:

NOTE: Please use Firefox to view the link.

Property Banks

Advanced applications like the bioinformatics Cytoscape has blazed the trail in the life sciences R&D areas. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization. In **Appendix II**, Cytoscape generates a visual that is a knowledge map derived from the SKA production in **Appendix I**.

Difficult Decisions

Most all decision making is not as simple as asking a question. Reasonably difficult decisions involve many variables in time and space. The SKA has very little to do with the complexity of difficult decisions. Using the SKA aids us in pulling together many facts related to the decision problem. In fact, knowledge may be drawn from across several domains. Aggregating the information coming out of our SKAs allows us to combine things into potential actionable intelligence to apply to the problem space of a decision. The databasing of RDF is known as a Semantic Bank[11].Consequently, as you gather knowledge that addresses the multiple variables leading up to a decision, you can store, publish, exchange, browse and manipulate your findings as a collection in the Semantic Bank.

In the security consulting business, this type of aggregation is called intelligence gathering. So no matter whether you are investigating a non-conformance or a crime, the principles are the same. Information of varied importance is retained. Clues start to bring things together and logical conclusions let the investigator build a case. The volume of related knowledge to a case can grow very quickly. Organizing the facts derived from Semantic Knowledge Assets is in itself a knowledge asset that may guide one toward different lines of inquiry. Ultimately, your decision may be to prosecute the crime.

The linked data infrastructure remains the same no matter how many SKAs are developed. After a large multitude of SKAs are in place these assets need to be managed. A Semantic Bank may be a good tool to enumerate or index the assets for corporate wide exposure. And the efficient management of knowledge assets has everything to do with its residual value and reuse, an important component in the valuation of a firms Knowledge Capital.

While the linked data infrastructure facilitates structured, semi-structured and unstructured information, much of the discussion is on leveraging legacy data and business schema's. However, since many firms today are creating knowledge through their research and development organizations, it seems like parlaying this directly into a SKA is

most cost effective early in the R&D process. In fact, some pharmaceutical companies are already realizing that structuring data may enable them to bring new product to market faster. This has implications for everything from research and development, to manufacturing, to pharmacovigilance and to sales and marketing. The shear volume of information in developing a new product often is the culprit for not meeting milestones or requiring additional labor costs. With such large amounts of information comes less understanding. With semantic enhanced data your chances of managing the resource is vastly improved.

The **Foundations for Decision Process Improvement** has touched upon the following key points:

- Better support for alternative decision paths.
- Improvement upon decision streams as laminar flows.
- Linked data infrastructure as a key element to improved decision processes.
- Semantic Knowledge Assets built upon the linked data infrastructure to yield intelligence.
- Valuation of SKAs as investments in Knowledge Capital yields value to a customer via optimal decisions.
- Finding information faster to prevent the decision process from stalling.
- Confidence in the intelligence queried from multiple linked data endpoints.
- Trust in the precision and accuracy of the intelligence from vetted structured resources, vocabularies, ontologies, etc.
- The knowledge asset as a means of verification on subject concepts.
- Multiple sources returned from the SKA and cross-reference may lead to reducing the number of decisions in the stream.
- Presentation not currently available for unstructured data yields new insight in the way we look at decision problems.
- Aggregation of knowledge artifacts necessary for analysis leading to a decision.

The mention of an enterprise that has complete integrated access to all data across the enterprise may imply to some that this bypasses security and access control. This is not the case, not even for the linked data infrastructure as described in this discussion. Access and control happens at the Http access and protocols. In fact these can be enhanced by the linked data servers such as OpenLink's Virtuoso Universal Server. If a given link points to a data object from a source that has limited or controlled access, its results will not appear in the final results graph for those users subject to access restrictions. This is also a component of managing the portfolio of Semantic Knowledge Assets.

Closing Remarks

The most amazing and remarkable thought coming out of this discussion is how absolutely easy it is to deploy the linked data infrastructure in the enterprise. The decision maker with Semantic Web applications can now interact with data at the conceptual level as opposed to the traditional logical level. Data about customers, suppliers, invoices, and orders, stored in almost any existing database can be referenced and presented in RDF form across database instance, machine, and network boundaries. This will afford your company a leadership position via an Innovative and Pragmatic use of Semantic Web Technology. Successfully achieving your initiative toward **Decision Process Improvements** across the enterprise, the Semantic Knowledge Assets facilitates powerful holistic views and traversal of enterprise data relationships whereby making great decisions for Corporate Performance Management ensures Key Performance Indicators promote the growth, resiliency and continuity of your business.

It has been said that "You cannot not manage what you cannot measure", implying the use of metrics is an absolute. Thus, when a strategic objective has been cast, the formulation of the strategy has a clear set of objectives for execution. Critical success factors are then linked to well chosen performance metrics, so-called key performance indicators. CPM is in union with or intersects with DPI. If operational performance was perfect, there would be no need for decisions to make adjustments by management. In fact, metrics alone does not guarantee optimal decisions. The decision maker also may not have the perfect expertise and skills through education or experience to make the best choice. Supplemental to metrics are two precepts derived from SKAs for DPI. Knowledge acquisition involves on the part of the decision maker complex cognitive processes: perception, learning, communication, association and reasoning. Knowledge is also used to mean the confident understanding of the subject with the ability to use it for a specific purpose. Identifying a good balance of critical success factors

should also include the yield from incorporating the use of knowledge assets. A deeper meaning to the metrics and concepts, highly relevant but not necessarily more facts and clarity of the bigger picture will lead to better decisions and less variation over the course and real-time stratagem execution.

Questions And Answers

Q. Fear! By making my data machine-readable, will it lose human comprehensibility?

A. No! It is falsely assumed that if one gains the semantics, you lose the visual understanding. Visual understanding is why so many people like Web pages. Thankfully, there are many ways to see data and work with it. If semantics help to describe the data, then your viewing options become more flexible and compelling.

Q. My company has been in business for 100 years. How in the world do you suggest we go about this?

A. I would suggest engaging people with the expertise and vision in this technology area to gain an orientation. The approach will vary depending upon the dynamics and evolutionary processes your company goes through over time, (i.e., such things as product life cycles, economic expansion or contraction in markets, etc.). Mapping your knowledge domains and potential knowledge assets in context with estimating valuations. In practical terms you would not invest in linked data infrastructures if you are going to sell off one of your business units or sunset a product line. On the other hand, if you are conducting R&D or product development, building the infrastructure will make a better financial case because the codified knowledge has reuse value over an entire product life cycle.

Q. How about my current IT investments?

A. Most IT investments today support their user community with proprietary desktop operating systems and applications. This means you are locked in and not easily adaptable to change nor does it let you improvise to meet new challenges in business. Realize, the semantic web and the discussions about linked data has largely come about with the help of Open Source Software and open standards as bio2rdf.org is a prime example, or Protege for developing ontologies. Consequently, you have to be open minded to the same extend that I am suggesting you to have an open attitude with regard to the valuation in Knowledge Capital. Fortunately, most Open Source Software can run on most of the operating platforms available today. Thus, it is a matter of better leveraging your current IT investment and perhaps doing away with expensive legacy applications as you deploy SKA technologies.

Q. Where are the skilled people in my organization going to come from?

A. In all likelihood, your people are already subject matter experts within the knowledge domains they work in. This question may be more a matter of whether you empower your folks to innovate and integrate. Not all SKA development can be born from one mold. Once you map and design the knowledge asset, it may not be so much a matter of programming skills as it is a task of integration of existing tools. And in so many companies, people are stifled from being creative which is exactly what you need in order to be successful in building the SKA.

Q. You said this was going to be easy, really?

A. I said it was going to be easy. I did not say it would not take a great deal of work. However, I think it will take less time to do than what companies go through today, lets say to bring in outside software applications into a regulated environment which requires extensive validation. By the time all the red tape is done (5 - 6 years) and the associated risk you have deal with, such as a critical component of the system like a server no longer supported by the vendor, the whole thing may have to start over before version 1.0 of the solution was ever deployed. I still think comparatively speaking it is easy given the Open Source availability of tools and companies like OpenLink Software servers.

Glossary of Terms and Acronyms

BI - Business Intelligence

CPM - Corporate Performance Management

DPI - Decision Process Improvement

faceted browsing - 'faceted navigation' gives the users the ability to find items based on more than one dimension, to see breakdowns and projections of the items along different axis, which helps users gather insights about the data they are exploring. The most useful 'faceted navigation' is 'context dependent', meaning that available facets, facet values and their count is based on the current set of results that the user is browsing.

Fresnel - is a vocabulary for displaying RDF. A Fresnel was also a type of lens invented by the French physicist Augustin-Jean Fresnel originally developed for lighthouses.

IEEE - *Institute of Electrical and Electronics Engineers*

KPI - Key Performance Indicators

LEAN - is the optimal way of producing goods through the removal of waste and implementing flow, as opposed to batch and queue.

RDF - Resource Description Framework

SKA - Semantic Knowledge Asset

SPARQL - A query language for linked data

UMBEL - Upper Mapping and Binding Exchange Layer - A lightweight, subject concept reference structure for the Web

URI - uniform resource identifiers

Zitgist - as in: **zeit•geist** \ tsīt **'** gīst \ *n.* [German fr. *zeit* time+ *geist* spirit]. The spirit of the time; the taste and outlook that is characteristic of a period or generation.
as in: **gist** \ jĭst \ *n.* The central idea; the main or essential part of a matter.
pronounced as in:

APPENDIX I

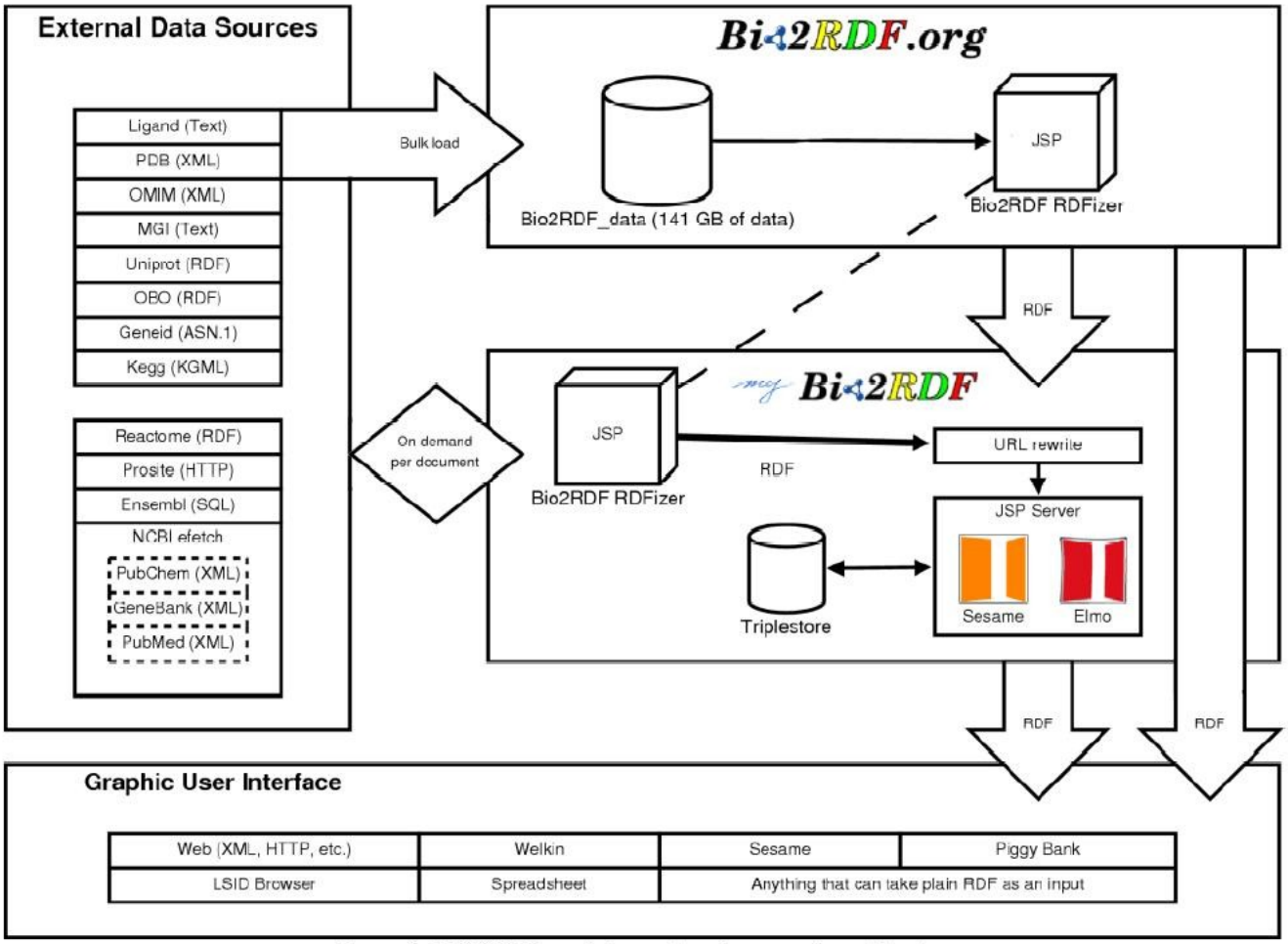
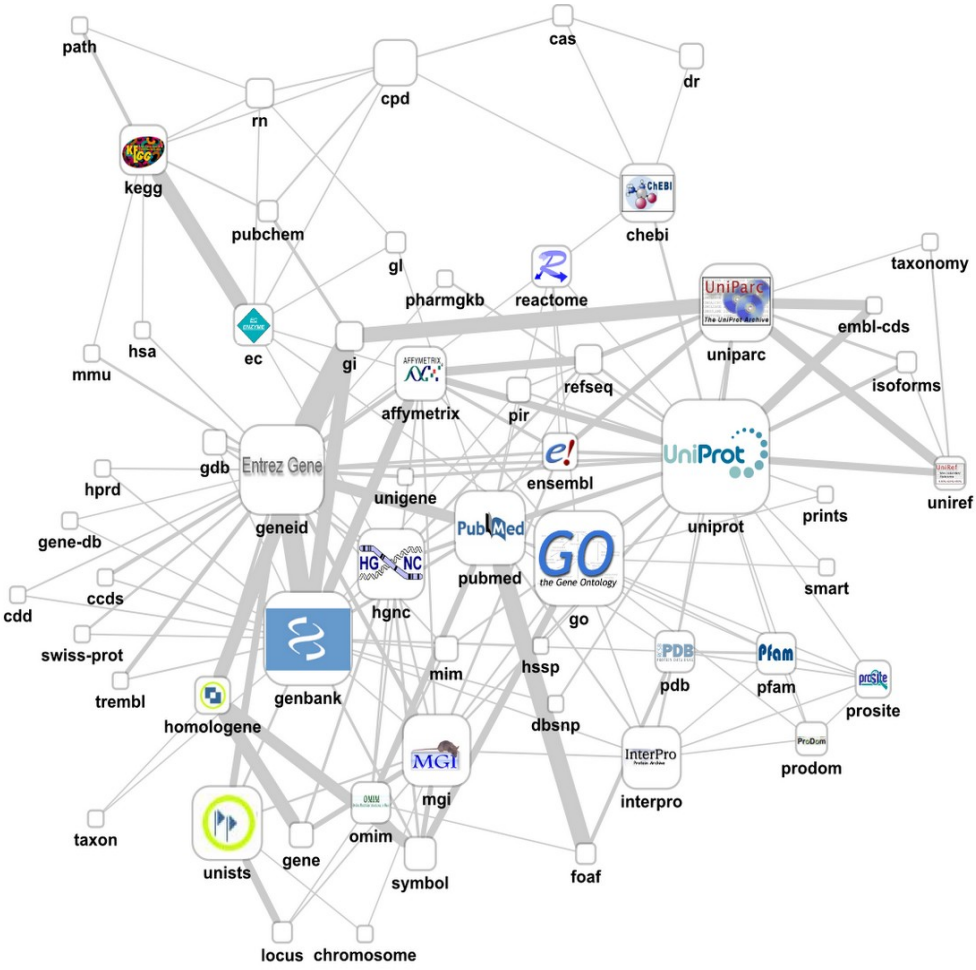


Figure 1: Bio2RDF knowledge system framework architecture

This diagram used by permission and is taken from a research paper by Francois Belleau, Marc-Alxandre Nolin, Nicole Tourigny, Philippe Rigault and Jean Morissette titled "Bio2rfd: Towards a Mashup to Build Bioinformatics Knowledge System".[12] See http://www.ncbi.nlm.nih.gov/pubmed/18472304

Appendix II



This graphic image is used with permission and is taken from the slideshare presentation by Francois Belleau, Marc-Alexande Nolin and Philippe Rigault given at the ISMB2008 Technology Track titled "How to use Bio2RDF with a Virtuoso server. See http://www.slideshare.net/fbelleau/bio2rdfismb2008/

Bibliographic References

[1] Strassmann, Paul A., 1998. *The Value of Knowledge Capital* [online]. Strassmann, Inc.
Available at: http://www.strassmann.com/pubs/valuekc/ [accessed 01 March 1998].

[2] Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, NY: 1990.

[3] Berners-Lee, Tim, 2006. *Linked Data* [online]. W3C World Wide Web Consortium.
Available at: http://www.w3.org/DesignIssues/LinkedData.html [accessed 02 May 2007].

[4] Zitgist, LLC. 2007/2008. *Definition of Linked Data* [online].
Available at: http://zitgist.com/labs/linked_data.html [accessed 2008].

[5] Bergman, Michael k., 2008. AI[3] Adaptive Information, Adaptive Innovation, Adaptive Infrastructure [online].
Available at: http://www.mkbergman.com [accessed 2008].

[6] Strassmann, Paul A., 1998. *The Value of Knowledge Capital* [online]. Strassmann, Inc.
Available at: http://www.strassmann.com/pubs/valuekc/ [accessed 01 March 1998].

[7] SugarCRM, Inc., 2008. SugarCRM Commercial Open Source [online].
Available at: http://www.sugarcrm.com [accessed 2008].

[8] Center for Drug Evaluation And Research, 2008. Current Good Manufacturing Practice (CGMP) Regulations [online]. US Food and Drug Administration. Available at: http://www.fda.gov/cder/dmpq/ [accessed 04 Aug 2008].

[9] OpenLink Software, 2008. *Virtuoso Universal Server* [online].
Available at: http://www.openlinksw.com [accessed 04 Aug 2008].

[10] UMBEL.org, 2008. *Upper Mapping and Binding Exchange Layer* [online]. UMBEL.org.
Available at: http://www.umbel.org [accessed 04 Aug 2008].

[11] Massachusetts Institute of Technology, 2008. Semantic Bank [online]. **S**emantic **I**nteroperability of **M**etadata and **I**nformation in un**L**ike **E**nvironments (SIMILE).
Available at: http://simile.mit.edu/wiki/Semantic_Bank [accessed 04 Aug 2008].

[12] Belleau, F.; Nolin, M.; Tourigny, N.; Rigault, P. & Morissette, J., "Bio2rfd: Towards a Mashup to Build Bioinformatics Knowledge System", Bio2rdf.org.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/18472304