Massachusetts Institute of Technology
Engineering Systems Division

Working Paper Series

ESD-WP-2006-10

ADVANCES IN SUPPLY CHAIN MANAGEMENT DECISION SUPPORT SYSTEMS: POTENTIAL FOR IMPROVING DECISION SUPPORT CATALYSED BY SEMANTIC INTEROPERABILITY BETWEEN SYSTEMS

Shoumen Palit Austin Datta

Research Scientist, Engineering Systems Division
Department of Civil and Environmental Engineering
Research Director & Co-Founder, MIT Forum for Supply Chain Innovation
School of Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
shoumen@mit.edu

July 2006

Massachusetts Institute of Technology (MIT) Engineering Systems Division Working Paper Series [ESD-WP-2006-10]

Advances in Supply Chain Management and Decision Support Systems:

Semantic Interoperability between Systems

Shoumen Palit Austin Datta

Research Scientist, Engineering Systems Division

Department of Civil and Environmental Engineering

Research Director & Co-Founder, MIT Forum for Supply Chain Innovation

School of Engineering

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

shoumen@mit.edu

version 4.0

TABLE OF CONTENTS

1.0 ABSTRACT

Globalization has catapulted ‘cycle time’ as a key indicator of operational efficiency [1] in processes such as supply chain management (SCM). Systems automation holds the promise to augment the ability of supply chain operations or healthcare networks to rapidly adapt to changes or respond, with minimal human intervention, under ideal conditions. Communities are emerging as loose federations or organization of networks that may evolve to act as infomediaries. These changes, albeit sluggish, are likely to impact process knowledge and in turn may be stimulated or inhibited by the availability or lack of process interoperability, respectively. The latter will determine operational efficiencies of supply chains. Currently “community of systems” or organization of networks (aligned by industry or business focus) contribute minimally in making decisions because collaboration remains elusive due the challenges of complexity. Convergence and maturity of research offers the potential for a paradigm shift in interoperability. It may evolve hand-in-hand with [a] the gradual adoption of the semantic web [2] concomitant development of ontological frameworks, [b] increase in use of multi-agent systems and [c] advent of ubiquitous computing enabling near real-time access to identification of objects and analytics [4]. This paper [33] explores some of these complex trends and related technologies. Irrespective of the characteristics of information systems, the development of various industry-contributed ontologies for knowledge and decision layers, may spur self-organizing networks of business community systems to increase their ability to sense and respond, more profitably, through better enter- and extra-prise exchange. In order to transform this vision into reality, systems automation must be weaned from the syntactic web and integrated with the organic growth of the semantic web. Understanding of process semantics and incorporation of intelligent agents with access to ubiquitous near real-time data “bus” are pillars for “intelligent” evolution of decision support systems. Software as infrastructure may integrate plethora of agent colonies through improved architectures (such as, service oriented architecture or SOA) and business communities aligned by industry or service focus (healthcare) may emerge as hubs of agent empires or agencies. However, the feasibility of the path from exciting “pilots” in specific areas toward an informed convergence of systemic real-world implementation remains unclear and fraught with hurdles related to gaps in knowledge transfer from experts in academia to real-world practitioners. The value of interoperability between systems that may catalyse real-time intelligent decision support is further compromised by the lack of clarity of approach and tools. The latter offers significant opportunities for development of tools that may segue to innovative solutions approach. A critical mass of such solutions may seed innovation and spawn the necessary systems architecture for intelligent interoperability, essential for sustainable profitability and productivity in an intensely competitive global economy and in healthcare, agriculture and security. This paper [33] addresses some of these issues, tools and solutions that may be applicable in areas as diverse as adaptive supply-demand networks [7] and innovation in public goods.

Keywords

Interoperability, SCM, Semantic Web, Agent Systems, AI, Supply Chain, Ontology, Business Organization, Adaptive Networks, CPFR, Collaboration, RFID, SDR, Healthcare, Microarray, Sensors, Complexity, Security, NORA, Ports, Customs, Global Public Goods, ebXML

2.0 INTRODUCTION

The ability to connect atoms (physical objects, goods, humans) with bits (data about objects or process) may be described as the Holy Grail of creating a real-time world model where data or information about objects or goods or humans are accessible on demand, anytime, anywhere. Diverse practices, such as healthcare and supply chain management [8] may be viewed as a subset of this over-arching concept of connecting bits to atoms. Real-world supply chains involve the flow of goods (atoms) and data (bits), in various combinations. A more descriptive version may characterize the supply chain as a network of players that transforms raw materials into distributed products [3] and services. Both, healthcare communities and network of supply chain partners, may share processes, data and information through various stages over an extended time frame. These partners are members of a value chain network. Obvious benefits of such collaborative principles were business drivers for the pioneering entrepreneurs of the 1990's who created a variety of e-marketplaces. The demise of several e-markets (for example, SAP Markets) may be rooted in lack of systems interoperability and trust. However, the core principles of e-markets are still viable. In this paper, we shall refer to them, in a generic sense, as community systems or business information networks (BIN). In essence, BIN is a system of information systems (SIS) that may promote sharing of (supply chain, healthcare) information and process knowledge by virtue of being positioned as a subset of an information hub or part of a greater network that may be connected through intelligent data agents to the ubiquitous data bus (UDB). BIN may be divided into: [a] business to business (B2B) collaborations, such as, RosettaNet and [b] information systems necessary for regulatory compliance and security, such as, customs or food and drug administration (US FDA). These systems of information systems (SIS) may act as hubs through which supply chain business partners may share time-critical information in largely event driven, asynchronous modes. Healthcare information systems may share architectural and operational similarities to BIN but use different process model(s).

SCM processes are multi-stage and interdependent [5]. The nature of collaboration implicit in these processes appears to suggest that supply *chains* actually compete with one another [16] rather than individual businesses. Few, if any, centralized supply chains exist where decision making is a shared or collaborative command and control operation between partners. Generally, decisions are made by supply chain partners, aiming to maximize their own profitability. Such decisions are autonomous, spatially and temporally distributed, layered and heterogeneous. However, the gradual dissemination of the virtues of vendor managed inventory (VMI) or collaborative planning, forecasting and replenishment (CPFR), are triggering some forward thinking businesses to explore sharing some data and/or information with supply chain partners, albeit selectively. To profit from globalisation, it is necessary for such collaborative practices and data sharing to occur in endemic proportions if global supply chains wish to respond or adapt to supply-demand fluctuations, often driven by outlier events in far corners of the world.

Evolution of information systems to serve supply chain processes are extensively documented. Electronic Data Interchange (EDI), now considered "archaic" on a technological time scale, is one such medium for inter-enterprise information exchange. Intra-enterprise exchanges are expected to augment the resource view to enable enterprise resource optimization, as claimed by the proponents of early ERP (enterprise resource planning) system developers. Off-the-shelf SCM systems are built to "fit" ERP systems but they often lack functional integration because planning, optimization and execution still are largely disconnected. One reason for the disconnect is that the decision space, in strategic supply chain planning and execution, is plagued with inadequate analytical tools and often lack real-time information. In addition, SCM systems were positioned for decision support within the four walls of the organization (local optimization) but profitability in a highly competitive global economy must respond in near real-time to the challenges of complexity [31] presented by global optimization.

This paper explores supply chains, healthcare and security issues that may benefit from innovation in inter- and intra-enterprise interoperability. Further, the paper reviews how some of the existing technologies (RFID or radio frequency identification, agents and ontology frameworks) may help deliver some beneficial solutions in this space. In effect this paper addresses the elusive quest for adaptable interoperability (Appendix 3) and convergence.

3.0 PROBLEM LANDSCAPE

3.1 Supply Chain

Profitable supply chain strategies must remain attentive to the dynamic interplay of adaptability and efficiency in order to balance product-centricity versus the focus on consumer's demand for choices with respect to variables such as cost, quality, service and cycle-time. A key strategic domain is the creation of a robust supply network plan (suppliers, production facilities, distribution centers, warehouses). Another domain involves logistics (sources, sinks, centralized vs decentralized, direct shipment, cross-docking, pull vs push, transport). The quantitative decision domain involves determination of quantity and location of inventory including raw materials, work-in-process (WIP) and finished goods. SCM information systems generally revolve around these decision domains. Efficiencies and profitability may depend on the extent to which systems and processes of the supply chain are enabled to share real-time information about demand, forecasts, inventory and transportation (all of which have a bearing on the above decision space).

Of interest in this space is the "Bullwhip effect" [10] reflecting demand volatility or amplification in supply chain domains. For example, small changes in consumer demand at retailers may cause large variations in the inventory at distributors or manufacturers. Thus, near-real time information about small changes along the chain must be accounted in supply chain models. Technologies, such as radio frequency identification (RFID), may be useful for acquisition of inventory data at the item level if there is sufficient business value for such granularity of data [9].

3.2 Regulatory Role

Supply chain facilitation offers opportunities for businesses to optimize profitability. For government agencies such as customs and border security administrations, the concerns are different. Availability of real-time accurate data, in advance, is one key element in their effort to "target" high risk shipments. Operational profiling (source of goods, personnel, routing) is an emerging paradigm for decision support systems that deal with risk management.

This problem space in the regulatory domain is in sharp contrast to the facilitatory view of information systems in business supply chains. Globalisation has forced supply chains to span multiple geographies and introduced a significant regulatory step for all physical objects and goods that cross geographic boundaries. This intersection of facilitation versus regulation falls squarely in the operational domain of customs administrations in each country [13]. Supply chains involve actual shipment of good across boundaries. The efficiency with which these goods are handled (shipment, receipt, distribution) are critical because delays will impact inventory (risk of out-of-stocks), quality of services (QoS), cycle-time, cash cycle, capital costs and transportation. It appears, therefore, the tax collectors of customs must evolve from the revenue domain to become an integrated part of the global supply chain if countries want to remain competitive in a global economy without business borders, at least in theory. In addition, the post-9/11 world necessitates that customs emerge as a border security force equipped to manage risk through targeted intervention of shipments that must be certified by customs, both in-bound and out-bound [14].

3.3 Healthcare

Systems interoperability is an obvious and natural phenomenon in the living world. For example, a human viewing a tiger in a field or woods may start breaking out in a sweat even if the external environmental temperature is cool or even cold. Healthcare and medicine must take into account the physiology of a system when addressing the use of biomedical research data or the needs of patients. However, data from biomedical research and point of care (POC) systems are largely disconnected or connected only through active human intervention (doctors, nurses, medical technologists). One such high volume data producing recent advance in biotechnology enables screening of gene expression using a technique known as microarrays [23].

The volume of data that must be analysed and correlated from a microarray screen is simply staggering. In physiology, genes or proteins rarely act in isolation to elicit a response (for example, sweating in response to viewing a tiger) or precipitating a disease state (for example, diabetes or cancer). The human genome project now offers the fruits of research that may be used to identify the location of a gene on the human genome map (and determine its neighbours and thus the potential for interaction). A ***combination*** of information reveals which genes are active (data from microarray) and where on the chromosome that gene is located (data from genome map). This reflects a monumental gain in valuable data but in reality the systems interoperability problems impede the "understanding" or translation of the data to actually improve the treatment or point of care service to the patient. Connected systems and collective understanding of the data from various new tools could better guide healthcare delivery and manage risk of future complications. In this example, genetic activity data from microarray and the "gene positioning system" (GPS) based on the human genome map may remain segregated or even isolated unless specifically trained individuals capable of relating one to the other "understands" the "meaning" of the data and explores its value or applicability to improve patient care.

To mine this data and make logical connections, we need agents in the semantic web with advanced abilities to look beyond the obvious, that is, not only individual genes in action but interaction between gene products or a network of genes involved in a *genetic circuit*. The type of interaction protocols mentioned above and the theory of distributed cognition, taken together, may be one necessary tool that deserves exploration from the point of view of molecular medicine and bioinformatics with respect to a number of disease states where the etiology often includes perturbation of gene-protein networks (for example, the tumour suppressor gene p54 or the oncogene, RB2).

Since the semantic web is a virtual space, it makes informational structures more relevant. In one view, the semantic web is a collection of knowledge, which, by definition, has machine-accessible meaning (in contrast to the 'dumb' collection of information on the syntactic web). This notion of the semantic web is currently being challenged in favour of the view of the semantic web as an action-enabling space [24]. Agents may play a central role in delivering the value of the semantic web to the user. For sake of clarity, it should be pointed out that "agents" are better suited to specific tasks while "agencies" or hierarchical network of agents may be endowed with intelligence. In other words, the theory of distributed cognition may apply to agencies but not to agents. Theory of distributed cognition departs from the mainstream of cognitive sciences in that it emphasizes the participation of external elements (manuals, databases) in the agent's thinking process [25]. Interaction of agents in the physical space may be facilitated by using mobile agents linked to devices (sensors, electrocardiograph, intravenous pump, software defined radio, PDA). Such interactions call for introducing ontologies for describing interaction protocols and agents that can recognize ontologies. For example, inform(door_open) and request(door_open) are different messages even though content(door_open) is the same. The former is about the state-of-affairs but the latter transmits an intention. These can be grouped as interaction protocols (conversation patterns). If agents are aware of this ontology, agents can learn new ways of communicating and by extrapolation, adapt to the needs of the system.

Distributing cognition in the semantic web through design of agents that can understand ontologies describing new interaction or communication protocols may find immediate use in healthcare and in the emerging confluence of nano-bio medicine. For example, a paramedic responding to an accident, must communicate to the nearest hospital trauma unit the status (answers) of the crash victim, not numbers. For example, report(heart_rate) should say 'normal' rather than 80/120 mm Hg. If we consider involvement of agencies with distributed cognition, the 80/120 mm Hg blood pressure may be linked to a context, for example, age. If a child is involved in the accident, surely report(heart_rate) should not transmit status 'normal' to the trauma center. The syntactic web and current software may be hardcoded to "report" "normal" for "80/120" values of blood pressure. This "normal status" report generated by the system unbeknownst to the paramedic attending the child at the scene of the accident could delay the scheduling and dispatch of an advanced pediatric trauma unit to the location of the accident. Delay in critical services may cause irreversible damage to or death of the child.

Missed connections, as explained in the previous paragraph, costs lives (mortality) or reduce the quality of life (morbidity). The decreasing workforce of trained personnel is evident even in the most affluent nations and almost non-existent in the rest of the world. Machine readable data and the use of "intelligent" systems to extract value from biomedical data is an important area that may be served by semantic interoperability and the tools discussed in this paper. The "success" of this confluence in terms of application, however, is dependent on the diffusion of the semantic web within hardware and software infrastructures as well as ontological frameworks. If these systems are in place, healthcare delivery may rapidly improve its quality of service. But, in order to extract the value from data, it will be necessary for the healthcare industry to invest in these tools and infrastructure to reap the benefits from ontological frameworks and the potential of the semantic web to connect the dots.

The potential of the semantic web to "connect the dots" is in fact a very tall order and one that is squarely dependent, at least in healthcare and biomedical applications, on the development and diffusion of mapping and translation functions *between* ontological frameworks. Two healthcare examples are cited below to emphasize the importance of this ontological mapping functionality.

Research and advances in nanoscale diagnostics can potentially save millions of women from the morbidity caused by breast cancer and linked mortality. Two genes, BRCA1 and BRCA2, identified nearly a decade ago, appear to be closely linked with familial early-onset breast cancer that constitutes about 5% of all cases. In addition to BRCA1 and BRCA2, at least eight other genes have been identified as contributors to breast cancer either directly or indirectly [AKT2, BWSCR1A, CDH1, ESR1, FKHR, PAX7, PIK3CA, ST8]. BRCA1 and BRCA2 are also linked to ovarian cancer and rhabdomyosarcoma (cell types are significantly different from mammary epithelium). Among the familial early onset group of women with breast cancer, 40% of them have mutations in BRCA1 and BRCA2. About 120 different type of mutations have been identified in these two genes and twice as many are actually thought to exist in the DNA (10,000 base pairs of A, T, G, C make up these genes). Any attempt at patient care management for this subset of afflicted women will be incomplete [or even wrong] without an understanding of the nature of changes [which one of the hundreds of possible mutations] that caused the carcinoma in the first place. The situation is analogous to repairing a leak in a water pipe that is a mile long. If you cannot identify the site of the leak, what are the chances of repairing the pipe? If you do not know 'where' or 'how' to look, what are the chances of identifying the site of the leak? The vast amount of research and patient data that is helping to build the knowledge base that may help identify the "site of the leak" is still not connected to healthcare providers or specialists. The "vocabulary" necessary for communication between data silos in research and practice is woefully inadequate and contextual interoperability of medical information systems are nearly non-existent. Hence, the value of creating ontological frameworks to benefit from biomedical research and improve healthcare.

Nearly a decade ago (1998) microarray was used to stratify patient populations with acute myeloid leukemia (AML). The drug Cytarabine produced remission in 78% of patients (even after a 5-year period) in patients selected by genotypes that were expected to work in synergy with the drug. In parallel studies with other AML patients, who were genotypically grouped and predicted to be less responsive, Cytarabine showed remission rates of 21%. In the pre-genomic era when genotypic stratification was not available, this drug could have produced a widely variable remission rate depending on the mix of patients and may not have received FDA approval. The convergence of two different tools enabled stratification of the patient population to reveal a potent drug, at least for a subset of patients afflicted with AML. Such stratification is not yet the norm due to high cost but nano diagnostics may soon make it routine. Genotypic profiling enables treatment of patients with genotypically matched drugs to maximize efficacy. The confluence of data and information necessary for the stratification strategy and genotypic matching of drugs to improve efficacy combines data, information and knowledge from diverse disciplines (organic chemistry, toxicology and pharmacogenomics, to name a few).

To make the desired "connections" between these fields without or with minimal human intervention, search systems must refer to ontologies and translational mapping *between* ontologies to enable semantic web users in medicine to make the appropriate links or build the hypothetical model from vast resources of decentralized information. Research data on AML, clinical data on AML, pharmaceutical data on chemotherapy, toxicologic profiles of drugs, epidemiological data, patient demographics, research advances in specific areas, FDA policy, human trial guidelines, safety compliance, resource availability and other important data are not and will not be organized in any one search space, in any one nation, in any one format or in any one specific ontological schema. Hence, the vital need for ontological mapping and the semantic web interface for healthcare.

4.0 MULTI-AGENT SYSTEMS

4.1 Evolution of Agents

Agents as automated software entities have been under development and use since 1960's (see Appendix 5). In the early days 'daemons' written in primitive UNIX could be fashioned as perpetually operating background processes. Agent technologies are currently used in remote management of IT infrastructure. Autonomous software agents monitor the health of infrastructure components (SNMP or Simple Network Management Protocol) or are evoked in simple processes such as RPC (Remote Procedure Call). This is an example of master-agent communication where the master delegates responsibility to autonomous agents for real-time sensing and reporting (to master). It was demonstrated that given proper task decomposition, agents could collaborate to provide coordinated output that would make sense to human users through the master controller. Agent technologies are evolving to include agent-to-agent (peer-to-peer) relationships communicating over mesh networks without master controller. These agents may have spatial mobility on the network (mobile agents migrating computing power closer to the data) and may have computing capabilities with rudimentary intelligence (using fuzzy algorithms, neural nets and methods that emulate machine learning). Industrial grade agents could be reliable, robust and fault-tolerant. Role-differentiated collection of agents may collaborate to act as intelligent multi-agent system. Such a group or swarm of agents (or agencies) will collectively exhibit behavior different from individual agents (swarm intelligence).

4.2 Structure of Agents

Traditionally, agent software may have five conceptual layers [6]. The outer envelope, generally, is the Communication Layer (for communicating with other agents). The Interface Layer is for sensing and effecting. In between the communication and interface layer is the computational payload, comprising the Definition Layer (that makes the connection with the Interface Layer). The Organizing Layer is usually responsible for core processing of information in conjunction with the Co-ordination layer which handles inter-agent inputs. Sycara [21] lists the characteristics of multi-agent systems as collectively autonomous, but task-specific and inter-dependent, systems that sense, seek and process information in their respective problem domains. In each problem domain there could be many deployment strategies [19] for multi-agent systems. Researches have explored some of the uses of agent technology in SCM [22] but there is much room for improvement based on bio-inspired mechanisms (Appendix 5).

4.3 Use Case: Track and Trace Agent

Profitability and security are both equally powerful drivers for access to real-time data and information about inbound and outbound consignments all the way from the original consignor to the ultimate consignee. In order to make this a reality, it is increasingly necessary to take advantage of automatic identification technologies and the devices (sensors, RFID tags, GPS) that may be placed on goods, carriers and fixed locations, such as, entry, transit and exit points in factories, warehouses, ports and public places. Embedded agents in the software layer may contain business logic (facilitation of supply chains for profitability) and/or risk profiles (regulation of supply chains for security) to continually monitor and/or analyse.

A typical use case involves sending RFID[1] or sensor data (translated into information) to a data receptacle or UDB or business application or secured access key such as the Unique Consignment Reference (UCR)[2]. Location awareness of objects is an over-arching theme in this physical world model where bits are connected to atoms. Progressive adoption of automatic identification tools (for example, RFID) makes location awareness possible but the value of such data may not be realized without advances in context aware applications. Use of such information may help prevent accidents if cross-reactive chemicals were being transported in containers. Information agents and message filtering agents can route context-aware object status either to automated decision systems or induce human intervention.

---

[1] RFID is a class of automatic identification technology that uses devices (passive radio frequency ID transponders or RFID tags) that may allow an item to have an unique serial number such as EPC or electronic product code for tracking purposes.

[2] UCR is a reference number for Customs use and Customs may demand it at any point during a Customs procedure or transport (www.wcoomd.org/ie/en/press/ucr_new_e.pdf).

5.0 SEMANTIC WEB

5.1 Core Principles and Ontology

The average user may never "see" the Semantic Web but the buzz about the semantic web is as intense as the internet itself. At the heart of the excitement is the realization of the fact that semantic metadata will enable users to do things with meaning. The massive amounts of data already available and the inevitable surge in data expected with spread of ubiquitous computing may be useless unless meaningful correlations and connections helps to make better decisions and drive innovations including the profitable ones or those that can save lives. But just because the semantic web may "hidden" from view does not mean that one can bypass the evolution of the semantic web. It is intended for computers to improve searches, viewing data, interacting with services and sharing information. It can offer process transparency across language and geographic boundaries to connect partners even if individual partners define or perform certain functions quite differently from others or use "standards" that are not standard.

The vision of the semantic web, first formally outlined by Tim Berners-Lee in 1995 [2], has matured. Progress has taken place in research communities around the world to demonstrate that semantic web may solve a variety of today's problems. Semantics is a collection of Resource Description Framework (RDF) data (or any other semantic language) which describes the meaning of data through links to ontologies, which act as decentralized vocabularies. Ontology is a term borrowed from natural philosophy. Therefore, a definition of ontology may state that ontology is a theory about the nature of existence (of what types of things exist). Artificial intelligence and semantic web researchers have co-opted this term to indicate a document or file that formally defines the relations among terms. Computers in future, empowered with this metadata, may be far more "meaningful" and "contextual" in their understanding of the data without human intervention, provided that the data is in machine readable format [26]. Human language thrives when using the same term to mean somewhat different things, but automation does not. Michael Dertouzos [27] and James Hendler [28] explain this central issue with a brilliant and simple example [30].

The real power of the semantic web will be realized when agents collect web content from diverse sources (for example, stock quotes from Bloomberg or microarray data), process the information (for example, in relation to your business or diagnostic test) and exchange the results with other programs or data (for example, demographic data or metabolomics index). The effectiveness of such agent based activities will increase exponentially as machine-readable web content, automated information services and real time-data become available. The semantic web promotes the synergy between agents that were not expressly designed to work together but can now transfer data among themselves if data is marked-up with semantic tags.

With ontology pages on the web, solutions to terminology (and other) problems begin to emerge. The meaning of terms or XML codes used on a web page can be defined by pointers from the page to ontology. Of course, the problem arises if you point to an ontology that defines addresses as containing a zip code and one that uses postal code or pin code. This kind of confusion can be resolved if ontologies (or other web services [32]) provide equivalence relations: ontologies may contain the information that a zip code is equivalent to a postal or pin code. In other words, transformational mapping functions between ontologies will be necessary as ontological frameworks begin to evolve, in the same 'organic' manner that characterized the explosive growth of websites from 0 to 25 million in the two decades since 1980.

There may not be any one 'standard' ontological format even for very closely related topics because the same format can be framed differently in a different language. For the semantic web to be useful it will be necessary to have layer(s) of mapping functions (analogous to adaptors and transformers that are necessary to use electrical appliances across geographic boundaries). Advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules. This mark-up makes it easier to develop programs that can tackle complicated questions whose answers do not reside on a single web page yet vital for certain users, for example, in healthcare. Access to these pages or data will demand security and authentication. Therefore, for agents to accomplish tasks that require data or information from multiple sources, an important facet of agent function will be exchange of "proofs" which may be written in the semantic web's unifying language using rules and information such as those specified by ontologies (see "semantic web layers" in the illustration below). Some programs already claim to exchange proofs in this way, using the preliminary versions of the unifying language but they are far from plumbing the depths of the semantic web's true unifying potential.

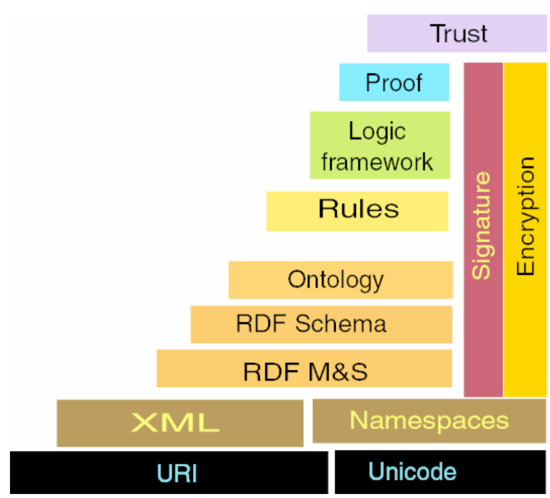


Illustration: Semantic Web Layers from Tim Berners-Lee [2]

Present day “discovery” engines and automated web services claim to discover and connect to various services. It is doubtful if at present the agents have found a way to locate a service that will perform a specific function. This process, that is, true service discovery, may happen only when there is a common language to describe a service in a way that enables other agents to "understand" both the function offered and how to take advantage of it. Services and agents can advertise or publish their function by depositing such descriptions in directories (Yellow Pages). Some low-level service-discovery schemes, such as Microsoft's Universal Plug & Play, focus on connecting different types of devices (information box in Windows XP: Found New Hardware). These initiatives, however, attack the problem at a structural or syntactic level and rely heavily on standardization of a predetermined set of functionality descriptions. Standardization can only go so far because we cannot anticipate future needs and disruptive changes.

One antidote to standardization and n:m relationships is semantics. The semantic web is flexible. Consumer and producer agents can reach a shared understanding by exchanging ontologies, which provide the vocabulary needed for discussion. Agents can learn new reasoning capabilities when they discover new ontologies. Semantics makes it easier to take advantage of a service that may only partially match a request. A typical process involves the creation of a "value chain" in which sub-assemblies of information are passed from one agent to another, each one "adding value" to construct the final product requested by user. The immense value of such information in healthcare cannot be overemphasized. To create complicated value chains, automatically, on demand, agents may increasingly exploit artificial intelligence techniques including tools from swarm intelligence, such as, ant-based algorithms. Semantic web will provide the foundation and framework to make such technologies more feasible. Its use will become ubiquitous and pervasive as context-dependent communication evolves successfully and deftly to address the many idiosyncrasies of the human language-dependent ontological frameworks through intelligent mapping functions.

5.2 Decision Support

Thus, the semantic web may be envisioned as an evolutionary improvement of the syntactic web [2] that is capable of “understanding” the meaning and by extension, eventually, the context, of information. In practice, the tools of the semantic evolution will enable software applications to “read” information presented on the web as easily as humans read web pages, process information contained therein and arrive at useful outputs in order to display meaningful information for other software applications or humans to use or act upon. Thus, a web of data, information and process silos can be linked depending on the contextual relationship between the data or process. Ontologies and the semantic web enable this contextual relationship to be recognized but the semantic web, *per se*, does not link the data silos or process streams. One way for such “connectivity” to occur is through the use of agents that can now search and find the necessary data or process because of a "common" language (ontology) that delivers the meaning (semantics) of the data or process, which in turn can "bind" together to generate useful information that may trigger an improved decision making process (not possible in the syntactic web framework where the meaning of the data or process cannot be "read" by other systems or agents).

For example, data from a shipping bill from Sony manufacturing plant in Penang (Malaysia) and near real-time location information from GE VeriWise automatic identification system [15] used by the logistics provider, Schneider Corporation, for track and trace purposes, may offer an estimated arrival time for the goods from Sony to its US distributor Merisol after US customs clearance in Long Beach, California. This information may generate an alert or advice to the retail store that the Sony PSP3 devices may not be on the shelves of the Circuit City store in Watertown, Massachusetts for the 4th of July sales event. The value of such information in this supply chain scenario is immense yet the current tools and the syntactic web makes it difficult, if not impossible, to extract and connect the relevant pieces of event-driven data to generate decisionable information. Agents in the semantic web charged with specific "discovery" tasks offers the potential to collect, analyse and provide answers (not merely numbers) based on access to a common process and data dictionary linked to an ubiquitous data bus (UDB) uploaded with real-time data from supply chain partners. The key is the machine readable data and the context of process information (for example, the event "shipping bill" from a Sony warehouse should be contextually relevant to trigger a "discovery" event (search) for the "receipt of goods" linked data that may be in the domain of the logistics provider, Schneider Corporation or still within the customs area awaiting clearance).

The ability of the semantic web to be increasingly functional and productive for use by non-experts depends on the underlying growth of ontologies and ontological frameworks. It is the latter that makes "machine readability" of data and "meaning" a part of the function that is expected to be delivered through the semantic web. The rate limiting function in the diffusion of the semantic web is the growth of the ontologies or at the next deeper level, the mapping between ontological frameworks. For example, ontologies contributed by the Japanese and German auto manufacturers may differ but through mapping transformations the differences may be made to "disappear" for a potential consumer who is comparing automobiles irrespective of their ethnicity. The ontological frameworks are also key for agents to "understand" process relevance and for ability of "discovery" or search services (true "web" services of the future) to link context of data, process or services within or between entities. For example, the business process "purchase order" in one company is business-relevant to "invoice" in another company based on the assumption that if Apple wants to buy Intel microprocessors then Apple will issue a purchase order and Intel will issue an invoice requesting payment from Apple.

As is inevitable in global trade, process names and descriptions change, often and without notice. For example, Lenovo, the Chinese PC manufacturer may use process descriptions not common to its US business partner, IBM. Syntactic discrepancies such as these are not expected to impede the semantic process relevance because the link between terms or descriptions (purchase order and invoice, for example) are not hard-coded or 1:1 programmed in an instruction set. The latter is the case, currently, in the syntactic web and is the method used by "directories" peddled by consortiums such as ebXML Registry, RosettaNET (PIPs) and UNEDIFACT, to name a few. These directories are very useful when only a few partners are involved in a generally stable industry but globalisation has shredded such norms. Validating and maintaining these directories is inefficient and will generate problems as process descriptions multiply in the multiple languages involved in business operations that characterize globalized supply chains. In the semantic context, no matter how the descriptions vary or evolve with globalisation, the meaning that is relevant to the process remains understandable (hence, machine readable) based on contributed ontological frameworks or mapping between ontological frameworks. The semantic web offers interoperability (and value) that is hitherto unimaginable in a syntactic world. The convergence of the semantic web with agents capable of accessing real-time data and right-time analytics is a step toward an "intelligent" interoperable decision system.

5.3 Web Tools

Traditional web tools (HTML) lack in the use of representation standards and do not provide for rich syntax and semantics. To overcome the lack of standard ways to exchange business information, consortium of businesses or organizations agree to pre-determined set of transactions (senders and receivers agree to common set of syntactic descriptions for processes) and thus sprout a common "standard" or directory. Such tedium was pursued under the auspices of the United Nations (UN) for three decades to reach one transactional standard: the UN EDIFACT.

UN EDIFACT has proved to be procedurally complex and difficult to implement despite the impressive body of work that includes attempts to create ontological frameworks. In this effort, XML was used for defining the structure and semantics of data through development of standard data structures and embedding the semantics in its feature-attribute mode. "Standard" ontologies were developed for different business areas under the guidance of UNCEFACT TBG (domain oriented working group within the UN ECE) to drive UNeDOCS, a form of web service for business transactions. The ontologies developed under these bodies, although comprehensive, are limited in their scope because they were conceived with a transactional orientation rather than true ontology frameworks that focus on meaning and context irrespective of the usage or orientation of the user. These ontologies, therefore, are but only a subset that covers transactional elements, commencing from order initiation to fulfillment, in a supply chain.

Neither the transactional limitations nor the inadequate semantic structure, however, prevented industry consortia (for example, OASIS), to build a suite of technical artifacts under the ebXML umbrella. This suite of "standards" are now a part of ISO 15000 series.

The example cited above is indicative of the chasm between experts with depth of understanding and the rush for industry to extract profitability from the fist-mover advantage even if the tools are inadequate. Use of XML as a tool in "defining the structure" may be one problem faced by groups and organizations rushing to "standardize" some predefined elements with certain preconceived business orientation. In the example below, the data pertaining to customer inquiries received at two branches of the same company are described quite differently in XML although the meaning and semantics are identical [19]. As is, the data (below) cannot be automatically merged for a quick report to the service manager to highlight the fact that 4759 customer calls were received on 4th of July 2006.

```
<CompanyData>
  <CompanyName>
       Fidelity
      </CompanyName>
  <CallData>
    <RecordDate>
                  Tue July 4, 2006
    </RecordDate>
    <CallsPerDay>
            3141
    </CallsPerDay>
  </CallData>
```

```
<CorporateRecords>
  <Company>
       Fidelity
  </Company>
  <Records Data= "Tue July 4, 2006" >
       <Calls Units="PerDay">
           1618
  </Calls>
</CorporateRecords>
```

5.4 ebXML Registry

The inadequacies of XML alluded to in section 5.3 makes it quite difficult to justify or provide credibility to the prediction that ebXML Registry will be the semantic web server of the future [17]. In the short term and for software peddlers seeking profitability, perhaps, an 'ailing' semantic web in the transactional domain of supply chain has been made possible by the ebXML Registry. It does provide standardized vocabulary, albeit limited, for generic supply chain purposes and may enable agent mediated automatic discovery, to an extent. The functionality of the "standard vocabulary" is dependent on the ontologies that are *supposed* to be shared by all participating entities.

The assumption that supply chain participants may not change or that new participants will adhere to the ontological frameworks is unfounded, based on the innumerable small and medium businesses (SME) from all parts of the world who could be a partner in a global network. These SME's may have processes or systems that must be overhauled or modified if it must "fit" the ontological frameworks of the ebXML Registry and its community implementations, such as UBL and RosettaNET. Despite its restrictive functionality on a truly global platform, the ebXML framework[3] deserves praise for ushering the use of ontology building blocks in use cases, such as, supply chain management.

Beyond the ebXML effort that is solely focused on the transactional terrain, others have attempted to use the principles of the semantic infrastructure to simulate models of the dynamic, information intensive and spatially distributed supply chain environments [11]. One approach encapsulates standard models of the supply chain (SCOR or Supply Chain Operations Reference Model), builds an ontology core to capture and distribute knowledge artifacts within the model (using ontology creation tools, for example, Protégé, developed by Stanford University) and provides decision support tools. In another approach, researchers have used the software 'ZEUS' developed by British Telecom to build software agents in a multi-agent-based supply chain modeling system [18]. Languages for modeling and creating agents (COOrdination Language or COOL) allow users to build agents capable of inter-agent interaction.[4]

6.0 USE CASE: PORTS

Supply chains involving goods transported via land, air and sea often suffer from uncertainty in cycle time when port authorities or regulatory intermediaries must issue clearance before the goods can flow back in the business operation. Therefore, ports are major institutions and a key node in the order fulfillment cycle.

The word "port" is an amphibole of different meanings – an aggregation of commercial enterprises seeking profits and operating a natural monopoly in providing a public utility service.[5] On the ocean side, the port serves the vessels represented by shipping agent and non-vessel operating common carrier (NVOCC). In addition to cargo

---

[3] ebXML was sponsored by UN CEFACT and industry consortia OASIS, representing software vendors (www.ebxml.org).

[4] COOL http://eil.utoronto.ca/profiles/rune/node7.html

[5] Towards Port Best Practice (March, 2000)

carrying vessels, there are port-based service providers such as tugs, pilot vessels and lighters. On the harbor side, there are terminal operators for facilities at wharves and berths for handling cargo, storage and warehouse facilities that serve inside and outside the customs bonded area. Port Authority (handling traffic and commercial operations), Customs and several government departments (Public Health, Transport and Security) have a presence at ports. The primary consumers of port services are the businesses and organizations involved with storage and movement of goods. These include stevedores, road and rail transport forwarders, warehouse operators, container terminal operators, storage providers, container repair operations and the various branches of Customs.

These organizations are involved in a complex web of interdependent relationships and conduct a great deal of intra-community business. The underlying element of this business is the sharing of information concerning cargo and logistics. Information related to payments, transactions and consignments are exchanged between these entities using a wide variety of different business systems and processes, databases and methods of exchange. The flow of data and information is expected, theoretically, to stay ahead of the flow of goods but in practice this is the key problem associated with port management [12]. Typically, cargo movement is a 2-part operation: (i) movement of cargo between the ship and the gate of the terminal, depot or wharf, and (ii) movement of cargo between the customer (shipper/consignee) and the gate of the terminal, depot or wharf. These segments may be sub-divided into steps or processes that include harbor entry (reporting the ship, arrival of ship, berthing), loading/unloading cargo, harbor exit (clearing the ship, departure), cargo clearance involving importer, customs, NVOCC (containers), cargo consignment involving road transport and inland haulage.

Integration of operational data flow between the different entities in a port is key to efficiency and security. Partners must have secured access to relevant data and systems to process transactions. The core services are equipment availability, delivery requirements, demurrage information, bookings, bills of lading, vessel, barge and rail manifests, load and discharge lists, vessel schedules, export receipts, work schedules, empty container returns, transfer between container freight stations and inland container depots, regulatory holds and permits and payment for port services (permits, release, customs clearance). “Delivery” is the successful outcome of these processes that must operate in synchrony and largely driven by documentation and information.

6.1 Port Systems

Systems architecture revolves around a data replication service where members of the port community may submit data in an asynchronous manner using EDI. The “status” is viewable through a web interface to partners.

Often, the available data is inaccurate, inadequate, unavailable or unsuitable for operational decisions. These systems often impose business rules (as in ebXML) dictated by the cycles of EDI message exchange based on an "one size fits all approach" which is inflexible and cannot accommodate global variances of normal business practice. Port systems are an archaic monopoly, with all powers associated with the cost, quality and composition of the service offering vested with the host which rarely offers personalized services to suit specific business needs beyond the rigid rules. It is here that globalisation and the global supply chain meets with the grave danger of being asphyxiated by the dead weight of old technology.

On the other end of the spectrum, between the feasible and fantastic, in the context of "community systems" in port operations, is the goal of real-time process visibility in a paperless environment. The 'paperless port' is a cliché. But somewhere in between, a couple successful examples are worth citing. The Felixstowe Community System[6] is a port community system that covers over 80% of container cargo moving in and out of UK. NACCS in Japan is an example of a port community, where major participants have dedicated relationships with one system (operating for customs). Port systems must view the "writing on the wall" and prepare for stringent risk management that security concerns demand. Systems interoperability and real-time access to accurate data and status will be pivotal. Soon, port systems may be required to respond to the scenario outlined in the illustration below and strategies such as non-obvious relationship analysis (NORA) may become an integral part of port security analytics.

6.2 Non-Obvious Relationship Analysis (NORA)

Neutering threats before terror can strike is one primary objective of government security czars. Ports play a dominant role in this respect. In addition to the "physical" aspects of port security, a significant amount of related information must be available to ports in order to execute the physical measures based on intelligence. Because "physical" elements in ports often involve goods, the question of security is inextricably linked to goods or delays in clearance that may negatively impact the supply chain and profits. The ability to detect threats, therefore, is of paramount importance and hence the demand for increasingly improved non-obvious relationship analysis (NORA).

[6] Maritime Cargo Processing Plc was set up to run on a continuing basis, the integrated port information system, known as the Felixstowe Cargo Processing System (FCPS).

Since we are no longer dealing with nation states as primary sources of threats, the World War II treatment of "intelligence" and decisions based on such information are practices in absurdity. In the 21st century we are faced with "global public bads" manifested by small groups or individuals that use the internet to communicate or transform commonplace objects or humans to serve as destructive tools. Traditional approaches to uncover such 'plots' are practically impotent. Hence, the US National Security Agency is intensely exploring use of non-obvious relationship analysis (NORA). To be even minimally functional, NORA demands systems interoperability.

For example, a retail agri-business store in Tulsa, Oklahoma, sells 1000 kg of nitro-phosphate to an individual who pays $2000 in cash. Days prior to this transaction, a specific account number releases through an automatic teller machine (ATM) $700 per day for 4 consecutive days. One day prior to the retail transaction, a small moving van is rented from U-haul and a local warehouse store (Costco) sells 10 large duffle bags to one customer. Two days after the transaction, an individual boards a flight to Newark, New Jersey with two large duffle bags as checked-in baggage and returns the same day to Norman, Oklahoma. After a few days, the individual from Oklahoma boards a flight from Denver, Colorado to Philadelphia, Pennsylvania. A few days later, a BMI airline check-in clerk reports a scuffle with a passenger who insisted to check-in an oversized overweight duffle bag for a flight from Leeds (UK) to Charleroi (Belgium). On a hot June afternoon, an explosion in the Deutsche Post mail sorting center at the Frankfurt Main Airport kills 314 people. Days before the blast, bell-boys at the Sheraton Hotel off the Frankfurt Main Airport had noticed and reported an individual who was sitting on the walk-way that connects the airport terminal 1 building to the lobby of the Sheraton Hotel. The individual was reading a book and was there for several hours each day. The police checked him out to be an US citizen and the German authorities alerted the local US Consul. The US Embassy was instructed by the US Department of Justice that the rights of the individual cannot be violated by ordering a search because the 'suspected' individual did not break any rules by sitting and reading a book in a public place.

The intersection of policy, rights and suggestive clues when combined with investigation, offers a very complex scenario. Policy and rights must be evaluated by law enforcement prior to any action, exploratory or investigative. Similar intersections are common in healthcare and insurance. The policy framework enabled by the semantic web (policy-aware web) may offer clues to solutions. The ability of agents to connect diverse decentralized information in the framework of the semantic web may improve the efficiency of such searches. The links and patterns may offer relevant clues that may aid non-obvious relationship analysis of data pertaining to random non-threatening events (sale of fertilizer in Tulsa) which may be danger (explosion in Frankfurt) in disguise.

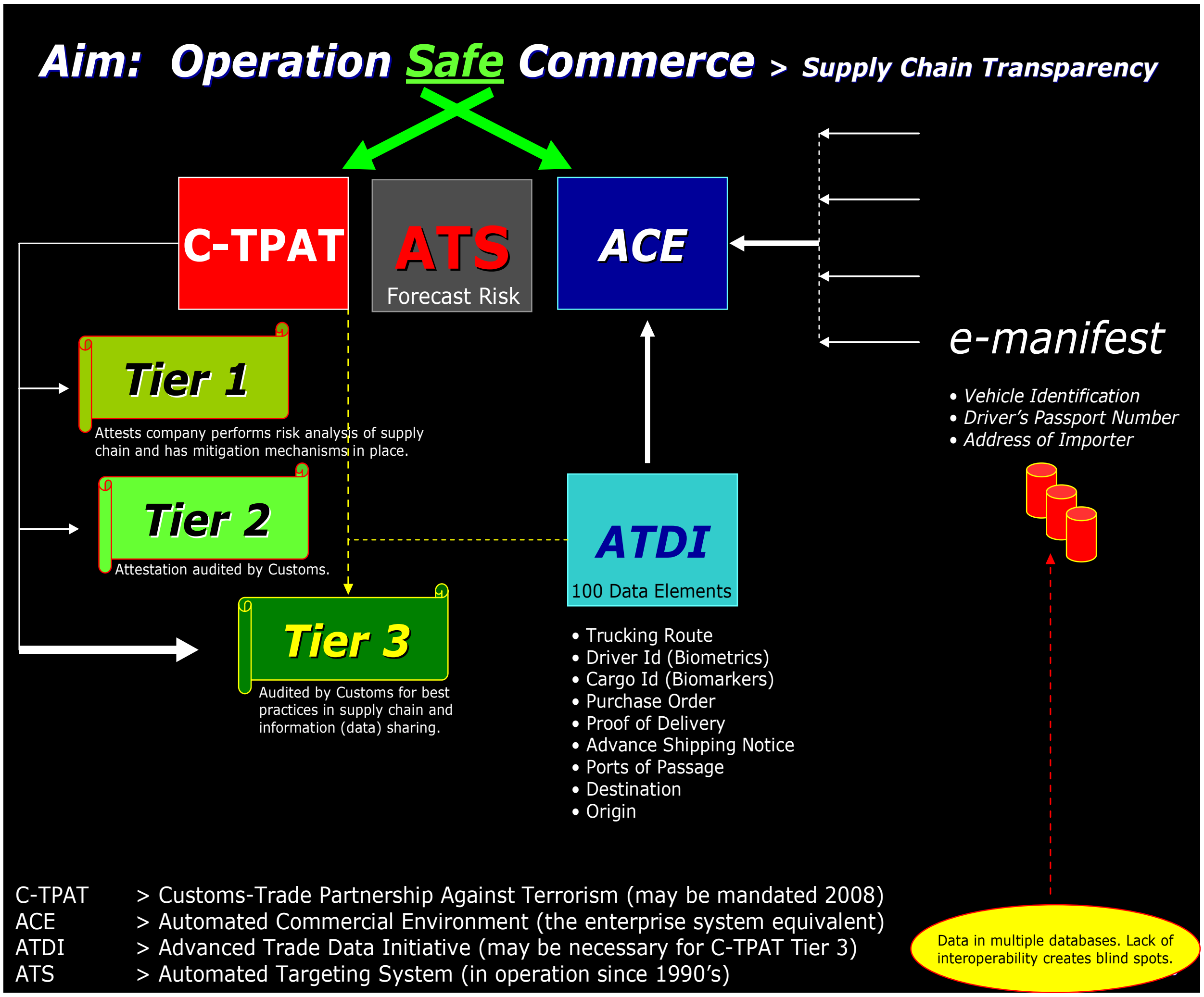


Illustration: Port Security Related Pilot Projects in the US

There exists a possibility of a mandate by the US in the form of Customs-Trade Partnership Against Terrorism (C-TPAT). Only Tier 3 certified companies may be allowed by US Customs to receive clearance without or with minimal inspection of their containers. To qualify for C-TPAT Tier 3 certification by US Customs, business must share data through the Advanced Trade Data Initiative (ATDI). Sharing sensitive data will add layers of data security. With data from ATDI, the customs "enterprise" system or Automated Commercial Environment (ACE) is expected to run analysis to spot anomalies, integrate biometric information (individuals, meat and agricultural products), perform non-obvious relationship analysis (NORA) and forecast risk profile associated with containers. Armed with this information, customs aims to selectively "target" cargo for inspections. A mere 60-day delay in customs clearance cost US businesses $58 billion in supply chain losses [13].

6.3 Agents in Port Community Systems

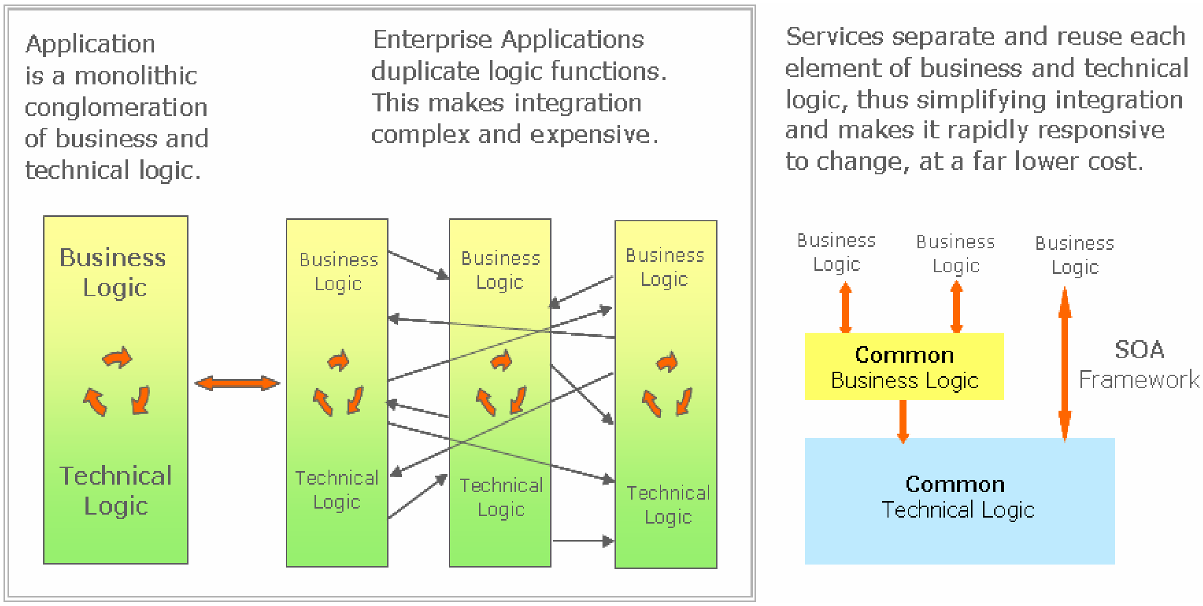
Ports are, therefore, a "cog in the wheel" of the global supply chain with the dual responsibility for trade facilitation and a prominent role in security regulation. Port community systems contain the necessary tools and infrastructure entrusted to execute these dual functions which are on opposite ends of the spectrum. The diversity of sources of data and information may make it impossible to thread together a multiplicity of systems and orchestrate systems level maintenance to continually synchronize real-time data, updates and changes, to be equally effective in facilitation and regulation.

Traditional systems demand harmonization and applications insist that data is derived from one central source which is expected to be accurate as well as authentic. This is antithetical to the multi-agent systems that thrive on data from multiple sources. Agent systems bring the program to the data and not vice versa (as is the case in traditional systems). Agent systems do not depend on data replication or synchronization techniques to maintain databases. On the other hand, port systems are designed to provide single point of access to data and control the transmission of requests for accessing data from the appropriate source. The assumption in this architectural strategy is the presumption that the data sources (that is, other systems in the port community system of information systems) are available online (as web-services) and can be accessed as XML responses to XML requests. Progressive port systems are increasingly exploring the internet as the key medium for use with web services and service oriented architecture (SOA) including the XML based ebXML framework and standards proposed by the W3C.

Key Standards for Service Oriented Architecture (SOA)

- Web Services Definition Language (WSDL)
- Simple Object Access Protocol (SOAP)
- Business Process Execution Language (BEPL)
- Universal Description Discovery and Integration (UDDI)
- Web Service Choreography Interface (WSCI)
- Blocks Extensible Exchange Protocol (BEEP)
- Web Services Addressing (WS-Addressing)
- Security Assertion Mark-up Language (SAML)
- Web Services Distributed Management (WSDM)
- Web Services Trust Language (WS-Trust)

The enthusiasm in the industry for Service Oriented Architecture (SOA) is justified because it may be a catalyst for enterprise-wide and inter-enterprise collaboration and integration. Illustrated below, it is clear to grasp why SOA may offer choices to deal with complexities and the future [20]. However, there are only a few implementations to date since the technology and its plethora of advantages are very poorly understood by practitioners, especially in public utility monopolies (such as, ports). In addition, public institutions are refractory to change and may continue to struggle to stay within their "comfort" zone with entrenched legacy systems rather than embrace new solutions.



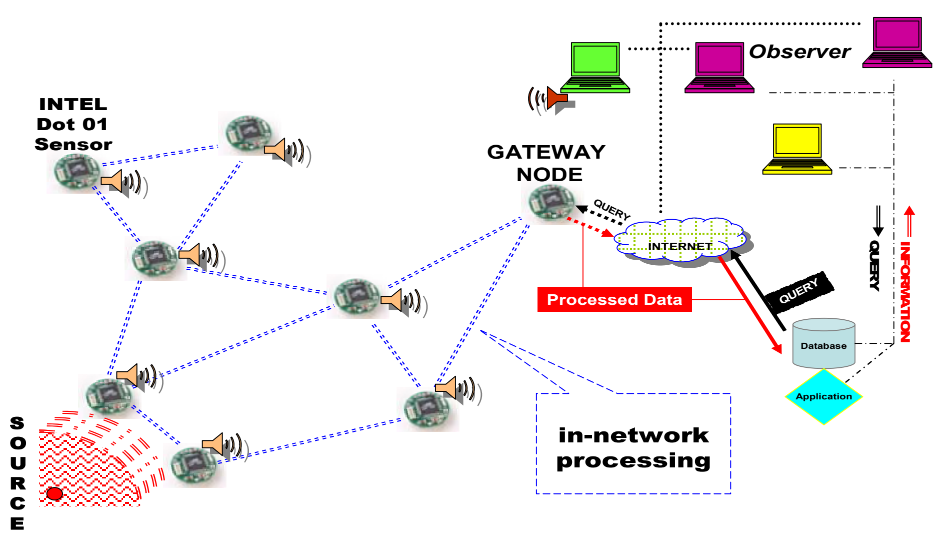
SOA is also attractive due to its synergy with agent deployment strategies. If high-level ontologies that deal with real-world concepts are systematically broken down to their core components, then regardless of the type of information system, it is possible to have an easy and seamless interchange of time-critical information. The granularity of the ontological framework will make agent specific tasks for data acquisition and/or discovery quite focused or targeted (high degree of process decomposition). Thus, the granularity supports hub and spoke oriented information systems used by ports as well as autonomous agent systems. Therefore, development of ontologies for each layer and the prospect of integrating port "community hub" systems using SOA and agent technology is as real as the "view" of this connectivity from reverse (agents or agencies considering such hubs as a part of their colony or empire). Needless to mention, ontological frameworks shall also feed the diffusion and function of the semantic web which will in turn boost the functionality of true web services, a medium and tool that is already preferred by users. Hence, port systems may offer a robust proving ground for systems interoperability that draws on the convergence of data sources such as automatic identification technologies, use of artificial intelligent based agent systems for connected discovery and the catalytic functionality of the semantic web to collectively make intelligent decisions.

7.0 CONVERGENCE of DATA, AGENTS and SEMANTICS

7.1 Sensors

The separation of business logic (process) from technical logic (device) outlined above in the discussion on SOA may be a key to harness value from ubiquitous computing. The nature of ubiquitous computing is still largely unknown but growth of wireless sensor networks may be an emerging example of pervasive ubiquitous computing. Application of sensors span the entire gamut that includes sensing of blood pressure and transmitting them to monitoring devices or to suggest trends of warehouse shelf occupancy or 'smell' hydrogen leaks. Sensors do not transmit identification data, such as electronic product code (EPC) or global trade identification number (GTIN), characteristic of auto-id technologies (RFID). Sensor data, therefore, cannot be used in the same manner as RFID. Sensors cannot be "plugged-in" directly as internet devices (InterDev) unless IPv6 (internet protocol version 6) is in use or the TCP/IP stack is subjected to architectural redesign with the much anticipated security layer (in progress).

Sensors are self-powered and may form wireless *ad hoc* networks that upload through specific nodes which may be then connected to data stores or the internet. Each sensor may have certain analytical abilities and due to in-network processing, some sensor networks transmit analyses of the data rather than the raw bits of data to provide "answers" instead of only "numbers" to the system. Sensor data may require different thinking in terms of "adaptive flow" or "streaming" databases. The data (analyses from sensor nodes) may stream through databases where the *query is stored*. For example, light emitting sensor network in a secure room sends positive light emission data on which the query (is anybody entering the room) need not act. Only when an obstruction causes a break in the network or occludes the light from a sensor(s), then, the query comes into effect and is answered.



Wireless sensor nets communicate via 802.15.4 but sensors may not directly connect to the Internet, yet.

Embedded sensors are likely to influence fields as diverse as healthcare and supply chain. Sensors attached to spindles in drilling machines may upload the status of the spindle in order that it is serviced or replaced within a reasonable time to avoid breakdown and downtime. Metrics, for example, meantime between failure (MTBF) and other parameters may be helpful to schedule preventive maintenance. Service supply chains (such as heating, cooling) may benefit from sensor-linked monitoring to determine when to send technicians to stem problems before they require emergency attention. The key is to integrate sensor data to improve performance. The flood of data from nanosensors may require agent integrated systems to extract intelligent information. Bio-nanosensors may evolve as an influential component of healthcare services and management.

7.2 Convergence of Bio-Nano-Info in Healthcare Decision Support

Information (info) from glucose nano-sensors [29], through semantic web services, may improve diagnosis and prognosis of type I diabetes (bio) that results from insulin-dependent disequilibrium of blood glucose. Diabetics use a host of methods to detect blood glucose but lack automated continuous monitoring tool. Catalysing convergence of "bio-nano-info" may spawn improvements in healthcare in real-time, delivery of services, resource management and the national economy. It is not inconceivable that confluence of tools aided by the semantic web may be quite disruptive. For example, glucose nano-sensors may reduce diabetic retinopathies in response to national healthcare policy to boost the quality of life and find cost-effective unassisted living for senior citizens diagnosed with diabetes. Convergence of ideas and tools, in the example above, is essential for fostering equity through global public goods. The later is a role for governments, global institutions and public-private-academic partnerships (Appendix 4).

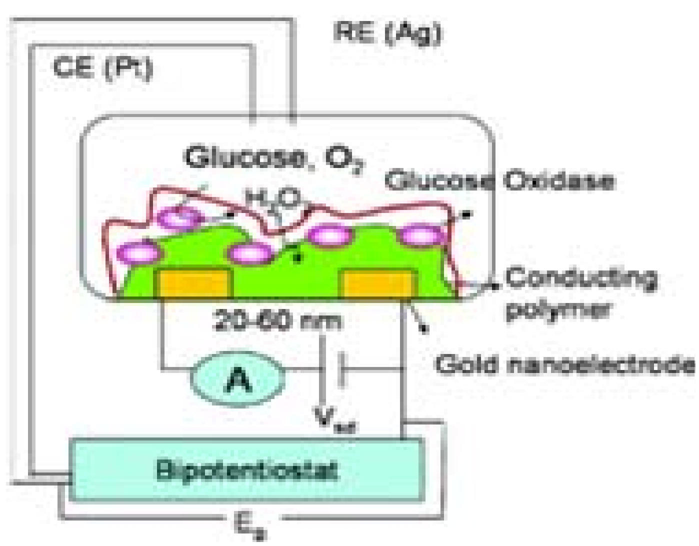


Illustration: Glucose Nano-Sensor: Disruptive Technology?

Let us explore a simple healthcare scenario where a local hospital in US zip code 90210 is the first point of contact for seniors aged 65 and above. About 200 seniors are identified to have some form of diabetes (type I and late onset). It is necessary for these individuals to have a monthly check-up and if the blood glucose is elevated the physician may prescribe a regimen of insulin. Family history reveals that a few seniors may have had parents or grandparents who had glaucoma, which could be diabetes-dependent. At present, the 200 seniors in 90210 must travel to the local hospital for monthly tests (outpatient clinic). Perhaps most will have normal blood glucose (120 mg/dl) or levels which may be controlled by modification of carbohydrate intake. However, a few seniors (assume 5) with a family history of glaucoma, require more careful monitoring yet only tri-annual visits are covered by their insurance. If a check-up reveals elevated blood glucose in this 'risk' group, then the clinic must schedule a physician appointment and take measures for insulin therapy. In practice, checks-ups for the 'risk' group are always delayed.

A generic process analysis may reveal costs associated with screening and monitoring of 200 individuals at the outpatient facility (includes cost of services for those who may be otherwise normal). The "supply chain" perspective will indicate inventory issues with insulin: it may have a short half-life (shelf life) and supplier-dependent (location of Genentech on US west coast) factors may make it necessary to stock enough insulin (inventory carrying cost) and risk expiry (waste) due to uncertainty of demand (how many of the 200 will require insulin). The quality of life analysis reveals that 195 seniors with diabetes but without known family history of glaucoma may be suitably cared for without the monitoring and services of the outpatient clinic personnel. Due to insurance and other cost factors, typically, the quality of service (QoS) and service level offered to all 200 patients are similar. The 5 at risk patients need more attention but due to the inefficiency, receives far less than desired service level. Lack of such services for these 5 diabetic seniors may reduce their quality of life, increase their risk of glaucoma with a concomitant increase in morbidity from poor vision. The latter situation, although preventable, creates a far greater burden on the social economy because these 5 seniors may no longer live unassisted and may require costly full-time assistance.

The alternative efficient scenario is one where patients with sub-cutaneous (implanted) glucose nano-sensors monitors blood glucose in real-time. The data is updated to a domestic node (WiFi 802.11b) and transmitted (WiMax 802.16a) to a web portal at the local hospital. Hospital policy kernels and authorizations will be linked to patient-specific data (made possible by the semantic web and web services). The data is analysed in applications with rules or agent impregnated systems which, if necessary, alerts nurse practitioners or doctors should any aberrant blood glucose fluctuation is detected in 'otherwise normal' or 'at risk' seniors. The frequency of insulin administration for the 'at risk' group is coordinated and that limits the onset of glaucoma or keeps the patient glaucoma-free for life! The ability to record, monitor and analyse the variation of blood glucose levels 24/7 in real-time in the 'otherwise normal' group may uncover subgroups with patterns of glucose utilization that may offer clues to variations in geriatric glucose-insulin interactions indicative of other related anomalies (insulinoma, autoimmune, insulin receptor dysfunction).

The petabytes of real-time data, if appropriately available through medical semantic web (semantic metrics), may reveal population genetic traits, which could guide future healthcare policy and funding for biomedical research based on the need of the population or self-service communities (for example, independent living for Alzheimer's patients). For the industrialized world (low birth rate nations) with an ageing population and soaring healthcare costs (>10-15% GDP), the ability to ask right questions may help extract revealing valuable information from real-time data. Patterns for insulin consumption may help hospital procurement managers adapt the 'supply chain' of insulin to diminish inventory carrying cost and reduce wastage without decreasing service level (quality of service). Elimination of the need for monthly blood glucose test will decrease costs for out-patient services and improve resource utilization. Taken together and summed over a number of services and drug procurement, real-time connected data offers potential for risk pooling of services and products by geography to reduce costs for managed healthcare and national health services (for example, the hemorrhaging NHS scheme in UK).

Nano bio-sensors for blood glucose, cartilage degeneration, blood pressure are already available or will soon become available, commercially. Wireless data transmission protocols are in existence along with the ubiquitous internet. The tools necessary to reduce costs, therefore, are at hand, as is, profuse skepticism and paranoia to change. Creation of medical ontologies and the diffusion of the semantic web is still a very slow process. It is further hampered by the sense of "denial" within organizations that has invested in "XML standards" but only after modifying XML to better suit their needs (for example, ebXML). The vast number of such modified-XML, thus, is no longer a "standard" that can catalyse interoperability (see section 5 and Appendix 2) which is crucial when dealing with very high volume data from divergent sources with innumerable devices, such as, sensors.

Creating global consensus to pursue ontological mapping in the biomedical domain may enable the future use of the semantic web to assist a single mother in Accra (Ghana) to learn from the village kiosk that addition of iodized salt to her baby's diet may prevent mental retardation or the inclusion of rice enriched in vitamin A (Golden Rice) may prevent blindness. The semantic web of the future may also reveal to a medical resident that the seemingly intractable pain from an apparent cervical spondylitis may not be an orthopedic case. The resident may uncover that such pain may be triggered by local inflammation reflecting autoimmune reaction with roots in the patient's teen-age years when she had a strep throat (*Streptococcus aureus*) but did not complete the course of the prescribed antibiotics. Epitopes on the cell wall of *S. aureus* are also present (shared) on human leukocyte antigens (HLA-B2).

In this classic tale of autoimmune mimicry, the symptom of pain is caused by inflammation due to autoimmune reactions. Connected thinking from distributed and decentralized information sources, thus, offers application guidance at the point of care (POC) to help the medical resident consider shifting the conventional focus that relates spinal pain to osteopathy or neurology to a deeper consideration of autoimmunity and immunology. Based on this improved understanding, facilitated through the medium of semantic web integrated application at the point of care, the medical resident may offer better quality of service. Rather than recommending generic analgesics, through better diagnosis, the medical resident may prescribe proper anti-inflammatory drugs to sufficiently relieve the discomfort from inflammation. The latter significantly enhances the patient's quality of life.

To improve healthcare delivery, the information technology community (bits) may wish to better understand some of the healthcare tools from advances in biotechnology, biomedical engineering and molecular medicine. In parallel, the biomedical community that interacts with the patient (atoms) may wish to understand the tools for delivery of 'bits' to aid in making better decisions and focus only on those who *need* healthcare services (see illustration below) to rein in costs. Thus, by abstraction, this is a 'bits to atoms' endeavour that fits well with the recursive concept of real-time data linked to processes at the right-time to help adapt or optimize decisions.

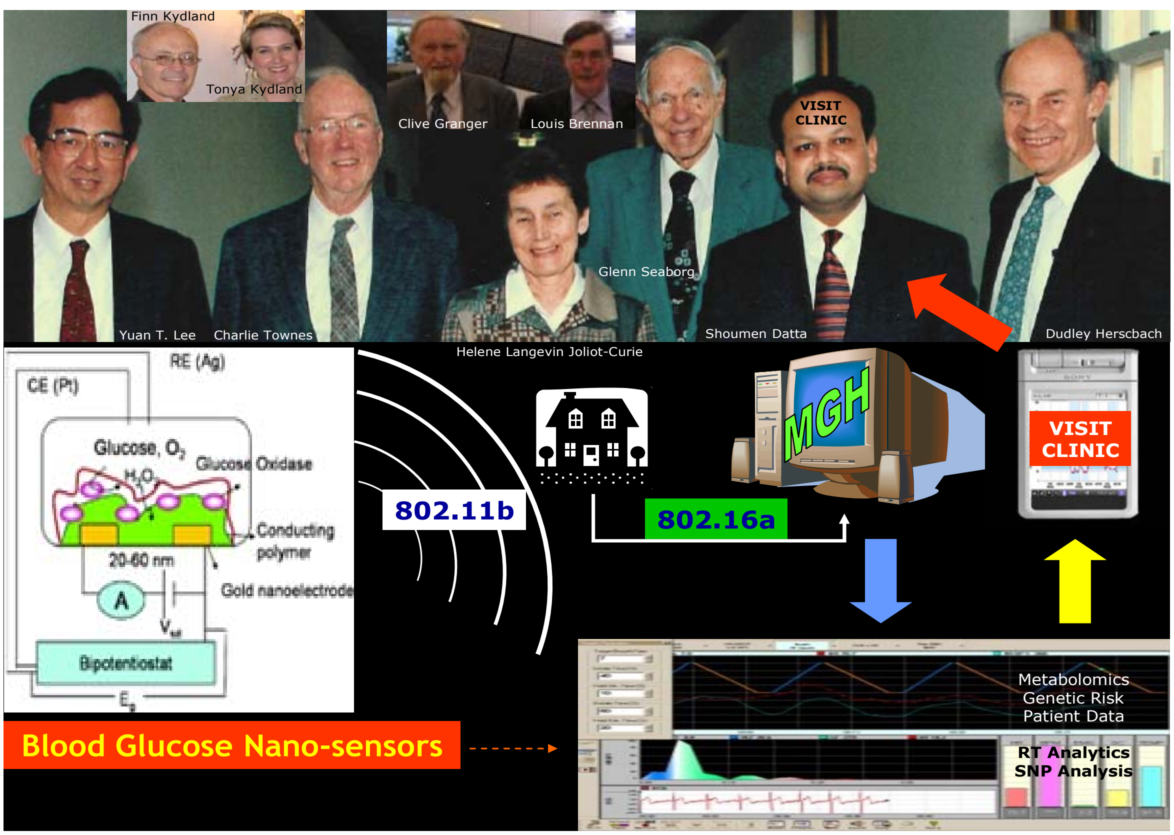


Illustration: Convergence Catalyses Healthcare Delivery - Improved Service Level at Reduced Cost

8.0 CONCLUDING COMMENTS

Development of ontologies that represent the knowledge of the problem space will facilitate use of agent systems within the semantic web infrastructure. Supply chain operations involving buyers and sellers separated by geography and political boundaries must waddle through a host of process intermediaries (finance, logistics, compliance, security) yet reduce cycle times to boost efficiency and hence, profitability. New approaches, especially the emergence of web services and Service Oriented Architecture, taken together with agents and the semantic web offers opportunities for interoperability in scenarios that include supply chain management and healthcare.

In addition to the topics discussed in this paper, the need for system of systems (SoS) interoperability permeates throughout daily usage and common observations (Appendix 6). Therefore, from a broader perspective, a reasonable confluence of these and existing concepts, tools, technologies and standards may collectively, improve adaptability of systems with little or no human intervention. It may even impart some degree of "intelligence" to decision systems to combat uncertainty or improve event-driven applications as diverse as profit optimization, response time in healthcare or hospitals [8], military readiness, emergency planning and detection of potential security threats or risks. To expedite the pace of improvements, an introspective look, at the issue of process illiteracy of technologists and the technical illiteracy of process specialists, may be warranted. Among other things, this paper [33] has also made an attempt, albeit feeble, to offer such a bridging function, as well, by oscillating between process discussions and advances in technology, viewed with the general perspective of offering real-world solutions.

Acknowledgements

This work was made possible by the financial support from the sponsors of the MIT Forum for Supply Chain Innovation. The corresponding author benefited from numerous discussions with academic, industry and government thought leaders, worldwide. Special thanks are due to Satya Sahu (Government of India Customs and Excise, New Delhi, India); Ralph Swick and Lalana Kagal (World Wide Web Consortium, MIT); Jon Stine, Martin Guttmann and Gary Haycox (Intel Corporation); Bob Betts (MainStreet Applications); Joseph Salvo, Lynn DeRose, John Carbon, Patricia Mckenzie and Charles Theurer (GE Global Research); Kimmo Lindqvist and Willem Duijf (GE Equipment Services); Franz Dill (Procter & Gamble); Andy Mulholland and Ard Jan Vethman (Capgemini); Peter Koudal and Ajit Kambil (Deloitte Research); Reuben Slone (OfficeMax); Jeff Wilke and Micah Samuels (Amazon); Samy Uthuruswamy and Shang-Tae Yee (General Motors); Paul Kern (The Cohen Group); Greg Parlier (SAIC); Kevin Larson (AAR Corporation); Robert Ciskowski, Grace Lin and Robert Mayberry (IBM); Alexander Renz (Microsoft); Franz Hero (SAP AG, Walldorf, Germany); Hideki Sakai (SAP Japan, Tokyo, Japan); Satoshi Arizumi (Sony SNC, Tokyo, Japan); Masami Kinugasa and Sam Nakane (PoweredCom, Tokyo, Japan); Susumu Takahashi (Agnie Consulting, Tokyo, Japan); Eelco de Jong (LogicaCMG, Hoofdorp, Netherlands); Michel Danet, Ray McDonagh and Dietmar Jost (World Customs Organization, Brussels, Belgium); Lion Benjamins (Stockway, Helsinki, Finland); Roongrote Rangsiyopash, Bhanumas Srisukh, Bunn Kasemsup and Pimchanok Chuengcharoen (Siam Cement Group, Bangkok, Thailand); Rajanish Dass (Indian Institute of Management, Ahmedabad, India); Louis Brennan, Kate Phillip Connolly and John Murray (Trinity College, Dublin, Ireland); Petri Kalliokoski and Santu Toivonen (VTT, Espoo, Finland); Vesa Salminen (Lappeenranta University of Technology, Lappeenranta, Finland); Raimo Pulkkinen and Heidi Lindroth (TEKES, Helsinki, Finland); Kenth Lumsden and Gunnar Stefansson (Chalmers University of Technology, Goteborg, Sweden); Yichen Lin (National University of Tainan, Taiwan); JrJung Lyu (National Cheng Kung University, Taiwan); Han Pang Hwang (National Taiwan University, Taipei, Taiwan); Dominique Estampe (ISLI, Bordeaux, France); Alain Lempereur and Michele Pekar (ESSEC Business School, Paris, France); Gerry Frizelle (University of Cambridge, UK), Yoon Chang (Hankuk Aviation University, Goyang-City, Korea); Alexander Zier (Hasso Plattner Institute, Berlin, Germany); Farokh Eskafi (TagArray Systems); William Reckmeyer (Stanford University and San Jose State University); Michael Barr (California Agricultural Leadership Foundation); Gita Swamy, Ananth Raman and Clark Gilbert (Harvard Business School); Howe Shrobe, Jonathan Bachrach, Sanjay Sarma, David Brock, Daniel Engels, Rory O'Connor, John Williams, Hui Li, Charles Fine, Mike Li, David Simchi-Levi and Joseph Sussman (Massachusetts Institute of Technology).

References

[1] Anderson, D. L. and Lee, Hau (1999) Synchronized supply chains <www.ascet.com/documents.asp?d_ID=198>

[2] Berners-Lee, T. (1998) Semantic Web Roadmap <www.w3.org/DesignIssues/Semantic.html>

[3] Bowersox, D. J., Closs, D. J., and Stank, T. P. (1999) Making supply chain integration a reality. CLM.

[4] Datta, S. (2006) Advances in SCM: Improving Forecasting Accuracy (http://esd.mit.edu/WPS/esd-wp-2006-11.pdf) Working Paper. Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA

[5] Chopra, S. and Meindl, P. (2001) Supply Chain Management: Strategy, Planning, and Operations. Prentice Hall.

[6] Collis J. and Nudumu D. (1999) ZEUS Agent Building Toolkit. Intelligent Systems Research Group, BT Labs

[7] Datta, S., *et al* (2004) Adaptive Value Networks in *Evolution of Supply Chain Management*. Kluwer (KAP).

[8] Datta S. (2004) Adapting Decisions, Optimizing Facts, Predicting Figures. Working Paper. Engineering Systems Division, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 http://supplychain.mit.edu/innovation/shoumen.htm

[9] Datta S. (2001) SCM in Real Time: An Introduction <http://supplychain.mit.edu/innovation>

[10] Lee, H., Padmanabhan, P. and Whang, S. (1997) The Bullwhip Effect in Supply Chains. Sloan Management Review 38 93-102

[11] Fayez M., Rabelo L. and Mollaghasemi M. (2005) Ontologies for the Supply Chain Simulation Modeling. Proceedings of 2005 Winter Simulation Conference

[12] Hakim S.W., Bowman G. W. and Siedenstat, P. (2005) Privatizing Transportation Systems. Greenwood.

[13] Worthen, B. (2006) Customs Rattles the Supply Chain <www.cio.com>

[14] Datta, S. (2006) Risk in the Global Supply Chain <http://events.wcoomd.org/wcoesgi_articles.htm>

[15] Joseph Salvo, GE Global Research, personal communication <www.trailerservices.com>

[16] Kalakota, R. and Robinson, M. (2001) e-Business 2.0: Roadmap for Success. 195-263. Pearson Education Asia.

[17] Farrukh, N. (2004) Demonstration of Web Content Management using ebXML Registry. XMLEurope 2004.

[18] Pathak S.D., Nordstrom, G. and Kurokawa, S. (2004) Modelling of Supply chain: Multi-Agent Approach

[19] David Brock, MIT Data Center, personal communication <www.mitdatacenter.org>

[20] Andy Mulholland, Capgemini, personal communication <andy.mulholland@capgemini.com>

[21] Sycara, K. P. (1998) Multi-agent Systems AI Magazine, Summer 1998, 79-92.

[22] Yufei Y., Liang, T.P. and Zhang, J.J. (2005) Using Agent Technology to Support SCM. Working Paper.

[23] Datta, S. (2000) Microarray http://supplychain.mit.edu/innovation/shoumen.htm (#2)

[24] Chandrasekharan, S. (2002) Semantic Web: a distributed cognition view. Technical Report 2002-13, Ottawaa

[25] Toivonen, S. (2004) Distributing Cognition in the Semantic Web *(personal communication)*

[26] www.csail.mit.edu/research/abstracts/abstracts03/web/02berners-lee.pdf

[27] Dertouzos, M. (2001) The Unfinished Revolution. Harper Collins.

[28] www.cs.umd.edu/~hendler

[29] Forzani, E.S., *et al* (2004) Conducting Polymer Nanojunction Sensor for Glucose Detection. Nano Letters 4 1785-1788

[30] Imagine that I hire a clown messenger to deliver balloons to my customers on their birthdays. Unfortunately, the service transfers the addresses from my database to its database, not knowing that the "addresses" in mine are where bills are sent and that many of them are post office boxes. My hired clowns end up entertaining a number of postal workers, not necessarily a bad thing, but certainly not the intention. With the tools of the semantic web, an address that is a mailing address can be distinguished from one that is a street address and both can be distinguished from an address that is a speech. Our current syntactic web is incapable of such distinctions (meaning and context). This is not the end of the clown story, because two databases may use different identifiers for what is, in fact, the same concept, such as zip code. A program that wants to compare or combine information across the two databases has to know that these two terms are being used to mean the same thing (see section 5.3 in this paper). Ideally, the program needs to discover such common meanings for whatever databases it encounters. For example, an address may be defined as a type of location and city codes may be defined to apply only to locations. Classes, subclasses and relations among entities are a very powerful tool for web use. We can express a large number of relations among entities by assigning properties to classes and allowing subclasses to inherit such properties. If city codes must be of type city and cities generally have web sites, we can discuss the web site associated with a city code even if no database links a city code directly to a web site.

Inference rules in ontologies supply further power. Ontology may express the rule "if a city code is associated with a state code and an address uses that city code, then that address has the associated state code." A program could then readily deduce, for instance, that a Cornell University address, being in Ithaca, must be in New York State, which is in the US and therefore should be formatted to US standards. The computer doesn't truly "understand" any of this information, but it can now manipulate the terms much more effectively in ways that are useful and meaningful to the human user.

Ontologies can enhance the functioning of the web in many other ways. They can be used in a simple fashion to improve the accuracy of web searches. Advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules. An example of a page marked up for such use is www.cs.umd.edu/~hendler. If you send your Web browser to that page, you will see the normal web page entitled: Dr James A Hendler. As a human, you can readily find the link to a short biographical note and read there that James Hendler received his PhD from Brown University. A computer program trying to find such information, however, would have to be very complex to guess that this information might be in a biography. For computers, the page is linked to an ontology page that defines information about computer science departments. For instance, professors work at universities and they generally have doctorates. Further mark-up on the page (not displayed by the typical web browser) uses the ontology's concepts to specify that James Hendler received his PhD from the entity described at the URI http://www.brown.edu (the web page for Brown University, Rhode Island). Computers can also find that James Hendler is a member of a particular research project, has a particular e-mail address. In the world of the semantic web, all that information is readily processed by a computer and may be used to answer queries (from where did James Hendler receive his PhD degree) that currently require a human to sift through the content turned up by a search engine.

Advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules. Suppose you wish to find Miss Cook you met at a trade conference last year. You do not remember her first name, but you remember that she worked for one of your clients and that her son was a student at your alma mater. An intelligent search program can sift through all the pages of people whose name is "Cook" (sidestepping all the pages relating to cooks, cooking, the Cook Islands and so forth), find the ones that mention working for a company that is on your list of clients and follow links to web pages of their children (assuming the children have web sites) to track down if any are in school, at the right place. Suppose Miss Cook's contact information is located by an online service which places her in Dublin. Naturally, you want to check this, so your computer asks the service for a proof of its answer, which it promptly provides (exchange of proofs) by translating its internal reasoning into the semantic web's unifying language. An inference engine in your computer readily verifies that this Miss Cook indeed matches the one you were seeking.

[31] The traditional control systems approach to deal with the challenges of complexity in industrial operations is to deploy model predictive control (MPC) to direct distributed control systems (DCS). Semantic interoperability and agent based systems may evoke a new paradigm in thinking about complexity and solutions to resolve complexity. However, understanding the roots of complexity is essential. The excerpt in Appendix 1 [below] is provided from a talk by Mujeeb Ahmed, senior member of ISA. This description of COMPLEXITY [below] is edited and adapted from a presentation at the 47th Annual Power Industries (POWID) Conference and the 14th Annual Joint POWID & Energy Power Research Institute (EPRI) Controls and Instrumentation Conference, 2004.

[32] See Appendix 7: Web Services (contributed by Dr Shang-Tae Yee, General Motors)

[33] This paper has attempted to be several "things" to serve several purposes. The attempt may not be successful and it is solely the fault of the corresponding author (S. Datta). The paper is deliberately over-simplified with an eye toward the "bridging" function. Exploration of ideas are incomplete and often plagued with digressions. All errors of content or coherence are due to the corresponding author. The paper may be unsatisfactory for many experts and practitioners but the amalgam of ideas may spark new thinking. This is an exploration to catalyse convergence and innovation. In addition to named contributors, the authors have freely used several sources of information to 'connect the dots' and show how distant disciplines, if coalesced, may offer new directions. The list of references is seriously incomplete. It may be amply clear that the original research is not due to the authors. Opinions and comments expressed here are attributable to the corresponding author and do not represent the views of MIT as an institution or the contributors or their organizations. For experts, there may be nothing 'new' in this article and for practitioners the vision may be too vague for implementation. But, it is the synthesis of ideas from a variety of sources, when presented in confluence, as suggested in this paper, that may be catalytic in the transformation of decision support systems to adapt or perhaps, with time, to develop predictive intelligence. Please email shoumen@mit.edu or write to Dr. Shoumen Palit Austin Datta, Research Scientist, Research Director & Co-Founder, MIT Forum for Supply Chain Innovation, School of Engineering, MIT.

APPENDIX 1: **COMPLEXITY**

Control systems rooted in feedback techniques lack the power to solve complex problems. Human operators with their experience and ingenuity can steer feedback control systems to the optimum point in the operating space. Relying on people for this complex task is not the most productive approach because of human cognitive characteristics and constraints in workload. A prudent and popular approach is to deploy model predictive control (MPC) to direct the distributed control system (DCS) but the MPC–DCS is under human supervision. Consequently, people perform decision and control functions. An ensemble of sophisticated systems can certainly boost the "intelligence" but the value obtained from the ensemble is critically dependent on three factors. 1st is the ability of the MPC. If it cannot manage computational complexity, nonlinearity and uncertainty is unlikely to add much value. 2nd is the synchronized functioning of the individual systems. MPC cannot add value if the DCS it is directing cannot maintain process variables at set point. 3rd is the complexity people face while interacting with the systems. A high level of combinatorial or dynamic complexity can make people error-prone during operations. Not all installations of MPC deliver the expected benefits. A primary cause for failure is the inability of the problem-solving agents to manage the inherent complexity during operation. The origins of such inability usually reflect insufficient attention given to complexity during the design phase.

Complexity's Roots

Optimal control solution to improve performance in multiple dimensions involves a higher level of mathematical and computational complexity. Obtaining the conditions for optimality requires mathematical manipulation of the system dynamic equations and the equation for performance measure. System dynamics relate controlled variables and manipulated variables. Naturally, these introduce constraints. The resulting equations are often nonlinear differential equations of second or higher order. Besides, they could be time varying. Furthermore, future information about the system's behavior and the performance measure is required. In other words, the reward must be clear before making the decision.

Evaluating performance on the basis of individual criteria can be relatively simple. But, determining what is best in a global sense *(global optimization)* presents greater complexity. Oftentimes a global performance criterion, such as cost expressed in mathematical form, is an aggregation of several distinct criteria, each of which epitomizes a particular dimension or performance. Some aspects of performance may be difficult to measure accurately or reliably. No solution may ever exist to entirely satisfy all the individual criteria. Compromises enter the picture. Ranking criteria influence the extent to which lower performance in one area is acceptable for higher performance in another.

Finding the right balance of inputs to minimize or maximize a comprehensive performance measure involves combinatorial complexity, which increases with the number of factors in decision-making. Some combinations of inputs *(parameters)* will be viable, others may be impractical. Still others may be ineffective. Some values of outputs and states just cannot happen. A further complication enters through the presence of global and local minimums in the search space. The challenges in "searching" for the best solution include identifying viable combinations, rejecting combinations that offer no improvement and conducting the search to converge on a global minimum. Besides, the solution must exist within the bounds of physical realities and the solution must happen in real time.

Dynamic complexity increases while interacting with different systems in a hierarchical deployment. Numerous time variant relationships, frequency of interactions over time and outcomes that are difficult to predict, are the principal factors that contribute to dynamic complexity. Correlations between causes and effects may not be simple. Effects may be disproportional to causes or counterintuitive. Effects may unfold over many time scales. System behavior may vary with operational point and time. One transition may force other transitions. Interdependency may propagate or amplify the effects of disturbances or failures. Hence, emergent behavior is the result of a mélange of actions. Consequently, predicting overall behavior is difficult and static characteristics of a system offer little help to forecast such dynamic behavior.

Control system behavior is primarily algorithmic. At a particular level of abstraction a collection of algorithms works together to achieve a higher-level objective. Intelligence in control systems comes from the pre-designed (static) algorithms. It is adequate to manage operational tasks anticipated and addressed by designers. All too often machines may not cope with events unanticipated by designers or even simple cases, for example, a MPC with no model between heat input and steam temperature is unlikely to exert control on steam temperature.

Humans, on the other hand, have superior abilities to cope with unanticipated situations. They accomplish operational tasks by automated behavior, conscious execution of procedures and conscious problem solving. They can change strategies instantaneously if new information refutes current beliefs or solve unusual problems by utilizing experience, intuition, imagination and creativity. In other words, humans can enhance the control algorithm online in real time.

People as well as control systems work in a web of interrelationships and interdependencies. Interaction between machines is generally limited to data exchange. Organize people to adapt. A centralized, functional hierarchy has long been a staple of organizational methods. Sure, such a structure assigns responsibility, establishes accountability, and provides stability. But, this is more likely to occur in routine situations only. In unusual situations, the organization's effectiveness may decline. Redefining strategy, reassigning functions, and maintaining coordination can suffer. Worse, the organization may be unable to quickly adapt to new demands.

APPENDIX 2: Failure to Standardize: Vast number of modified-XML

4ML	ARML	BiblioML	CIDX	eBIS-XML	HTTP-DRP	MatML	ODRL	PrintTalk	SHOE	UML	XML F
AML	ARML	BCXML	xCIL	ECML	HumanML	MathML	OeBPS	ProductionML	SIF	UBL	XML Key
AML	ASML	BEEP	CLT	eCo	HyTime	MBAM	OFX	PSL	SMML	UCLP	XMLife
AML	ASML	BGML	CNRP	EcoKnow	IML	MISML	OIL	PSI	SMBXML	UDDI	XML MP
AML	ASTM	BHTML	ComicsML	edaXML	ICML	MCF	OIM	QML	SMDL	UDEF	XML News
AML	ATML	BIBLIOML	Covad xLink	EMSA	IDE	MDDL	OLifE	QAML	SDML	UIML	XML RPC
AML	ATML	BIOML	CPL	eosML	IDML	MDSI-XML	OML	QuickData	SMIL	ULF	XML Schema
ABML	ATML	BIPS	CP eXchange	ESML	IDWG	Metarule	ONIX DTD	RBAC	SOAP	UMLS	XML Sign
ABML	ATML	BizCodes	CSS	ETD-ML	IEEE DTD	MFDX	OOPML	RDDl	SODL	UPnP	XML Query
ACML	AWML	BLM XML	CVML	FieldML	IFX	MIX	OPML	RDF	SOX	URI/URL	XML P7C
ACML	AXML	BPML	CWMI	FINML	IMPP	MMLL	OpenMath	RDL	SPML	UXF	XML TP
ACAP	AXML	BRML	CycML	FITS	IMS Global	MML	Office XML	RecipeML	SpeechML	VML	XMLVoc
ACS X12	AXML	BSML	DML	FIXML	InTML	MML	OPML	RELAX	SSML	vCalendar	XML XCI
ADML	AXML	CML	DAML	FLBC	IOTP	MML	OPX	RELAX NG	STML	vCard	XAML
AECM	BML	xCML	DaliML	FLOWML	IRML	MoDL	OSD	REXML	STEP	VCML	XACML
AFML	BML	CaXML	DaqXML	FPML	IXML	MOS	OTA	REPML	STEPML	VHG	XBL
AGML	BML	CaseXML	DAS	FSML	IXRetail	MPML	PML	ResumeXML	SVG	VIML	XSBEL
AHML	BML	xCBL	DASL	GML	JabberXML	MPXML	PML	RETML	SWAP	VISA XML	XBN
AIML	BML	CBML	DCMI	GML	JDF	MRML	PML	RFML	SWMS	VMML	XBRL
AIML	BML	CDA	DOI	GML	JDox	MSAML	PML	RightsLang	SyncML	VocML	XCFF
AIF	BannerML	CDF	DeltaV	GXML	JECMM	MTML	PML	RIXML	TML	VoiceXML	XCES
AL3	BCXML	CDISC	DIG35	GAME	JLife	MTML	PML	RoadmOPS	TML	VRML	Xchart
ANML	BEEP	CELLML	DLML	GBXML	JSML	MusicXML	PML	RosettaNet PIP	TML	WAP	Xdelta
ANNOTEA	BGML	ChessGML	DMML	GDML	JSML	NAML	PML	RSS	TalkML	WDDX	XDF
ANATML	BHTML	ChordML	DocBook	GEML	JScoreML	xNAL	P3P	RuleML	TaxML	WebML	XForms
APML	BIBLIOML	ChordQL	DocScope	GEDML	KBML	NAA Ads	PDML	SML	TDL	WebDAV	XGF
APPML	BIOML	CIM	DoD XML	GEN	LACITO	Navy DTD	PDX	SML	TDML	WellML	XGL
AQL	BIPS	CIML	DPRL	GeoLang	LandXML	NewsML	PEF XML	SML	TEI	WeldingXML	XGMML
APPEL	BizCodes	CIDS	DRI	GIML	LEDES	NML	PetroML	SML	ThML	Wf-XML	XHTML
ARML	BLM XML	CIDX	DSML	GXD	LegalXML	NISO DTB	PGML	SAML	TIM	WIDL	XIOP
ARML	BPML	xCIL	DSD	GXL	Life Data	NITF	PhysicsML	SABLE	TIM	WITSML	XLF
ASML	BRML	CLT	DXS	Hy XM	LitML	NLMXML	PICS	SAE J2008	TMML	WorldOS	XLIFF
ASML	BSML	CNRP	EML	HITIS	LMML	NVML	PMML	SBML	TMX	WSML	XLink
ASTM	BCXML	ComicsML	EML	HR-XML	LogML	OAGIS	PNML	Schemtron	TP	WSIA	XMI
ARML	BEEP	CIM	DLML	HRMML	LogML	OBI	PNML	SDML	TPAML	XML	XMSG
ARML	BGML	CIML	EAD	HTML	LTSC XML	OCF	PNG	SearchDM-XML	TREX	XML Court	XMTP
ASML	BHTML	CIDS	ebXML	HTTPL	MAML	ODF	PrintML	SGML	TxLife	XML EDI	XNS

(Source: David Brock, MIT)

APPENDIX 3: Elusive Quest for Adaptable Interoperability

Globalisation forces cultural amalgam that depends on efficacy of change management to reap the benefits. Fueled by resistance and paranoia, changes to embrace or accommodate globalization are few and far between. Drivers of change management are amorphous and differences in policy, standards, trust and financial inequalities create further hindrance. Governments, organizations and businesses are exposed to these changes without adequate training, tools or frameworks to enable them to deal with global commerce. Regulatory compliance may be often divorced from business needs and decisions may be made based on inaccurate or corrupt data. Processing of data to yield valuable actionable information remains largely unexplored and is often plagued by lack of visibility due to inconsistent interoperability. These problems are compounded by the inability of systems to adapt or respond in near-real time despite the advances in technology and progress toward intelligent decision systems.

Interoperability between systems and adequate operational transparency may help stem some of the frustration of businesses unable to fully enjoy the fruits of globalisation, for example, outsourcing or offshoring. On the other hand, regulatory agencies must remain vigilant to ensure security through tracking and tracing of goods to prevent disenfranchised individuals from taking advantage of the movement of objects between geographic boundaries. Therefore, tracking data for a sealed container from the Port of Hong Kong must be visible in systems of different countries while the container is at sea or if the vessel makes stops *en route* before arriving at its final destination. Currently, the occurrence of "black holes" of information may not be surprising due to operational and/or technical inadequacies between systems. The integrity of the physical seal of the container and its location is as important as the identity and source of the goods. While efforts are underway to optimize secure container shipments, the operations are often at the mercy of the local logistics providers who may horde the data about the source and identity of goods. Ideally, goods must be traced further back into the business supply chain to ensure credibility of the source. At present, visibility of the "chain" is quite restricted if not completely unavailable.

There are no easy "one shoe fits all" solutions to these problems. There is also room for debate as to and whether the depth of collaborative visibility of the supply chain can be turned into a profitable advantage rather than compromise true competition. The investment necessary to gain visibility and transparency both in terms of cost as well as change management can only flourish if it is a collaborative venture that includes regulatory agencies, such as, customs. All parties must be equally determined to remain cognizant about operational efficiency. However, even for pre-agreed issues, the ability to generate a bird's eye view of the sequence of processes are plagued by the lack of interoperability between customs and business systems. Shared data models, common process descriptions (ebXML Registry) and alphanumeric serialization attempts (EPC) are aimed to offer some degree of standardization. It follows that any attempt at standardization requires sufficient adoption of the so-called standards in order to harvest the anticipated efficiencies from economies of scale of adoption.

Adoption of any standard may be broadly described as a convergence of the inclination to change the *status quo* with the availability of the tools or infrastructure that enables the adopted standards to be practised. The impact of software in this quagmire is inescapable since it is one key infrastructure for operation of systems. Hence, it is clear to grasp why lack of systems interoperability may be precipitated by deployment of a heterogeneous mix of proprietary software in systems that are expected to communicate and interoperate. Operations in the transactional domain in some countries tend to align with UN EDIFACT standard. The processes that were "standardized" by this "directory" are restrictive in scope and syntactically describe a limited number of operations deemed "classical" amidst the heterogeneity. The increasingly uncertain emergence of complexity in goods, products and services calls for innovative methods that are flexible, adaptable, reconfigurable and responsive to volatility of supply-demand. Agility of infrastructure is a pre-requisite for nimble systems. In addition, the core tools of the system, for example, software, must be able to understand and communicate between systems in order to enable interoperability.

The unrelenting emphasis on the need for interoperability may evoke the thinking that we must all use the same software as infrastructure to ensure communication and interoperability between systems. It is true that the latter may be one way to reduce uncertainty but it is certainly not a reasonable *modus operandi* nor can we recommend any specific software packages. The diffusion of the semantic web may facilitate adaptable interoperability between systems without the need for complete elimination of heterogeneity of systems.

In the world of the syntactic web, interoperability is possible but at a higher cost of maintenance. The ability to connect between unrelated systems through the use of "connectors" that may use (adopt) one or more standards, formats or frameworks, specified by groups or associations, may offer functional interoperability that may serve useful purposes. For widespread use, connectors must be rapidly implementable, capable of data or information exchange, preferably possess some "intelligence" or analytics and must be easy to upgrade but remain adaptable.

The adaptability part of the "adaptable interoperability" paradigm refers to processes that are used in a variety of operations. If the processes remain unique then businesses will incur a high cost to interact, globally. If the processes are rigid then changes will be slow and painful. Because, nations are unlikely to agree on any one process map and its financial manifestations, it is imperative that process descriptions must entertain diversity with the greatest design flexibility to rapidly adapt to change. In the syntactic web, interoperability may be accelerated if processes are designed in a manner that can be easily "translated" in terms of the semantic content even though the syntax of words in the description may vary between operations. For example, ports in one country may specify "containers per vessel" while the same process may be described as "unit containers in each cargo ship" by another port system. In the semantic sense they are identical but the syntax is sufficiently different to create barriers for interoperability between two software systems hard-coded to one process or the other.

The technical issues cryptic in the above example may find some similarities with human language. Translation from English into French achieved by the "brain" of the translator. It is equivalent to (in this case) the software in the system that extracts the semantic context from the syntactic description of the process (without interruption or human intervention) to execute the desired functionality.

Currently, systems fail to understand that "containers per vessel" and "unit containers in each cargo ship" refer to the same criteria. The latter begins to create interoperability barriers. It is also possible to query such data ("unit containers in each cargo ship") to obtain statistic, such as, "containers loaded per day" or "containers loaded in cargo bay one between 0800 and 1000" but only if the relationship between "unit containers in each cargo ship" and "containers loaded in cargo bay one between 0800 and 1000" is "understood" by the system. Such queries are quite complex if syntactic software is used because each query may be process dependent or described by a set of words that must be hard-coded without room for adaptability. To obtain the statistic "containers loaded in cargo bay one between 0800 and 1000" from the data set "containers per vessel" requires the system to analyse loading or unloading data for containers by cargo bay numbers and link with the frequency of operation by the hour (which implies understanding that 1000 is 10am). If all of these links were to be programmed (hard-coded) for every possible query a human operator may request, the complexity and its cost of maintenance may be astronomical.

Lack of systems level "understanding" could be a blow to the success of non-obvious relationship analysis as well as efficiency of searching connected information. For example, even the advanced search algorithms used by Google may fail to identify the news article ("*China will not buy chips from Taiwan"*) if an user searches for "semiconductor industry business news in Asia" as the search phrase in the syntactic world of Google and other search engines.

Current software systems offer a limited number of descriptions and few analytical tools in order to maintain system complexity at a manageable level. These systems use 1:1 links, requires expensive resources and often calls for re-programming since they are based on fixed processes. The US Government Accountability Office predicts in a recent report that US Department of Defense may spend nearly $14 billion on software changes in 2006 (Federal Computer Week, 31 July 2006). The latter are detrimental to the economy as a whole and in particular to the dynamic necessity of process change. It stands to reason, that adaptable interoperability is essentially a culmination of meaningful convergences that offers the potential to catalyse innovative solutions. This scenario is not an 'IT' development. This is a fundamental advance in the ability of science to foster machine intelligence which builds, in this case, on the evolution toward the "open source technology" of the semantic web, provided data and processes are machine readable. Semantic systems are emerging and may be implemented without any new discoveries.

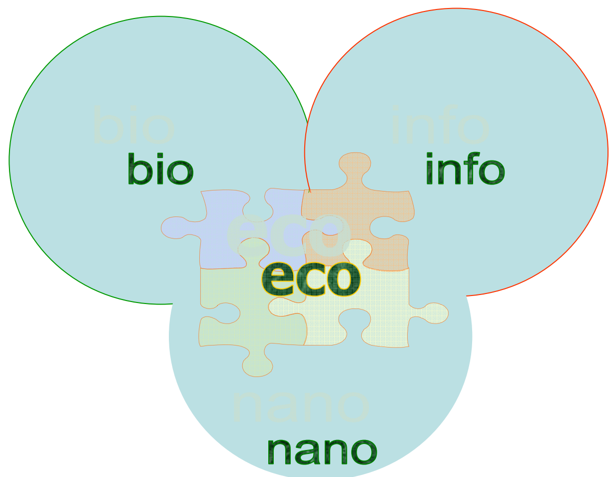
For fruition of interoperability it is vital to appreciate the breadth of convergences that may be necessary yet rarely found under one roof. Horizontal imagination and the ability to construct fruitful associations to 'connect the dots' or build on past convergences, are essential elements of this vision. For the past few centuries, the Dutch have used auctions for flower trade. English auctions, which are strikingly different from Dutch auctions, were used by traders and real estate agents in Great Britain and North America. But ordinary people in the 20th century did not trade through auctions. In sharp contrast, 70 million people in October 2005 used auction pricing to buy-sell goods.

To most readers, the mention of eBay may not seem unusual but eBay is rapidly evolving as a text-book example of interoperability through convergence. Despite its novelty, eBay did not offer us anything new. This conundrum may be traced back to Henry Ford, who said, "I invented nothing new. I assembled into a car the discoveries of other men behind whom were centuries of work." eBay assembled into its business model the fruits from the development of the transistor, internet, world wide web and auction pricing. The participants on eBay use interoperable mechanisms that connect the millions, seamlessly. Pre-existing tools and technologies created the novelty of eBay, which is a sum of parts, where none of the parts are new or were inventions due to eBay.

Convergence enabled Pierre Omidyar, the founder of eBay, to innovate a profitable business process using existing tools. The power of innovation latent in the confluence of concepts may trigger exploration of some of these unorthodox juxtapositions and unchartered territories peppered with imperfect ideas. Such endeavours may help shape or catalyse adaptable interoperability in modern environments that are guaranteed to change. Because we cannot direct the winds of change, it is better to have the knowledge of when and how to adjust the sails.

APPENDIX 4: Convergence Catalyses Global Public Goods to Enhance Sustainable Economic Growth

CONVERGENCE



Global Public Goods

Disclaimer: Suggestions in this article are by no means comprehensive for any national growth strategy. These suggestions are only minor segments of any economic plan, albeit incomplete and inadequate. Important issues such as, infrastructure (land, sea, air) and defense are not discussed, yet, but are vital. The topics (bio, info, nano, eco) mentioned here are generic sketches, not a roadmap of their true potential or future scope. Eco (energy) issues pertaining to hydrogen and ethanol are aimed to provoke debate. One purpose of this appendix is to suggest some of the exciting possibilities that may emerge from enabling strategies that seek and exploit convergence of ideas, tools, technologies, concepts and standards. None of the original research are that of the author. It is a mere collection of diverse facts laced with the author's version[7] of how their use may be beneficial if used in confluence with other ideas. This may only serve as a provocative guide for public-private-academic partnership or explorations. Opinions and comments in this article are solely due to the author. Views expressed in this article do not express the opinion of MIT or any individual, institution or organization mentioned in this appendix.

[7] The physicist Leo Szilard once announced to his friend Hans Bethe that he was thinking of keeping a diary: "I don't intend to publish. I am merely going to record the facts for the information of God." "Don't you think God knows the facts?" Bethe asked. "Yes" said Szilard. "He knows, but He does not know this version of the facts." (Hans Christian von Baeyer in *Taming the Atom*)

1.0 Ideas and Innovation: Past, Present and Future

Biotechnology (bio) and information (info) technology are rewarding directions for economic growth (basic and applied) but traditional approaches are over-subscribed. However, it appears that nano science and technology is not yet a core in the traditional vein nor is the concept of eco which, in part, targets elimination of non-renewable fossil fuel in favour of using hydrogen with provisions for usage of ethanol. Profound economic benefits for poor nations and political changes for affluent nations may result from the hydrogen-ethanol economy and from nano.

Commercial impact from nano may be significant for nations geared to take advantage of nano-engineering. For example, Wilson's Double Core tennis balls use nano-composite clay to keep them bouncing 2 to 4 times longer. Used at the 2002 Davis Cup, they are now the official balls of this championship event. Babolat has introduced super-strong carbon nanotubes in tennis rackets for improved torsion and flex resistance. Nanotube enhanced golf clubs may be arriving soon. Drug delivery by nano-encapsulation is nearing completion of human trials. Use of nano in the food industry may boost security and safety.

Convergence is key to creating 'innovation' societies in 'knowledge' economies. Workforce creation through national policies and gender-blind encouragement in secondary and post-secondary schools for pursuit of science and mathematics, are key pillars for sustainable economic growth leading toward quality of life improvements. However, educational curriculum that lacks analytical rigour but promotes learning of facts and regurgitation in tests are incapable and insufficient to address the needs of any 'innovation' society[8].

Structured inclusion of experimentation from pre-teen years is essential to foster creativity. The latter demands teachers who have had exposure to basic research in core scientific principles. Policies to attract and retain teachers with some research experience to teach in secondary schools are essential for long-term sustainability of any innovation society. It is also necessary to create programs that allow cross-pollination between secondary school and university faculty[9] as well as secondary school student project activities that may be supervised by university students or faculty involved in research.

---

[8] Innovation is not a novelty of our times nor is it the monopoly of the 20th century. A clear example of innovation achieved through planned systemic improvement in analytical rigour may be traced to the educational reforms (10-year gymnasium school) instituted around 1810 by Wilhelm von Humboldt of Prussia (modern Germany). The results were impressive. According to John Derbyshire (in *Bernhard Riemann and the Greatest Unsolved Problems in Mathematics*) a list of dozen greatest mathematicians at work in 1800 may include: Argand, Bolyai, Bolzano, Cauchy, Fourier, Gauss, Germain, Lagrange, Laplace, Legendre, Monge, Poisson and Wallace. The most striking feature of this list is the near-total absence of Germans except Gauss (but nine French mathematicians). A similar list in 1900 includes: Borel, Cantor, Carathéodory, Dedekind, Hadamard, Hardy, Hilbert, Klein, Lebesgue, Mittag-Leffler, Poincaré, Volterra. Four Frenchmen, one Italian, one Englishman, one Swede and ***five*** Germans!

[9] In 1724, Peter the Great issued a decree establishing an Academy at St. Petersburg which was designed to be much more than the "research" institution that academies were supposed to be in the 18th century. This Academy was to conduct research (the "traditional" role) as well as teach (traditional function of the "university" in the 18th century) and host a secondary school.

It is imperative that nations do not replicate the mistakes of K-12 education system of US. A seminal report from the Carnegie Foundation (1996) suggested reasons for declining effectiveness of US basic education. The report reveals that 51% of mathematics teachers in US public schools (K-12) never took math as a part of their college curriculum. A third of the science teachers never took science as a major in college. A national survey of US high school physics found 18% of physics teachers had degrees in physics while 11% had *some* degree in physics education but not in core physics while 27% teachers had neither a degree nor experience in physics.

Innovation germinates through confluence of education, respect, discipline and tolerance. However, the classical definition of education, respect, discipline and tolerance is insufficient to gel the confluence necessary to shape imagination or stimulate innovation (Albert Schweitzer in *The Decay and Restoration of Civilization*). In *Aims of Education*, the mathematician-philosopher Alfred North Whitehead observes, "the University of Cambridge which had done best at teaching mathematics is the one from amongst whose graduates have come more of the English poets, while Oxford which has specialized in the humanities, has tended to turn out writers who have attained, on the whole, a high level of mediocrity. I suppose that by the time one has discussed literature with a witty and learned professor, you know what has been achieved and how good it is. You become respectful and begin to wonder who am I to do better?''

The element of leadership as well as the ability to lead in unchartered territories, thus, is a complex trait to cultivate and must start with strengthening primary and secondary education. Leadership and leaders must inculcate vision but generally they end up only with "views" that are often skewed because it is a rather difficult balancing act between learning to learn how to acquire pragmatic knowledge and indulging in informed intuitive imagination.[10]

In 1830, Ferdinand de Lesseps, a French diplomat in Cairo, dreamed of linking Europe and Asia (Mediterranean and Red Seas) by cutting a canal through 118 miles of arid land at a cost of FFR 200 million. In November 1869, the Suez Canal opened and fewer than 500 ships passed through in 1870, far below the projected ROI (return on investment). Dividends failed to materialize on 'shares of passion' and Egyptians desperate for cash sold the shares to Great Britain for GBP 4 million. In 2002, 15000 ships passed through the Suez Canal generating US$2 billion in revenues for Egypt (*Parting the Desert: The Creation of the Suez Canal* by Zachary Karabell). In 1880, Ferdinand de Lesseps formed a company to replicate his feat in Panama. This travail was undone by weather and tropical diseases (yellow fever, cholera, malaria) that killed 22,000 labourers. In 1904, through the urging of US President Theodore Roosevelt, the US Army Corps of Engineers re-started the work, which would cost US$352 million and 5609 lives. In 1918, after four years in operation, less than 5 ships passed through the Panama Canal daily. Since 1970, more than 15,000 ships pass each day (*Path Between the Seas: The Creation of the Panama Canal* by David McCullough).

[10] In 1910, General Billy Mitchell proposed that airplanes might sink battleships by dropping bombs on them. The US Secretary of War, Newton Baker, remarked, "That idea is so damned nonsensical and impossible that I am willing to stand on the bridge of a battleship while that nitwit tries to hit it from the air." The editor of the prestigious Scientific American wrote, "to affirm that the aeroplane is going to 'revolutionize' naval warfare of the future is to be guilty of the wildest exaggeration."

In the 20th century, the 31 mile tunnel linking England to the Continent got started in 1987 fueled by GBP 5 billion from banks and 112,000 British investors. In its first 3 years since May 1994, The Chunnel saw fire and operating expenses around GBP 2 billion. In 1997, the original investors suffered losses when the banks seized shares in exchange for restructuring the crushing debt. By 2001, the winds changed, more than 16 million passengers and nearly 2.5 million tons of freight passed through the tubes (*The Chunnel: The Amazing Story of the Undersea Crossing of the English Channel* by Drew Fetherston). Perhaps, soon, this endeavour may become as profitable. Big ideas offer big dividends but have stormy beginnings. Even worse is when people cannot visualise the future because their imagination is bent out of focus by short-term ROI or resistance to change *status quo* or politics.

Creating analytical reports are essential but analysis should also bolster the need to change. Sometimes the latter is under-emphasized and may be reflective of authors who may be inclined to believe in core operating principles that 'should' remain unshaken or the continuity of the environment in which they operate (*Built to Last* by Jim Collins). In fact, the environment is under constant change and they may be unable or too late to recognize the need for change integration. This issue of "continuity" in the context of a future vision may be simply wrong (*Creative Destruction* by Richard Foster and Sarah Kaplan).

Economy is not a continuous process but rather a series of flash or strategic inflection points when the operative procedures may change suddenly and without warning (Andrew Grove in Financial Times, 20 August 2003). In this respect, economics draws from the study of evolution. Stephen Jay Gould, the famed Harvard evolutionary biologist and prolific author who influenced his field for decades, is perhaps best known for his idea of "punctuated equilibria" in which he argued that evolution consisted of relatively rapid spurts of species evolution rather than gradual, continuous transformations. A nation must continually take actions to reinvent itself to keep up with the competition and gain advantage. The economy must show resilience through a continuum of implementation, anticipation and adaptation to the future before the 'writing on the wall' becomes visible (Gary Hamel and Lisa Valikangas in *The Quest for Resilience* Harvard Business Review, September 2003). Similar ideas need to percolate within global organizations and governments, especially in view of the global uncertainty and threats to security.

It is vital to ensure systemic integration of basic education with workforce creation to stimulate economic growth and create jobs suitable for the knowledge economy's innovation society which could demand a standard of living improvement at least 2-10 fold higher than the current per capita income. Although creating policy is a preface to implementation as well as a legislative necessity for appropriations, true progress often demands 'out-of-the-box' leadership. Policy is like music, silent unless performed. Thus, it is prudent to develop and implement systemic models (based on policy suggestions) that are economically sustainable and likely to be adaptable or may lend itself to replication, if successful, for dissemination and adoption.

Systemic vision requires individuals who can visualize future issues through a dynamic and analytical convergence lens which balances the odds of probabilistic decision making in a generally uncertain world. It is such 'inclusive' thinking that is less common because institutions still offer prizes only for depth of expertise, almost exclusively. However, a few erudite individuals such as Murray Gell-Mann and Nicholas Negroponte, continue to comment on the need for individuals with 'horizontal' understanding. But, academic *status quo* and paucity of such enlightened views are discouraging individuals to pursue a broad spectrum of 'horizontal' understanding. In "Darwin's Middle Road" that Gould wrote for the Natural History magazine, Gould once said, "if genius has any common denominator, I would propose breadth of interest and the ability to construct fruitful analogies between fields."

2.0 Leadership and Leaders: Few and Far Between

Biotechnology is currently, almost, synonymous with biopharmaceuticals. In other words, establishing corporate-academic liaison to generate basic research that can be scaled for applications and production of drugs. As of 2001, funding for six key US biopharmaceuticals is estimated at US$6 billion (Exelixis, HGS, Lexicon, Millennium, Tularik and Vertex). In this "supply chain" where is the real profit? The "design" phase harbours the intellectual property and long term potential for premium jobs. Specialized technical employment (>5 fold per capita income) demands a pipeline of expertise that segues to concerns for mathematics and science foundation from pre-elementary to post-secondary education. Biotechnology initiatives may soon require new direction (nano-bio) and coordination for workforce development, competent enough to attract investment either internal or FDI (foreign direct investment).

Country	R&D personnel per million people	R&D Spending	GDP in 2002 Billions USD	Patents per million people	Population In 2015
Ireland	2,132	1.5 %	111	106	4.4 million
Israel	1,570	3.7 %	125	74	7.7 million
Singapore	2,182	1.1 %	93	8	4.8 million
UK	2,678	1.8 %	1,404	82	60 million
US	4,103	2.5 %	9,612	289	321 million
Japan	4,960	2.8 %	3,394	994	127 million

Table 1: Favourable R&D statistics?

It appears that nations are keen on the post-genomic progress despite the complexity of genomic-proteomic approach that requires cross-functional multi-disciplinary programs to act in *synchrony*. Strategists are aware that screening 10,000 potential compounds may reveal only one which may find approval as a prescription drug. The process may take about 15 years and costs on average US$1 billion to develop. In 2001, worldwide pharmaceutical R&D topped US$50 billion. Is there a potent policy to deal with the spectrum of biotechnology opportunities? Is there investment to address the far reaching potential of apomixic agri-biotech, RNAi (Alnylam), nano-biomedicine and biometrics? Following the US model may not be feasible for most nations. Perhaps one reason is the extent of basic research funding in US (for 2005, approaches US$100 billion) which trickles down to pre-college education. Taken together, the US invests more than $500 billion in research. In 2002, US Congress appropriated about $23 billion for the National Institutes of Health and US biotech companies invested more than $25 billion for R&D. In 2002, US healthcare cost about 14% of GDP or $1.3 trillion. By 2050, 25% of the US population will be 65+ years.

Can biotechnology enhance sustainable economic growth through creating new markets? It may be prudent for the few developed economies to venture beyond their solipsistic bliss and begin to weigh the saturation of Western markets, fierce competition in the far-East versus the potential of the Third World as emerging economies with increasing need and buying power. Today, number of people living on less than $2 a day is about 3 billion or nearly half of the world population. While biotech offers great promise (see below), it is equally important to remind ourselves of the 'mundane' measures that still begs attention. There are still at least 48 countries where more than half of the population cannot afford to use iodized salt. More than 40 million infants in the developing world are unprotected from iodine deficiency, which is the most common cause of preventable mental retardation in young children.

Country	GDP Growth 1975-2000	GDP Growth 1990-2000	Per capita 1973	Per capita 2002	Health Expense % of GDP	Health Expense per capita	TB reported per 100,000
Ireland	4 %	6.5 %	$ 7,023	$ 28,662	7 **	$ 1234 **	
Eq Guinea *	10.4 %	18.9 %		$ 2,087			
China	8.1 %	9.2 %	$ 1,186	$ 4,671			
Thailand	5.5 %	3.3 %	$ 1,750	$ 6,575	6.1	$ 112	
Botswana	5.1 %	2.3 %		$ 7,792	4.2	$ 219	303
Vietnam	4.8 %	6 %		$ 2,072			
Sri Lanka	3.2 %	3.9 %		$ 3,202	3.4	$ 99	
Mozambique	1.5 %	3.9 %		$ 892	5.8	$ 50	104
Ghana	0.1 %	1.8 %	$ 1,260	$ 1,946			

* Equatorial Guinea ** 1998

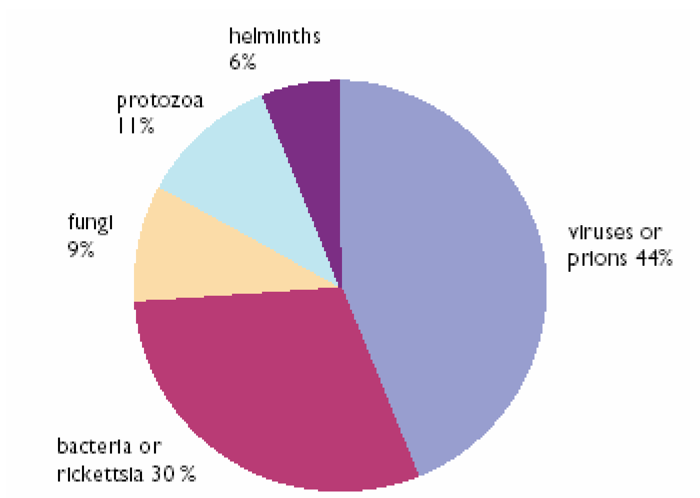
Table 2: Economic Growth Comparisons and Health (WHO)

What is perhaps alarming is the unregulated global public deception carried out in the name of biotech to promote public hysteria by groups with 'earth friendly' names. They invent data to paint Frankenstein-esque images to provoke deliberate media sensationalism and force governments to enact policies that produce death by starvation or malnutrition. For politicians to act based on reports from such advocacy groups, is akin to saying that the Pope is unbiased on contraception. Advocacy is welcome and essential to extract truth and keep us honest but it is detrimental when advocacy groups are unreasonably biased against reason.

In *The March of Unreason* by Dick Taverne, Lord Taverne provides the example of Golden Rice Project. It was demonized by environmentalists while researchers concluded that consumption of 200g of Golden Rice a day prevents blindness and measles in 14 million children under 5 years old. The scientific world hailed it as a breakthrough in genetic engineering of rice to create strains that produce additional Vitamin A in the body. 80% of the world's population needs and deserves the benefits from advances in science for society. The political illiteracy of most scientists and scientific illiteracy of politicians coupled with a commercial media allergic to science but focused on selling sound bytes enables the unbounded irrational clamour of a few to drown out the voices of reason in abject disregard for the plight of the hungry.

In the global health vein, it does not require much training to grasp that the number of people affected by cancer, diabetes, hypertension, neural dysfunction and cardiovascular disease is far less than the mortality and morbidity from malaria (parasite), TB (bacteria), meningitis (bacteria and virus), to name a few. The economic loss is simply staggering. Many of the major aetiological agents of infectious diseases identified in the past 30 years either originated or severely affected sub-Saharan Africa and are either viruses or prions (see Chart 1, below).

Chart 1: Distribution of Emerging Disease Agents (WHO)



Year Agent-Disease		Year Agent-Disease
1972 Small round structured viruses Diarrhoea		1989 Hepatitis C virus (non-A, non-B hepatitis)
1973 Rotaviruses Infantile Diarrhoea		1990 Human herpesvirus-7 Exanthema subitum
1975 Astroviruses Diarrhoea		1990 Hepatitis E virus Enteric non-A, non-B hepatitis
1975 Parvovirus B19 Aplastic haemolytic anaemia		1991 Hepatitis F virus Severe non-A, non-B hepatitis
1976 *Cryptosporidium parvum* Acute enterocolitis		1992 *Vibrio cholerae* O139:H7 epidemic cholera
1977 Ebola virus Ebola haemorrhagic fever		1992 *Bartonella henselae* bacillary angiomatosis
1977 *Legionella pneumophila* Legionnaires' disease		1993 Hanta virus pulmonary syndrome
1977 Hantaan virus Haemorrhagic fever		1993 Hepatitis G virus Non A-C hepatitis
1977 *Campylobacter spp.* Diarrhoea		1994 Sabia virus Brazilian haemorrhagic fever
1980 (HTLV-1) Adult T-cell leukaemia/myelopathy		1994 Human herpesvirus-8 Kaposi's sarcoma
1982 HTLV-2 Hairy T-cell leukaemia		1995 Hendra virus Castleman's disease
1982 *Borrelia burgdorferi* Lyme disease		1996 Prion (BSE) Meningitis, encephalitis
1983 HIV-1, HIV-2 AIDS		1997 Influenza A virus New variant Creutzfeldt-Jakob
1983 *Escherichia coli* O157:H7 Haemorrhagic colitis		1997 Transfusion-transmitted virus
1983 *Helicobacter pylori* Gastritis, gastric ulcers		1997 Enterovirus 71 Epidemic encephalitis
1988 Human herpesvirus-6 Exanthema subitum		1998 Nipah virus Meningitis, encephalitis
1989 *Ehrlichia spp*. Human ehrlichiosis		1999 Influenza A virus Influenza (Hong Kong)
		1999 West Nile-like virus Encephalitis (New York)

Table 3: Disease Agents Identified (1972-1999, WHO)

Table 4: Morbidity and Mortality Facts about Tuberculosis (TB) and Malaria

Tuberculosis	• TB kills approximately 2 million people each year • Between 2002 and 2020, nearly 1 billion people will be newly infected • 150 million people will get sick and 36 million will die of TB • Someone in the world is newly infected with TB every second • Nearly 1% of the world's population is newly infected with TB each year • One third of the world's population is currently infected with the TB bacillus • More than 8 million people become sick with TB each year • About 2 million TB cases per year occur in sub-Saharan Africa • Around 3 million TB cases per year occur in south-east Asia • Over a quarter of a million TB cases per year occur in Eastern Europe • HIV and TB is a lethal combination, each speeding the other's progress
Malaria	• Malaria kills an African child every 30 seconds • Malaria kills 1 million people each year (90% in sub-Saharan Africa) • Malaria causes more than 300 million acute illnesses • 20% of the global population is at risk of contracting malaria • Malaria cost Africa more than US$ 12 billion in annual GDP • Constitutes 10% of Africa's overall disease burden • Key challenge is drug resistance (chloroquine, sulfadoxine-pyrimethamine)

Can post-genomic biotechnology offer pharmaceutical solutions to ameliorate the affliction caused by infectious diseases? The example of Novartis in combating leprosy is a prime example. Yet, in 2001, about 20% of US global pharmaceutical R&D supported efforts for cancer drugs and an equal amount was devoted for neural disorders. Less than 10% was spent for infectious diseases, of which, the major share went for HIV/AIDS drugs.

Leadership is necessary to develop world health solutions from biotechnology and nano-biotech through genomic-proteomic emphasis. Can biotech investments create a legacy through innovation in utilizing the modern biotech tools to produce therapeutics that shall improve the plight of millions, those who are downtrodden and forgotten?

Rather than mimicking the US model, the theory of global club goods may be applied in a 'center of excellence' *modus operandi* to spawn global collaboration, for example, to create a novel designer mosquito that cannot carry *Plasmodium falciparum*, the principal malaria parasite. Map of the malaria genome and bioinformatics can be instrumental in creating drugs to treat resistant strains. A global distributed technology focus may hold the key. The profit from policies to enable such progress is of the "retail" variety, in other words, great profits are reaped not from high margins but from high volume. Innovation to produce low cost drugs for simple afflictions that impact billions are likely to result is far greater profits compared to high margin low use products marketable only in the affluent industrialized nations.

The critics and the pharmaceutical industry may point toward financial losses from the drug for river blindness, which most African countries could not afford. Recent disputes with HIV/AIDS drugs are perhaps even better known. But the economics of "lack-of-profit" based opposition from pharmaceutical behemoths may be inherently flawed when it comes to "disruptive" models stemming from agile, nimble companies using collaborative investments and taking advantage of innovation partnerships. The lesson (see below) from disruptive technologies may support this perspective (*Innovator's Dilemma* by Clayton Christensen).

Apple Computer's early entry into the hand-held computer or PDA market, later popularized by 'Palm' Pilot, offers a definitive example of large companies in small markets (from *Innovator's Dilemma* by Clayton Christensen). In 1976, Apple Computer sold 200 units of Apple I for about $600 each. In 1977, Apple II was introduced and 43,000 units were sold within two years. Based on this success, Apple Computer went public and its 1980 IPO was an industry triumph. By 1987, Apple Computer had grown into a $5 billion company and sought large chunks of annual revenue to preserve its shareholder value. In 1990, the PDA market was emerging and Apple Computers invested millions to develop the "Newton" PDA which sold 140,000 units in 1993-1994. It amounted to only 1% of Apple Computer's revenue and hardly made a dent in Apple Computer's need for new growth. While selling 43,000 units was viewed as an IPO-qualifying triumph in the smaller Apple Computer of 1979, selling 140,000 "Newtons" was viewed as a failure in the giant Apple Computer of 1994.

The principle is based on observations that small markets (profits) fails to support financial growth needs of large companies. As a consequence, the larger an organization becomes (Merck, Amgen, Abbott, Pfizer, Novartis), the weaker the argument that emerging growth opportunities (markets) can be useful engines for financial growth.

The math is simple: a $40 ($4000) million company can grow profitably at 20% with an additional $8 ($800) million revenue in the first year, followed by $9.6 ($960) million the second year. Pharmaceutical companies posting annual revenues of $10 billion, then, require $2 billion in additional revenues to grow at 20%. Disruptive models facilitate the emergence of new markets but there aren't any $2 billion emerging markets or even $800 million emerging markets!

It is precisely when emerging markets are small and the least attractive to large companies that entry into them is critical. This issue may be worth consideration and reflection for distributed technology policy and investment by small but affluent nations as well as institutions such as the World Bank, IMF and Asian Development Bank and organizations such as the Third World Academy of Sciences (TWAS) or the Inter Academy Council (IAC) that may act as catalysts to stimulate opportunities for the "south" nations.

The criterion of civic responsibility[11] is divorced from analysis of markets that don't exist. In formulating (and more importantly, implementing) policy in biotech mediated growth, nations may wish to seek the zeal to innovate to alleviate suffering and provide a moral example to be judged by history based on actions, amidst the fervour of globalization and its discontents. The 'south' nations are struggling to achieve such capabilities and deserve help.

3.0 Info

The revolution in use of information technology may have been the viral spread of the internet using the medium of the world wide web which witnessed 0 to 25 million websites in 2 decades (1980-2000). An inaudible evolution of the 25 year old web began more than a decade ago (1994) with the proposal to create the semantic web by Tim Berners-Lee, the innovator who synthesized the world wide web by creating software to allow electronic documents to link to each other. The semantic web, similarly, is a vast open source area where both start-ups and behemoths can play critical roles. The impact of the semantic web is described in specific scenarios throughout this paper.

High tech start-ups and small firms in US, launching innovative disruptive IT products, logged about $62 billion in revenues (1976-1994) but those who followed after the market was established had revenues of only $3.3 billion, a *twenty fold* difference. Investment in ICT as an economic growth area must shift from repetitive manufacturing to "intelligent" products and services. Some nations may no longer remain interested in making "boxes" for Dell but wish to further academic-industry partnerships to create disruptive products and services for new markets. But,

- are there qualified primary and secondary school teachers to foster innovative math-science education?
- how robust are the systemic endeavours to nurture the talent required for economic growth?
- are their tolerant "incubators" for "out of box" thinkers and their non-conformist blue-sky ideas?

US colleges awarded 24,405 bachelor degrees in computer science in 1996, 50% *less* than 1986 [30,963 in 1989]. Engineering graduates dropped from 66,947 in 1989 to 63,066 in 1996. With globalization of technology, 60% of all new jobs created require IT related skills but only a fifth of workers possess such skills. Utilizing ICT services as an economic engine for the ambitious youth will require different thinking where (business) systems-related virtual products and services can be offered globally.

---

[11] Here's a titillating scoop: Bill Gates has an assignation on Wednesday (24 September 2003) with Botswana prostitutes, not just one of them but a whole team. Wait, it gets kinkier: He's bringing his wife, Melinda. The encounter between the world's richest man and some of the world's poorest prostitutes is part of Gates' new passion: doing to AIDS and malaria what he did to Netscape. I'm tagging along with the couple during their tour of Mozambique, South Africa and Botswana, and frankly, one of the best things to happen to Africa is their fervor to alleviate Third World fevers. The buzz among African aid workers is that Gates will be remembered more for his work fighting disease than for Windows. Certainly the wealth of the Bill and Melinda Gates Foundation (nearly $100 billion) is improving the prospect that vaccines will be found for malaria and AIDS. The foundation's most banal work is with vaccines, but those programs have already given out vaccines that will save 300,000 lives. AIDS, malaria and tuberculosis are all worsening in the Third World and now kill 6 million people per year. This slaughter is one of the moral challenges we face today, yet Western governments have abdicated responsibility and Western medical science is uninterested in diseases that kill only poor people. Many times more money addresses erectile dysfunction than malaria. (Nicholas D. Kristof nicholas@nytimes.com for International Herald Tribune of 25 September 2003)

Services today are not about ISP or web hosting. A surfeit of ISP companies or websites do not indicate progress. Ukraine has more internet service providers that any other non-English speaking country and in Germany, for every 1000 people, there are 85 websites. Being able to buy Gucci handbags in Moscow is not an indication that the country is a market economy.

Businesses in US generate >1 terabyte of data per second. With the deployment of auto identification technologies, the US packaged consumer products industry alone may be faced with >1000 terabytes of static data storage each year. Imagine the sheer data volume for all global industries and businesses, taken together. The current internet holds about 1 billion pages with 10 petabytes of data. The internet, as we know today, will be dwarfed. The future demands radically different (nano) mechanism of data handling and software, to make sense of data (semantic web) to extract useful information and shunt intelligence to improve processes.

The economics of data-rich societies create other issues of which the problem of unemployment is key. It will affect those with basic skills (check-out clerks in retail industry, warehouse labourers). Track and trace automation technologies (for example, radio frequency identification of objects affixed with RFID tags) are likely to drastically reduce the requirement for low skill labourers. This wave of pervasive automation will further shift the yardstick of "skills" and current estimates are likely to be flawed. The ability to automatically identify objects using radio frequency id (RFID) tags must be implemented on some products by some suppliers by if they wish to supply Wal*Mart. But is that a model for the world to follow? About two-thirds of the nearly 1 million employees in 3000 Wal*Mart stores, in US, are item checkers. Can automation of object identification potentially eliminate millions of jobs? The cost savings for Wal*Mart ($8.35 billion, claimed by *Fortune* of 17 November 2003) will not be ignored by shareholders. Imagine the cumulative effect on workforce with gradual adoption of auto id technologies (AIT).

Quarter century ago, robotics created a divide between Japanese and US automobile industry that drove Chrysler to bankruptcy. Detroit witnessed riots as UAW (United Auto Workers) unions fought lay-offs and job terminations from automation. The economic cycles have shifted and major grocery store chains in California experienced a significant financial impact from the strike and lockout of 70,000 United Food and Commercial Workers Union members (14 October 2003). UFCW in California is concerned about elimination of jobs since all three major chains are also involved in pilot studies to explore future operational cost savings from use of RFID tags. This bandwagon is not fueled by systemic improvements but only short sighted *ad hoc* gains from the "low hanging" fruit.

The unfathomable volume of data will cause data handlers to declare war on the *status quo* software and pamper an irreverent group of thinkers who may innovate software to handle gargantuan tasks without traditional "lines of code" programming fraught with error (100 to 200 errors per 1000 lines). Charles Simonyi (WYSIWYG) and Mitch Kapor (Lotus-1-2-3) are pioneering these fields. Steve White is on the trail of self-healing autonomic software by exploiting the nexus of artificial intelligence and artificial biology (IBM). Nowhere more will the lack of strength in elementary and secondary education, if unattended, will become painfully obvious. Lack of academic rigour in pre-college education will erode the "supply chain" of talent and exclude the economy and youth from contributing in this field. As a consequence, businesses must start as followers once the fields are established. If the talent is available but the vision, will and infrastructure to nurture "out of the box" thought is in short supply, then the dissipation of talent is inevitable. (Finland, Ireland, Taiwan and Chile are small but potentially powerful nations worth reviewing with respect to their innovative policies for educational reform and entrepreneurial zest.)

Impact of "bad" software is fueling innovation to create programs without programming or software that "senses" tools the user needs (just-in-time programming, tangible user interfaces). In 2000, a quarter of all commercial software projects in the US were terminated due to poor quality of software, at a loss of about US$67 billion. Bad data caused estimated losses of another US$61 billion. In 2005, simple software maintenance issues cost US Department of Defense $14 billion, alone.

New models for data, databases and data movement are promising fields for intelligent services that include use of Agent based models embedded in the semantic web. The ability to innovate (small scale innovation) and extract intelligence from data is required to meet the emerging applications of sensor networks and ultrawideband (UWB) applications. Sensors may soon be transformed by nanotechnology. *Ad hoc* sensors networks communicate via multi-hop broadcast protocol (802.15.4) which can transform telecommunications, particularly where infrastructure is lacking[12]**.** With the gradual advent of IPv6, sensors may be directly plugged to the internet. Hence, sensor network data from security installations may be useful in non-obvious relationship analysis (NORA).

Bio-sensors are increasingly important in healthcare, biometrics and military. The abundance of data may make current database architectures impotent. The field requires new thinking (streaming database, adaptive data flow) to create next generation of products and services. Innovative *convergence* of behaviour-based millibots with frequency agnostic interrogators (software defined radio or SDR) for analytical data mining through the semantic web is one unambiguous example of what the future demands.

Wireless sensor networks are a prime example of pervasive computing that ranges from blood pressure sensors in arteries transmitting data to monitoring devices or sensors suggesting trends in warehouse shelf occupancy or a plethora of security and biometric applications. Sensors are self-powered and form wireless *ad hoc* networks that upload through nodes connected to internet. This will change in the next few years (2007-2008) with adoption of IPv6 which will enable individual sensors, not only nodes, to be connected to the internet or grid (fabric layer). Each sensor has certain analytical and in-network processing abilities. Such networks can transmit analyses of data (not only raw bits) in order to provide 'answers' rather than numbers.

---

[12] A Boston nonprofit founded by Nobel Laureate Bernard Lown has ambitious plans to build a nationwide, wireless computing network for Uganda's impoverished health-care system. The project will be built "on the cheap" using the country's existing cell-phone network, Palm handhelds and new battery-powered, wireless Linux servers. Satellife, a nonprofit organization specializing in medical technology, hopes the network will be based on 3,000 to 5,000 Palm handhelds given to doctors and health-care workers in the field. The handhelds will be used for routine health administration, ordering and tracking medical supplies, delivering new treatment guidelines and, of course, communication. In the field, the handhelds will connect to inexpensive, battery-powered Linux servers set up across the country. Built by WideRay, a San Francisco startup, the Jack servers have built-in GPRS radios, which afford them an always-on connection to Uganda's near-ubiquitous cell-phone network. About the size of a thick hardback book, the Jack servers act as "caching" servers, storing content sent to them over the cell network from the administration's computers in Kampala. In turn, reports and e-mail received from handhelds are relayed wirelessly back to the capital. The servers communicate with handhelds using an infrared link. The servers are powered by industrial-grade batteries and a single charge lasts up to a year. "It's all you need to provide connectivity in remote locations," said Saul Kato, WideRay's founder and CEO. "The big issue is connectivity. There are no landlines. The only real infrastructure is cellular. Some of the facilities are literally huts, but you don't need any other hardware on site." WideRay makes versions that are equipped with infrared as well as Bluetooth and Wi-Fi wireless technologies. "It's an incredible project," said Kato. "There's no other way to deploy this kind of thing. It's a moneymaker for us, but also serves a critical need. It's good every way you look at it." www.wired.com/news/wireless/0,1382,58296,00.html (01 April 2003)

Sensors are bio-inspired since living organisms are a network of sensors (smell, taste, sight, sound, touch). In the 19^{th} century, Humphrey Davy developed a miner's lamp that sensed the presence of methane in coal mines. Today we have sensors already embedded in some scenarios but nanotechnology will soon make a whole new class of ultra-sensitive sensors possible as well as feasible for pervasive deployment. Bio-sensors for detection of hidden landmines or liquid explosives have life-saving applications. But, in general, what to do with sensor data and how to derive value from sensor networks are important issues for every industry, organization and government.

4.0 Nano

The first nanotechnologists may have been the glass workers in medieval forges who may have made the Lycurgus Cup (Roman Empire) now in the British Museum. The process of nano-fabrication, in particular the making of gold nanodots or quantum dots, is old. The colour in stained glass windows found in medieval and Victorian churches depend on the fact that nano-scale properties are quite different from microscale properties. At the nanoscale, the yellow colour of macroscale gold, gives way to orange, purple, red or greenish, depending on the size of nanodots and their density per unit area (which makes the colour visible to the naked eye). Nano gold does not act like bulk gold. The different coloured gold nanodots will look yellow again if they are pushed back and allowed to join. If enough (number) nanodots are close to each other but not close enough to combine, then, we can see, with the unaided eye, different colours, red, for example, in red stained glass (glass contains 40 parts per million of gold).

1nm (1 billionth of a meter) is a magical point on the dimension scale. Nanostructures are at the confluence of the smallest human-made devices and the largest molecules of living things. Nanoscience is the study of molecules and structures roughly between 1 and 100 nm. A human hair measures 50,000 nm across and the smallest features etched on commercial microchips were 130 nm in 2002 but recent claims are below 90 nm (in 2005). By 2010, almost all the principles involved in making microprocessors (chips) must be re-thought. The industry must shift from micro- to nano-chips. CMOS chip fabrication will change. Integrated circuits (IC) depend on the science described by Ohm's Law. In the nanoscale regime, Ohm's Law may not work. The fundamentals of charge motion are then described by quantum mechanics. In the next few decades, we may witness the emergence of chemically assembled electronic nanocomputers (CAEN) that uses molecules (rotaxanes) to switch the current passing through a molecular wire from a high to low current state. This switching is the basis for an entire integrated architecture (switches placed in an array of logic gates that can actually perform computation). The basic research on CAEN work was published nearly a decade ago in *Science* by Stanley Williams from Hewlett-Packard (HP Lab). Recent advances offer more hope for transistors based on carbon nanotubes that could run many times faster than anticipated future generations of silicon-based devices but uses less power. Researchers at IBM[13] have selectively arranged carbon nanotubes on a surface to make transistors, a step toward arranging them into complex circuits. This development overcomes the practical obstacles to CAEN (difficulty of deliberately arranging the molecules so that they can form complex circuits).

Today, we are at the dawn of the pre-nano era, comparable to the pre-silicon era of 1950s when TV was black and white, small and limited, unreliable and fuzzy.

---

[13] Klinke, C. *et al* (2006) Field-Effect Transistors Assembled from Functionalized Carbon Nanotubes. Nano Letters **6** 906-910

Nano is highly convergent, multi-disciplinary yet a fundamental principle. Nano will influence every facet of science, technology and business to transform life in a manner that will have visible and economic impact ranging from use of suntan lotions to the future of space travel on an elevator (lift). For any nation, rich or poor, to under-emphasize the importance of nano may be analogous to making vacuum tubes or vinyl records, today.

Unlike information technology, where almost anyone can train to be a technician (for example, Cisco Networking Academy[14] in US high schools since 1996) and become productive in a matter of months, nanoscience research and business stemming from nanotechnology may require in-depth scientific knowledge and PhD training. Therefore, the estimates of PhD's required to become an 'innovation' society may require re-evaluation. Short sighted estimates by some nations are indicative of a failure to understand the imminent winds of change given that such reports are still being published half a century after the founding speech of nanotechnology (Richard Feynman, 1960) and even after nano made the cover of *Forbes* and *Science* magazine in 2001, ten years after the discovery of carbon nanotubes (1991), for which several Nobel Prizes have been awarded.

Nanoscience and nanotechnology require us to imagine, make, measure, use, design and innovate at the nano-scale. Because the nanoscale is unimaginably small (one nanometer is one billionth of a meter), it is clearly difficult to do the imagining! In the short term, globally, there will be a shortage of talent. It is an opportunity to lead or succumb to brain drain. Researchers in universities may be offered significant compensation if they move to private sector or start-ups. Unlike the bursting of "dot com bubble" where an asset-less dot com (eBay) may command a market value several times more than a business with goods (Sotheby's), the "nano dot" boon may offer actual goods and products. Nano skin creams and suntan lotions (Sol Gel Technologies, a start-up from Hebrew University, Israel) now co-occupy checked-in luggage with nano-coated stain resistant khakis from Levi's Dockers.

Nano is big business and priority for HP, NEC, IBM, Merck, Dow, 3M, Boeing. In US, spending on nano research in 2005 is estimated at nearly US$2 billion. The US National Science Foundation predicts that nano-related goods could be $1 trillion market by 2015. It is likely that nano may become a bigger economic force than the combined US economy from software, cosmetics, drugs and automobiles (*Nanotechnology* by Mark Ratner and Daniel Ratner).

However, a percentage of economic growth emerging from nano will remain within existing markets, for example, use of nanopore structures (Zeolite) by Mobil for directed catalysis or its use for domestic water softeners. Zeolites represent the first broad scale profitable application of nanotech (saves 400 million barrels of petroleum per annum that may be soon worth $40 billion or more) but cryptic under the petroleum or water industry. Nanostructures for nanofiltration are being pursued by Air Products and Dow Chemical. Efficient ways for drinking water purification (Generale de Eaux) is certainly good business and friendly for the environment (nano-catalysts for hydrogen economy). Nanostructures for separation will aid dialysis patients (the blood need not be pumped out/in for dialysis to occur). Toyota uses steel produced by NKK Japan that incorporates nano-particulate carbon during the rolling process, thus offering fuel efficiency through weight savings without compromising structural integrity.

---

[14] The author was responsible for germinating the first Cisco Networking Academy, which commenced at the Thurgood Marshall High School in San Francisco during his tenure (1995-1998) as the Special Assistant XV to The City and County of San Francisco and through a Mayoral appointment as a Special Assistant to The Superintendent of Schools (San Francisco Unified School District of 70,000 students). This model was then replicated nationally and later, globally, by Cisco Systems Inc of San Jose, California.

4.1 Nano Materials

This is probably the area with most aggressive reach (milk cartons to ship building) and business-wise in the short term (nano-catalysts). Examples include homogeneous nanomaterials and bio-inspired[15] synthesis of thin films:

[a] Refractive polymers – tints glass to adjust to sun light or laminates to block UV

[b] Self-healing automobile tires – polymer forms a bridge over punctured surface

[c] Corrosion prevention – lowers bridge and railroad maintenance

[d] Automobile windshields that do not get wet – ice cannot form, rain does not impair visibility

[e] Bathroom tiles and hospital sheets – self-cleaning, biocidal (Nanogate Technologies)

[f] Flexible display – digital paper (Ntera of Ireland and E-Ink, a spin-off from Xerox Palo Alto Research Center)

[g] Thin-film material – semiconductors, batteries and solar panels (photovoltaic and electrical storage devices)[15]

Heterogeneous nanostructures or nanocomposites may evolutionize construction, aerospace and ship building. The basic building block, reinforced concrete (concrete poured over a framework of metal rods), may soon find stiff cost competition from plastic being poured over framework of strong, firm, rigid carbon nanotubes, which are 60 times stronger than steel (tensile strength). A carbon nanotube that is narrower than human hair can suspend a semi-trailer (lorry or cargo truck with 18 wheels). A single nanotube could stretch from earth to the stratosphere and be able to support its own weight. The latter spurred NASA (1999) to review ideas proposed by Konstantin Tsiolkovsky (1895) a Russian visionary and Arthur Clarke in *The Fountains of Paradise* (1978). The idea is to build an elevator (lift) that will travel 60,000 miles from the earth's surface into space carrying cargo and humans. It is predicted that the 'space elevator' will lower the cost of positioning a satellite in space from $10,000 to $100 per pound. NASA provided $570,000 to Bradley Edwards to explore.

Edwards' proposal calls for a single nanotube about 75 cm wide and thinner than a piece of paper that will stretch 60,000 miles from the surface of the earth or the ocean floor (*The Space Elevator* by Bradley Edwards, 2002). A 3-day NASA conference in Santa Fe, New Mexico (September 2003) drew 60 scientists and engineers working on the concept. The outcome of the meeting: "it is plausible" (Gentry Lee, Chief Engineer, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California). The 'space elevator' apparatus would lift up to 13 tons of cargo in a week to reach the geosynchronous orbit (22,300 miles). The necessary underlying technologies exist except the carbon nanotube material (ribbon). The estimated cost to build the first space elevator may be $6 to $12 billion. Subsequent elevators may cost $2 billion. The estimated cost of building and operating the international Space Station is expected to exceed $100 billion. Children born at the turn of the 20th century may indeed take an elevator to space to celebrate their 50th marriage anniversary.

---

[15] Following mechanisms inspired by marine sponges, researchers have created thin-film materials (without harsh chemicals at room temperature) that have novel electronic and optical properties (may lead to higher power batteries & efficient solar panels).

Schwenzer, B. (2006) Kinetically Controlled Vapor-Diffusion Synthesis of Novel Nano-structured Metal Hydroxide and Phosphate Films Using No Organic Reagents. Journal of Materials Chemistry **16** 401-407

Self-healing nanocomposites will prevent Titanic-esque tragedies from a mere gash and may allow an aircraft to recover from the sort of fuel tank damage that downed Air France's Concorde flight 4590 or even the space shuttle Challenger. The future of clothing (Nano-Dry from Nano-Tex) and combat wear for the US military promises to combine nano-composites with (bio)sensors to offer a host of monitoring as well as protective features (Institute for Soldier Nanotechnology, Center for Material Science, MIT).

Can the US and EU re-invent shipbuilding with nano-composites? Cheap, disciplined labour and Taylor-esque efficiencies catapulted Hyundai Heavy Industries, Daewoo Shipbuilding and Samsung Heavy Industries to emerge as the main beneficiaries of the South Korean shipbuilding boom. The West may have relinquished leadership as traditional shipbuilders (Newcastle, UK) but nanomaterials may offer a come back strategy for the industry, if leaders are not afraid of risks that innovation demands.

Korean (and Taiwanese) shipbuilding operations are booked through 2006 (International Herald Tribune, 25 September 2003). Every 35 days three ships go out and three get started at Hyundai's Ulsan shipyard. Ships sell for about $67 million for a 300,000 ton crude oil carrier and $200 million for carriers capable of transporting liquid natural gas. In 2002, South Korea produced ships totaling 12.4 million gross tons (up 40% from 2001) with a backlog of 27.5 million gross tons (up 36% from 2001). In contrast, the numbers for Japanese shipbuilders are 11.5 and 24 million, respectively. EU shipbuilders ranked 3rd with 4 million gross tons and China is 4th with 1.6 million gross tons but is a rising competitor banking on disruptive technologies.

4.2 Nano Opto-Magneto Electronics

The confluence of nanoscience and information technology leads to the field of nano opto-magneto electronics. The present state of electronics may be nearing the end of its ability to continue to improve based on lithography. Nanotechnology offers solutions and even allows for computation in clothing, wallpaper, paint (William Butera, PhD thesis; 2003, MIT) or anywhere else. Nothing need to be without processing power and nothing need be left unlinked. Pervasive computing is within grasp and the fruits (of nanofabrication) may be within the reach of Third World nations.

The macroscale radio frequency identification (RFID, UWB) tags of today may be replaced with tiny frequency agnostic nano-tags that can identify jewelry, books, ammunition and zoological specimens. The $50 billion annual memory disk business is bound to grow with penetration of pervasive computing. Chris Murray (IBM) has shown that bits can be stored as magnetic nanodots and the size can be reduced up to the super paramagnetic limit. Using dip-pen nanolithography, it is possible to reduce individual features down to a few nanometers. If each nanodot contained one bit of information (0 or 1) and if they were spaced at distances equal to 10 times their own size, then, on a piece of paper 8.5x11 inches (US standard letter size), one could store 200,000 sets of the *Encyclopedia Britannica*. Nanotechnology also holds promise for artificial photosynthesis, thus, converting light into energy. The current annual energy consumption of US is about 100 quads or 1×10^{17} BTU (1 quad = 1×10^{15} BTU). The energy contained in the sunlight that shines each year on the 48 contiguous states is about 45,000 quads.

4.3 Nano-bio Medicine

This area lies at the confluence of nano science and biotechnology. Biotech and pharmaceuticals shall readily admit that the most difficult problems in therapeutics are bio-availability and variable efficacy. It is estimated that $65 billion of current pharmaceuticals suffer from poor bio-availability. For example, if chemotherapeutic agents can be delivered to a tumour site prior to metastasis then the prognosis of the patient's quality of life improves. Nano-medicine may have answers for improving bio-availability through advances in nano-encapsulation, magnetic nanoparticles and photodynamic therapy. Variations in efficacy are linked with understanding the genetic profile in the same way that screening identifies mutations in breast cancer or neurofibromatosis (NF). The ability of nano-particles to cross the nuclear membrane is an area that deserves attention to prevent undesirable side-effects.

However, success of this merger in terms of applications must wait for diffusion of the semantic web infrastructure (see paper). The ability to reliably detect DNA or RNA based disease agents (anthrax, TB, AIDS) or medical risks (breast cancer, Tay-Sach's disease) may offer benefits. Current techniques (micro-array, DNA sequencing) are expensive in developed nations and are prohibitive in the 3rd World. The emergence of bio-terrorism and threat to agriculture, water and consumables that cross geographic boundaries may need these technologies to detect risks at levels even lower than the microscale.

Robert Letsinger and Chad Mirkin of Northwestern University founded Nanosphere and developed the quickest and most accurate test for anthrax and heart disease. Tests for AIDS are in progress. Integrating bio-sensors with "lab-on-a-chip" technology (system-on-a-chip or SOC) from micro-array developments (such as Affymetrix) will offer POC (point of contact) tests that can perform parallel tests for several diseases at one time, instead of individuals tests for TB, AIDS, Hepatitis, for example. These integrated BSOC (bio-system-on-a-chip) systems are expected to be far cheaper and will be feasible even for the under-developed nations (with the highest disease burdens) to take the first step forward through affordable diagnostics, bio-terror detection and some forms of biometrics.

Nanotechnology holds the potential to turn this vision into reality and nations prepared to invest in the ground work may reap the collective economic benefits. Investment in nano-biotech as an economic growth engine need not be stirred by altruism. In many developed nations, there are enough medical reasons to justify serious focus on nano-scale bio-sensors for screening and diagnostics. Developments (including biometrics) will be accompanied by debates ranging from privacy, insurance and legal methods to prevent Third World immigrants or visitors from entering a nation by determining infection profiles at ports of entry (TB, malaria, AIDS) or genetic risks.

4.4 Nano in Food Industry

Nanotechnology business in the food industry is projected to grow from $410 million in 2006 to $5.8 billion by 2012, but the impact will not be even (Cientifica). Some sectors will see rapid adoption, while in other areas uptake will be limited. Food packaging (use of nano-materials) may see improvements from use of anti-microbial or self-cleaning coatings. Security may be enhanced due to advanced ability of nanosensors to detect pathogens, chemicals and contaminants at levels currently undetectable. Quality and shelf-life monitoring are likely to be far more accurate.

5.0 Energy Economy

In 2003, when General Motors (GM) demonstrated HydroGen3 at the Tokyo auto show, the $1M price tag for the minivan may have evoked nostalgia in those old enough to remember that DEC priced its desktop computer, PDP-1 at $120,000 in 1960. Michael Ramage, former Executive VP of ExxonMobil Research and Engineering commented that, "we face a chicken and egg problem that will be difficult to overcome." Bernard Bulkin, former Chief Scientist at BP echoes the 'chicken and egg problem' as the "need for a massive new hydrogen infrastructure to deliver the goods."

In other words, the pipe dream about filling stations and fueling services – the bread and butter (with jam) for the global oil merchants. Perhaps the latter may explain, in part, the tone from US National Academy of Science (NAS) and American Physical Society (APS) reports that the hydrogen economy challenges are enormous and "the transition to a hydrogen economy, if it comes at all, will not happen soon." Predictions are difficult to make, especially about the future, but the recent report (chaired by Michael Ramage of ExxonMobil) from NAS, an august society of thinkers, stopped short of saying it is impossible. The facts may be genuine but one must wonder about the quality of vision of the leadership that is bound to retard the progress toward the brilliant economic prospects and environmental benefits from use of hydrogen as a fuel.

Today, the world uses about 13 terawatts of power, of which, approximately 80% is derived from carbon dioxide emitting fossil fuel. To keep the Earth's average temperature low enough to prevent eventual sea level increases (projected to be from 8 metres to 35 metres) and sustain 3% annual economic growth, we will need between 10 and 30 terawatts of new carbon-free power by 2050.[16]

For fossil fuel enthusiasts, the Middle East spells doom not only because it fails to contain its metaphysical zeal but also because it cannot sustain the global demand for oil even if peace was offered a fighting chance in between its many wars. A decade ago economists were confident that demand for oil was stagnating at 70 million barrels per day (bpd). Our current global consumption of about 85 million bpd (US at 21 million bpd) vastly exceeds such estimates. It is now clear that global growth, especially in India and China, will push demand for oil over the 120 million bpd mark by 2030, according to the International Energy Agency (IEA). For such demand to be met, one assumes "boundless Middle East oil" output to grow by more than 30 million bpd. Saudi Arabia's comforting "trust me" statements about its oil reserves (jumped to 280 billion barrels in 1988 from 110 billion barrels in 1978) and production capacity (expects to sustain 10-15 million bpd for at least another half century) are highly suspect since these numbers have never been subject to any third party audit or any report on how the reserves stack up on a field-by-field basis. In 2000, 90% of Saudi Aramco's oil was produced from 6 fields. The three most important fields producing 80% of Saudi oil (total production is about 10 million bpd) were discovered in 1940, 1948 and 1951 (Matthew Simmons). In the next 25 years, additional supply of oil may grow at a rate less than the estimated increase in demand (20 million bpd). In other words, oil output of about 100 million bpd may result in oil shortages of about 20% or more by 2030.

---

[16] Jim Hansen, Goddard Institute for Space Studies, NASA in MIT Technology Review (07-08/2006)

Discovery teams in search for new resources are accelerating their missions but success is rather scarce and expensive. A bit of hope comes from the oil sands (tar sands) of Fort McMurray (Alberta) that cover an area of 58,000 square miles. It is expected to yield 174 billion barrels of oil (active through 2025). At the current rate of consumption of oil by the US (21 million barrels per day of which 60% is imported and represents 25% of global demand), the Canadian reserves will be depleted in about 22 years if Canada chooses to sell every drop of oil from tar sands to the USA but none to China.

The discovery of "Tahiti" deep sea oil field by Chevron in the Gulf of Mexico may add about 500 million barrels[17] or a 3 week supply for the US. The Minerals Management Service of the US Government estimates about 44 billion barrels of oil (~5 year supply for US) remain to be discovered in the Gulf of Mexico.

Emergency oil stores are most certainly inconsequential. Rokkasho, an oil-storage site in Honshu holds 30 million barrels of oil. It is barely enough to supply one week's worth of Japanese demand for oil. Rokkasho is also dominated by giant wind-turbines but Rokkasho's fresh sea breeze can operate the turbines 20% of the time and may not produce enough power to make up for the energy consumed in their construction. Japan was keen to invest in yet another hydrogen venture – that of nuclear fusion reactor – in Rokkasho. However, the project (ITER) is now slated to be housed in Grenoble, France. The industrial world is vulnerable to oil supply disruptions in the Middle East. The economic progress in developing countries will be retarded by energy crisis and it will jeopardize the UN Millennium Development Goals (MDG) to reduce global poverty by half by 2015. The promise of oil from the Middle East and UN's plight to reduce poverty may be utopian dreams unless the vision of hydrogen as a fuel is a part of the answer to global demand for energy.

A report published in 2001 by Sir David King, Chief Scientific Adviser to British PM Tony Blair advocated a fast track for fusion development. Fusion "should not be a place to play short-sighted international politics" says veteran nuclear researcher Tomabechi of Japan and expects "fusion-generated electricity will be sent to the grid within the next 35 years" but admits it will be expensive.

One proposed fusion reactor uses Hydrogen and Boron-11 as fuel. Hydrogen is obtained by electrolysis of water and Boron deposits are plentiful (140 million tons in California, 500 million tons in Turkey). A 100 mega-watt plant would burn 200 grams of Boron a day, as opposed to 700 tons of coal[18] to power a similarly sized coal-burning

---

[17] Chevron claims that it could cost $3.5 billion to develop the "Tahiti" oil field that may yield 500 million barrels of crude petroleum. The exploratory vessel, *Deep Seas*, leased from Transocean, costs $250,000 per day (including team of 170 is $500,000 per day).

[18] The annual amount of carbon dioxide released from burning fossil fuel is projected to increase from 24 billion metric tons in 2002 to 33 billion metric tons in less than 10 years (by 2015). Coal presents the world's single largest opportunity for carbon dioxide mitigation. Coal generates 37% of the global fossil-fuel related emissions, in second place, after oil (42% from oil). In the US, coal contributes 51% of electricity but 81% of carbon dioxide related to power generation. Overall, 40% of global electricity is generated from coal-burning plants that spews twice as much as carbon compared to natural gas (kilowatt for kilowatt). World Coal Institute estimates there are 164 years worth of coal in the ground compared to 40 years store of oil (through 2050). According to the Natural Resources Defense Council, the enthusiastic coal mining countries are US, India and China. More than fourteen hundred 500-megawatt coal powered plants are planned worldwide by 2020, of which 140 are in the US. (Source: MIT Tech Review 07-08/2006)

plant. Better yet, this type of fusion reactor emits no radiation. In addition, because the reactor is safe and clean, it is possible to build small neighborhood power plants or even have a portable domestic fusion reactor right in your back yard, eliminating wasteful long distance electricity transport.

The investment necessary to transform this vision into reality is not quite there, even though evidence suggests billions of dollars worth of government investment in hydrogen fuel projects in US, EU and Japan. A bit of this was spent on perfecting procedures to break down natural gas into hydrogen and carbon dioxide with substantial wastage as heat (~15% of energy). According to Pete Devlin of the US DoE (Department of Energy), it costs $5 to produce the amount of hydrogen that releases as much energy as a gallon of gasoline ($2 in 2004 Q2 with crude over $50 a barrell in 2005 and now creeping toward the ominous $100 per barrel with escalating Middle East violence, July 2006). It is increasingly attractive to think about dumpster size conversion equipment costing about $375,000 (2004 Q3) according to Sandy Thomas of Alexandria (VA, USA) based H2Gen Innovations, manufacturers of hydrogen generators.

Dr Joseph Romm, former Acting Assistant Secretary for Renewable Energy at US Department of Energy (DoE) may have sounded like the doomsday prophet when he told the US Congress (House Science Committee) in March 2004 that, "if we fail to do so because we have bought into the hype about hydrogen's near-term prospects – we will be making an unforgivable national blunder." Perhaps Dr Romm had the best intentions of promoting hybrids or alternative fuels such as biomass usage (carbon sequestration) and ethanol. Unfortunately, we have not made any significant strides with ethanol (US annual production is just about 2 billion gallons, mainly from corn) and the same holds for biodiesel and hybrids. JoAnn Miliken, who currently heads the hydrogen storage research for US DoE agrees that hybrids, "can't solve the problem."

The gobal hydrogen endeavour is lacking direction and leadership. It is necessary to articulate an unambiguous goal unencumbered by the imminent geo-political ramifications that surround any such profound economic change. Much to the chagrin of the wealthiest industrial nations and energy industry behemoths, hydrogen may be that elusive bridge between the 'haves and have nots' which was once conceived to be built bit by bit with information technology. The industrial revolution and the information age provided some incremental quality of life benefits to the developing world but the energy economy increasingly yet silently drains resources and dampens productivity gains. The energy genie is still in the bottle and it has hydrogen written all over it.

5.1 Electrolysis

A simple process that can catapult the hydrogen economy to the forefront of global progress comes from the same man who invented the electric motor, nearly two centuries ago. Michael Faraday, born on 22 September 1791 in Newington, Surrey (England), invented the dynamo in 1831 which led to the invention of the electric motor. In 1832, he started work on electrochemistry that led to the discovery of the principle of electrolysis. He lived to see the first isolation of elemental Lithium by electrolysis by his mentor, Humphrey Davy and later in 1855 by Bunsen and Mattiessen. Michael Faraday died on 25 August 1867 in Hampton Court and is buried in the Highgate Cemetery in London. What we shouldn't bury is the idea of generating hydrogen from electrolysis of water, in every garage!

It may have intrigued Faraday to see his discovery of electrolysis being practiced safely and successfully since 1875 to remove unwanted hair. Charles Eugene Michel (1833-1913), an ophthalmologist in St. Louis, Missouri (USA), helped pioneer a technique for removing wild eyelash hairs (cilia) by means of electrolysis. It is the only process that is universally and medically approved, documented and accepted by the US Food and Drug Administration, as a means for permanent hair removal. It is perhaps time to add another universally accepted process to the credit of electrolysis[19] – portable hydrogen generation from water to replenish solid hydrogen storage in automobiles.

Electrolysis of water in your garage to generate hydrogen to replenish hydrogen storage tank in your automobile ("gas tank") could eliminate the trillion dollar investment necessary to re-tool the infrastructure for generation and delivery of hydrogen fuel. If electrolysis is successful and portable hydrogen generators are in every garage, there won't be any "oil industry" behemoths to reap the financial gains from that trillion dollar pool. These behemoths exercise global clout and exert pressure on organizations (governments) that plan and promote the pathways to hydrogen economy. Few, if any, nations of the world have leaders who are willing to lead at the expense of their personal popularity at home and abroad. John F Kennedy, threatened by the Sputnik progress of the then USSR, challenged the nation (USA) to put a 'man on the moon' within a few years. It did happen, soon enough. We lack leaders who can articulate such unambiguous universal agenda to challenge the global research community to concentrate their focus on developing efficient commercial portable electrolyzer compatible with solid hydrogen storage systems for automobiles. Such a vision will ruffle some feathers, especially raising the ire of the petroleum industry and OPEC. If automobile usage, alone, could become independent of fossil fuel, imagine the impact on the global economy caused by decreasing oil prices and the potential economic boom for the developing nations.

It is often said that we humans are our own worst enemies. The economic revolution possible through the use of hydrogen fuel for automobiles harbours the potential to trigger resistance from the uneducated and uninformed about the source of energy for electrolysis. Renewable energy (solar, wind) systems may supplement a part of the energy required for electrolysis but for commercial uses (car rental agencies, fleet operators) the use of nuclear energy (nuclear fusion?) to run the electrolyzers will be inevitable, at least in the near future. Some nations have politically banned the use of nuclear (fission) energy.

Chernobyl has been sensationalized by the media yet the incident-free accidental nuclear (fission) plant meltdown at Three Mile Island (USA) is virtually unknown. Chernobyl did not use the recommended safety precautions but the operation at Three Mile Island nuclear power plant was monitored according to the highest safety standards. Safe nuclear (fission) energy has been a reality for several decades yet the ignorance of the public is still cultivated only to worsen global warming (*The March of Unreason* by Dick Taverne).

UK obtains one fifth of its electricity from nuclear plants and all but one (Sizewell, UK) will be decommissioned by 2023. Sir David King and Mr Tony Blair may no longer stand up for the "green" mirage with North Sea oil reserves declining and greenhouse gas emissions on the rise. Even politicians know that politics may temporarily derail long term vision but long term needs can annihilate political frameworks and fail to protect disingenuous individuals from social wrath.

---

[19] Richard Bourgeois and his team at GE Global Research (Niskayuna, NY) has built a low-cost mass-manufacturable electrolyzer using a GE plastic (Noryl) resistant to corrosive electrolytes. The prototype produces hydrogen via electrolysis for $3 per kilogram (functionally equivalent to a gallon of gasoline) which is substantially lower than current estimate of $8 per kilogram.

Cheaper and safer nuclear fusion energy is on its way, too, as discussed earlier and perhaps sooner than expected (*Colliding Beam Fusion Reactor* by N. Rostoker, M. Binderbauer, H. Monkhorst in Science (1997) 278 1419). Is there a national leader who can promote science literacy to enable the public to evaluate the risk to reward ratio of safe use of nuclear energy (fission and fusion) as one mechanism to declare independence from fossil fuel? The simplest way to usher in the hydrogen economy is not a 'chicken and egg' problem, it is a political problem.

5.2 Scientific Hurdles

Simple ideas are often complex problems and this is not an exception. Several questions need answers through focused research. By using energy (solar, wind, nuclear, hydroelectric, geothermal) we can convert water to yield hydrogen in an electrolyzer and reverse the process in a fuel cell to obtain electrical energy from hydrogen. Energy required to produce hydrogen by electrolysis is 32.9 kWhr/kg. For 1 mole (2g) of hydrogen the energy is about 0.0660 kWhr/mole. For commercial electrolysis systems that operate at 1 A/cm^2, 1.75 volts is required, which translates to 46.8 kWhr/kg and an energy efficiency of 70%. Lowering the voltage for electrolysis, will increase the energy efficiency of the process and is one important area for research. R. P. Viswanath and his team at the Indian Institute of Technology in Madras, claims to split water into hydrogen and oxygen at a lower potential (0.9V) by using a compartmentalized electrolytic cell. Current efficiency works out to 135%, a key advance, if reproducible.

In an ideal case, fuel energy from hydrogen is converted to electrical energy at an efficiency of 80% or more. This is greater than the ideal efficiency of a generating facility which burns hydrogen and uses heat to power generator. Fuel cells currently may not approach >80% efficiency but are still more efficient than electric power plants which burns a fuel. In comparing the fuel cell process to its reverse reaction (electrolysis of water) it is useful to treat the enthalpy change as the overall energy change. Gibbs free energy is necessary to drive the reaction. In electrolysis and fuel cell pair, 237 kJ energy is required to drive electrolysis and the heat from the environment will contribute 48.7 kJ. In the fuel cell, 237 kJ is regained as electrical energy (48.7 kJ escapes as heat but part may be recaptured for use). The benefits from catalysis and search for nano-materials (nano-catalysts with greater surface to volume ratio) that can improve efficiencies in the electrolysis/fuel cell paradigm are key areas for research.

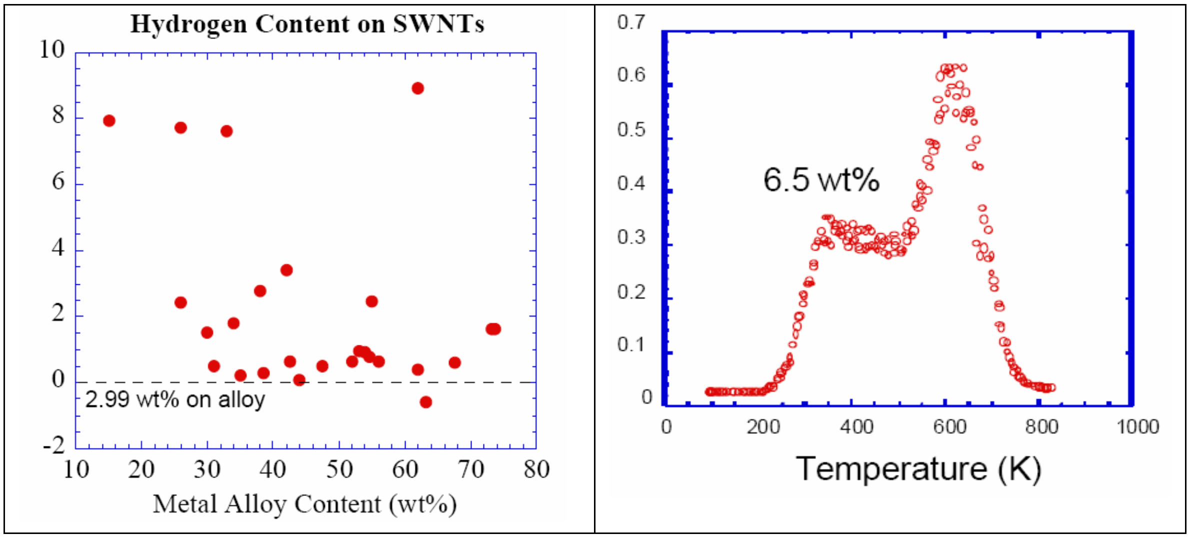
Can we use bio-catalysts for hydrogen generation? John Peters and Lance Seefeldt of Utah described the structure of an enzyme found in the soil microorganism *Clostridium pasteurianum*. CpI is a hydrogenase, which uses iron atoms to catalyze hydrogen (waste) production from protons and electrons. The enzyme active site, tethered to a substrate, may improve hydrogen production efficiencies. Non-platinum catalysts are also of increasing importance (G.W. Huber, J.W. Shabaker and J.A. Dumesic. 2003. Raney Ni-Sn Catalyst for H_2 Production from Biomass-derived Hydrocarbons. *Science* **300** 2075–2077).

Safe, non-toxic, cost-effective on-board solid hydrogen storage solutions are emerging from traditional materials (sodium borohydride) as well as nano-materials (nanotubes). Catalysts (doped nanotubes) that can improve storage capacities above the US DoE recommended 6.5% (storage by weight) will be a boon to this fuel system. Solid sodium borohydride currently can store 10.5% hydrogen[20] by weight.

---

[20] One kilogram of solid $NaBH_4$ reacted with 950gm of water yields 213.5g of hydrogen gas or 24,230 BTU = 7.1 kWh = 25,560 kJ (1 kg of hydrogen = 119,600 kJ which is a close equivalent to 1 gallon of gasoline = 121,300 kJ).

Carbon single-wall nanotubes (SWNT) and other nano-structure materials exhibit hydrogen storage capacities at room temperatures. Capacity for adsorption of hydrogen by SWNT is about 8% by weight. A method for growing SWNTs by vapor deposition from methane holds promise when scaled-up, to produce SWNTs for $1 per kilogram.



HYDROGEN STORAGE IN CARBON SINGLE-WALL NANOTUBES

Proceedings of the 2002 US DoE Hydrogen Program Review

A.C. Dillon, K.E.H. Gilbert, P.A. Parilla, J.L. Alleman, G.L. Hornyak, K.M. Jones, and M.J. Heben

5.3 The Road Ahead

Standardization of form, for fuel cells, and mechanism to replenish with hydrogen from portable electrolyzers may be as simple as inserting a tube to feed hydrogen to adsorbent in the on-board storage tank or solid material that can be stored in the hollow frame of an automobile. When the domestic electrolyzer is not available (vacation to distant location), electrolyzer outlets may be ubiquitous in stores, grocery chains, restaurants or charged fuel cells can be exchanged just as cooking gas is available in tanks in most corner stores (in Europe). GPS-RFID-UWB linked hydrogen sensor tagged fuel cells can help track and trace customer issues, inventory, billing operations, as well as monitor fill-status and safety considerations.

Perhaps this discussion about hydrogen as a renewable fuel for the near future focuses disproportionately on fuel consumption as a function of automobiles and transportation. In most industrialized nations, energy consumption in this sector is about 20-25% of total energy. Successful implementation of hydrogen as a fuel reduces need for fossil fuel by the same amount. However, residential and commercial use is another 20-25% while industrial uses claim almost half of the energy output. Fusion reactors are a definitive answer but are still a few decades away. The 'south' nations cannot continue to wade through squalor while scientists develop ways for mass deployment of fusion reactors somewhere between the years 2025-2050. One doesn't need the brilliance of Jeffrey Sachs to grasp that energy is the key that may change the tide for sustainable growth in 'south' nations and lend credibility to Mary Robinson's penchant to pursue ethical globalization. However, politically unpalatable it may be, the 'fast food' model of energy is needed.

There may be a "STAR" solution, albeit, a partial solution. Sealed, transportable, autonomous reactors (STAR) can generate power from nuclear fission without refueling or maintenance. It is a choice that is available now and those who will keep it out of the hands of those who need it the most may wish to consider the odds. By the time you finish reading the rest of this article, several children will have perished from malaria. In Africa, a child dies every 30 seconds from malaria, alone. Improvements in sewer and sanitation can eliminate this morbid mortality statistic. Lack of energy is the gatekeeper in the 'south' nations who are reliant on the proactive benevolence of the 'north' even for their very basic survival. Disposable nuclear fission reactors may be an interim answer to this agony.

STAR can meet the immediate energy demands of under-developed and developing countries without the risk of malicious use by disenfranchised individuals to use the by-products for weaponry. Taking a lesson from modern submarines, STAR was developed by the Lawrence Livermore National Laboratory in California, to be transported to a site and generate power for decades (current estimate is 30 years). When the fuel is spent, it can be retrieved from the site and disposed under proper supervision since there is no option in a sealed STAR for recharging the fuel rods depleted of fissile isotopes (usually commercial operators replace fuel rods every few years). The latter coupled with a tamper-proof casket eliminates the risk of extracting fissile material. An even better reason to use STAR in 'south' nations is the ability to produce versions capable of generating 100 or 10 megawatts of electricity compared to the conventional nuclear stations that produce about one gigawatt. Without an extensive electricity grid, the output from a conventional nuclear station is wasted or unused due to lack of distribution infrastructure. STAR units producing 100 megawatts may be 15 metres tall, 3 meters in diameter and weigh 500 metric tons. A lighter, 200 metric ton version may produce 10 megawatts of electricity. Nuclear fuel, liquid lead coolant and a steam generator is sealed in the housing along with steam pipes ready to hook up to an external generator turbine.

In the 5th century BC, Herodotus noted in his *History* that every Babylonian was an amateur physician, because the sick were laid out in the street so that any passerby might offer advice for a cure. Nearly 2500 years later, we are all environmentalists offering advice, apostles of platitude without power, uniformly impotent, to cure the energy sickness. Yet the 'penicillin' from the mold in Dr. Florey's coat was gifted as the cure for the energy-addicted world, over half a century ago by Lise Meitner, Otto Hahn and Frederic Joliot-Curie, among other notables.

5.4 One Shoe Fits All ?

Safe use of nuclear fission for energy *cannot* become the 'one shoe fits all' solution for the impoverished nations to climb out of their poverty capsules. It is a solution at hand and one that will empower the 'south' nations to see some light at the end of the tunnel. It is important that national policies commit to concomitant exploration of other renewable energy sources that can stem the tide of global warming, even if the validity and reliability of such warming trends are often steeped in scientific controversy and subject to incessant political spin in the media.

Encouraging advances include one by UK-based Intelligent Energy and its product – a motorcycle - that runs on hydrogen fuel cells, attains speeds of upto 50 mph and travels for 100 miles before refueling is necessary. Tokyo Gas launched (2005) a residential fuel cell project where a home-owner can lease an unit that extracts hydrogen from natural gas and uses it to generate enough electricity to meet about 60% of the demand for a four-person household. Each unit may reduce a home's annual greenhouse gas emissions by 40%. A 10 year lease costs JPY 1 million (<$10,000) but the savings from reduced energy usage, today, may not cover the cost of the lease. The annual shortfall is estimated to be about JPY 40,000 (<$400) per home (MIT Technology Review, March 2005).

Developments from super-conductivity research are helping to produce better fuel cells. New thin film solid oxide fuel cells (SOFC) offers catalyst-independent operating temperatures of less than 500^0 C. At less than 1 micron thin and an output of about 1 volt, a stack of SOFC equivalent to two soda cans may produce more than 5 kilowatts (enough to power one or more typical households). Connected to a homeowner's natural gas line, this stack operates at an efficiency of about 65%, a two-fold increase in efficiency over conventional power plants.

Plankton fuel cells, energy from spinach, biodiesel, carbon sequestration (Craig Venter Institute) and natural forces (air, water, solar) to generate energy are all likely to be more or less viable in specific use cases and environments. The latter is demonstrated in part by Costa Rica which claims to derive 92% of its energy from renewable sources. It is vital to pursue these and other emerging 'green' sources of energy while we continue to boldly support options available at hand to immediately provide energy for the emerging economies that are politically responsible. The sooner the impoverished countries are economically mature, the sooner they can contribute to invest in the global plight for alternative 'green' energy to reduce fossil fuel dependency and reduce the burden of carbon emissions.

5.5 Ethanol Economy

With 40% of operating automobiles **not** running on petroleum, Brazil has demonstrated that it is the global leader in the use of ethanol without government subsidies. Alcool (ethanol) as an alternate fuel for cars, buses and other motor vehicles is in use because 75% of all vehicles sold in Brazil are flex fuel vehicles that can run either on ethanol or gasoline or a mixture. Commencing with ideas and idealism that sprouted during 1970-1975, Brazil now manufactures 4 billion gallons of ethanol per year. Brazil is ***manufacturing*** energy from sugar (sugarcane) and pays for plant operation using energy obtained from burning the fiber from sugarcane (biomass). This is a remarkable paradigm shift because till recently energy was traditionally associated with discovery and mining (oil, coal, natural gas). While "sugar farmers" in the EU are sparring over the size of "hand outs" (subsidies), sugar is effectively used as an energy cash-crop in a novel entrepreneurial zeal just a few thousand miles, south.

The necessary detour to use ethanol as an alternate bio-fuel remains low on the strategic agenda for many nations despite the projections that gallon for gallon ethanol is competitive with gasoline in terms of fuel efficiency even if crude prices drop to $40 per barrel. One wonders why national energy policy wonks are yet to grasp that production of ethanol can be tuned by manufacturing at a cost comparable to or lower than that of gasoline. Agriculture and farming can reinvent its financial lustre from corn, sugarcane, sorghum and the oil-weed, *Jatropha curcas*, to name a few common agricultural raw materials. Although ethanol does not significantly reduce carbon emissions, if viewed in the narrow sense of emissions alone, it adds little to the *total* carbon in the atmosphere. The carbon dioxide given off while burning a gallon of ethanol is roughly equal to the amount absorbed by the plants to produce the next gallon. Thus, ethanol may be a true beacon of hope since it may help to wean away the war on peace by reducing the dependency on Middle East oil reserves.

Commercially, ethanol is far more lucrative as a cash-crop due to the fact that small, industrially advanced and truly innovative knowledge economies (nations) may now manufacture energy (ethanol) more than its requires. Hence, these countries may sell ethanol plus its associated knowledge services to the fuel-guzzling Dragon (China) and the Elephant (India). It is as simple as manufacturing and selling Taytos or tikka masala or wonton soup!

The commercial appeal cryptic in ethanol may be further stimulated by the understanding that ethanol can be obtained from *any* cellulose source because basic sugars are the key building blocks of cellulose. In practical terms, cellulose is present in all plant materials – wheat and rice straw, switchgrass, paper pulp, agricultural waste, leaves. This fact doubles the potential to squeeze twice as much as fuel from the same unit area of land.

The convergence of energy scientists with biologists are likely to unveil new vistas only limited by our imagination. Biologically, converting cellulose to ethanol is a two-step process. First, the long chains of cellulose must be broken down to basic units (glucose, fructose or other sugars) and second, fermenting those sugars into ethanol. In nature, fungi and bacteria secrete enzymes (cellulase) that hydrolyzes cellulose to "free" the sugars. Yeasts ferment the sugars to produce alcohol. With the tools available from biochemistry, molecular genetics, recombinant DNA and bio-engineering, it is possible to improve the efficacy of the microorganisms for production of cellulosic ethanol. Genetic engineering to yield strains of yeast (*Saccharomyces cerevisiae*) that can tolerate higher concentrations of ethanol[21] in a fermentation reactor and can survive on cellulose alone[22] may be a "disruptive technology" in the ethanol energy economy. The strides made by genomics research makes it possible to engineer an existing small microorganism[23] with an artificial chromosome to harbour genes for the enzymes necessary to direct a high yield manufacturing process to produce ethanol from cellulose.

Manufacturing energy may lend itself to the practices of near-shoring, off-shoring and outsourcing. For example, recently Singapore leased an island for 999 years from Indonesia to set up a processing facility. Leasing a few islands from Indonesia may not be an absurd idea given that it is the world's largest archipelago with 13,667 islands nestled between Asia and Australia, spanning 3200 miles along the Equator from east to west (almost the expanse of US) and 1100 miles from north to south. In addition to an abundance of plant materials (cellulose), these tropical islands are suitable for growing sugarcane or sorghum. The entrepreneurial and industrialised nations of Finland, Ireland and Taiwan, for example, may off-shore ethanol manufacturing in Indonesia (or partner with the State of Bihar, the arid and poorest state in India) in the true spirit of confluence of globalization and innovation. The potential for significant profits from such investments in bio-energy may be only limited by politics and/or the inefficiencies that often plague the supply chain of information arbitrage.

---

[21] Greg Stephanopoulos of MIT has developed a yeast strain that claims to tolerate 50% more ethanol.

[22] Lee Lund of Mascoma, a start-up in Cambridge, Massachusetts, has engineered a thermophilic bacteria to optimise the kinetics of cellulase and whose only fermentation product is ethanol.

[23] Synthetic Genomics, a start-up in Rockville, Maryland, founded by Craig Venter, is exploring *Mycoplasma genitalium*, a microbe which dwells in the human urinary tract and has the smallest genome (517 genes) of known life form (except viruses), to produce task-specific genetic pathways (for example, the two steps or tasks necessary to breakdown cellulose to produce ethanol) in much the same way that software is loaded on to a computer's operating system. Instructions from the software could be used to create spread sheets or word processing. Similarly, the "biological software" introduced in the genome of *Mycoplasma genitalium* would instruct the microbe (the cell) to break down cellulose to produce ethanol.

5.6 Ethanol Value Chain[24]: Hypothetical Example of Supply (Ireland) versus Real Demand (India)

	Sorghum bicolor (Sweet Sorghum)	*Saccharum officinarum* (Sugarcane)	Corn	Annual Demand India (2004)	Ireland
Growth Cycle	4-5 months	12-16 months			
Crops / year	2	1			
Water / crop	4,000 m^3	36,000 m^3			
Biomass	70 tons / hectare	90 tons / hectare (TPH)	1.4 – 6.5 TPH		
Ethanol	40 litres / ton	70 litres / ton (LPT)	400 LPT		
Yield / crop	2800 litres / hectare	6300 litres / hectare (LPH)	560-2600 LPH		
Annual Yield	5600 litres / hectare	6300 litres / hectare (LPH)	560-2600 LPH		
Production Cost	USD 0.30 / litre	USD 0.29 / litre	USD 0.37 / litre		
Cost / US Gallon	$1.14 per 3.785 L	$1.10 per 3.785 L	$1.40 per 3.785L		
90EBG*				1 billion litres	
95EBD**				3 billion litres	
Other				1 billion litres	
Production				2 billion litres	
Import Potential				3 billion litres	
3 billion Litres	1.1 million hectare	0.83 million hectare			
Land (Hectares)					6.9 million
Total Agriculture					4.4 million
• 80% Grass					3.5 million
• 11% Grazing					0.5 million
• 9% Crops					0.4 million
Total Farmers					130,000
Farmers 55+					52,000
Weekly Disposable					€149 Income-rural
Weekly Disposable					€160 Income-farming
Weekly Disposable					€195 Income-urban
Earnings Potential Per hectare/year					€1300 $0.60 / litre

90EBG* = 90% gasoline plus 10% ethanol and 95EBD** = 95% diesel plus 5% ethanol

[24] If my theory of relativity is proven successful, Germany will claim me as a German and France will declare that I am a citizen of the world. Should my theory prove untrue, France will say that I am a German and Germany will declare that I am a Jew. (ALBERT EINSTEIN, Address at the Sorbonne, 1936)

APPENDIX 5: AGENTS - *Where Artificial Intelligence meets Natural Stupidity*

1.0 Are Agents Necessary?

Linearisation of real world conditions are often necessary in order to "fit" mathematical models. This approach is not conducive for scenarios that demand real-time adaptability. Businesses experience non-linearity. An example of that is the Bullwhip effect which represents wide fluctuations in the supply chain. The discrete, dynamic and distributed nature of data and applications require that supply chain solutions not merely respond to requests for information but intelligently anticipate, adapt and actively support users. Agents can support clearly discernible tasks or process, interact with each other in a specified environment (for example, inventory management), interact with other Agents directly or via a message bus, continuously harness real-time data (RFID, UWB, GPS, sensors, actuators) and share this data with other Agents to offer real-time adaptability supply-demand value chains or networks.

Computer-based modeling has largely used system dynamics based on ordinary differential equations. However, a multitude of industrial and businesses, including supply chain management, are struggling to respond in real-time. Eventually this transition may emerge as real-time adaptable business network. This paradigm shift will make it imperative to model software based on Agents and equations. The question is no longer whether to select one or the other approach but to establish a mix of both and develop criteria for selecting one or other approach, that can offer real-time solutions. The "balance" is itself subject to change. For experts in supply chain management, the situation is analogous to "push-pull" strategy where the push-pull boundary may continuously shift with changing demand.

These potential benefits are at the core of the excitement to explore Agent[25] based systems in decision support. Real-time adaptability may affect a vast array of static or pre-set business processes. It is likely that many processes must change in the evolution toward the adaptable business network. In particular, real-time adaptability may revolutionize supply chain management, fostering supply chain innovation through deployment of Multi-Agent Systems (MAS).

---

[25] The idea of Agent originated with John McCarthy in the 1950's at the Massachusetts Institute of Technology (MIT). The term "Agent" was coined by Oliver Selfridge (circa 1959). McCarthy and Selfridge were colleagues at MIT along with AI expert, Marvin Minsky. The recent trends in Agent research began in 1977 with contributions by scholars including Nicholas Negroponte and Patti Maes of MIT, Katia Sycara of Carnegie-Mellon University and H. Van Dyke Parunak of the University of Michigan at Ann Arbor, to name a few. This appendix on Agents is not an original work by the author in any form, whatsoever. It is only an attempt to discuss Agents in "simple English" for appreciation by non-AI experts. However, failure to represent the ideas with clarity is entirely due to the inadequacy of the author. The original work in this appendix is collected from various papers by number of experts at MIT, Carnegie-Mellon University (CMU) and University of Michigan, Ann Arbor. Papers by H. van Dyke Parunak deserves special mention and are lavishly used in this discussion. In particular, the content from the following reference is often quoted in this appendix: H. Van Dyke Parunak (1997) Engineering Principles from Natural Multi-Agent Systems in Annals of Operations Research **75** 69-101

Commercial supply chain software (i2, SAP, Oracle) defines process in the traditional terms of rates and flows (consumption, production). System variables (cost, rebates, transportation time, out-of-stock) evaluate or integrate sets of algebraic equations (ODE, ordinary differential equations or PDE, partial differential equations) relating these variables to optimise for best results (best price, shortest lead time, minimal inventory). The process based on equation-based modeling (EBM) assumes that these parameters are linear in nature and relevant data are available. In the real world, events are non-linear, actions are discrete and information about data is distributed.

Uncertainty often influences profitability and static optimisation fails to add value, hence increaseing the demand for adaptability. The "sense and respond" type solutions will demand Agents-based software that functions continuously and autonomously without human intervention to "understand" the process environment in order to respond appropriately or alert human operators. The latter is the grand outcome expected of Multi-Agent Systems[26] in intelligent decision support.

Agent-based models (ABM) draws clues from natural behavior of biological communities of ants, wasps, termites, birds, fishes and wolves, to name a few. The "natural" use of continuity and autonomy indicates that Agents are able to execute processes or carry out activities in a flexible, intelligent manner that is adaptive and responsive to changes in the environment (without requiring constant human guidance, intervention or top-down control from a system operator). An Agent that functions continuously in an environment over a period of time may "learn" from experience (patterns). Agents that inhabit an environment with other Agents (Multi-Agent Systems) are expected to communicate, cooperate and possess mobility between environments. The mobile, networked, autonomous, self-learning, adaptive Agent may have radically different principles compared to those that were developed for monolithic systems.

---

[26] H. Van Dyke Parunak (1997). Engineering Principles from Natural Multi-Agent Systems. Annals of Operations Research **75** 69-101 Although it still remains a paradox, it is undeniable that simple individual behaviors of bugs like ants and wasps, ***collectively***, may offer intelligent models of complicated overall behavior. In fact, this may have been known for centuries. One ancient observer, King Solomon, knew from his father, David, of the elaborate court organizations of oriental kings and preparations needed for military campaigns. He marveled that insects could accomplish both these tasks without any central control. Thinking of the complex systems needed to maintain the palace commissary, he wrote, "Go to the ant, consider her ways and be wise. Having no guide, overseer or ruler, she prepares her bread in summer and gathers her food at harvest time." He knew the complexity of a military organization and was impressed that "locusts have no king, yet all of them go forth by companies." Nearly 3000 years later, a participant in the NCMS Virtual Enterprise Workshop (1994) commented, "we used to think that bugs were the problem. Now we suspect they may be the solution!"

2.0 Design Principles of Agent-based Models (ABM) versus Equation-based Models (EBM)

Examination of naturally occurring Agent-based systems suggests design principles for the next generation of Agents. While particular circumstances may warrant deliberate exceptions, in general:

[1] Agents should correspond to "things" in the problem domain rather than to abstract functions.

[2] Agents should be small in mass, time (able to forget) and scope (avoid global knowledge action).

[3] Multi-Agent Systems should be decentralized (no single point of control/failure).

[4] Agents should be neither homogeneous nor incompatible but diverse.

[5] Agent communities should include a dissipative mechanism (entropy leak).

[6] Agents should have ways of caching and sharing what they learn about their environment.

[7] Agents should plan <u>and</u> execute concurrently rather than sequentially.

Difference in representational focus between ABM versus EBM has consequences for modularisation. EBMs represent the system as a set of equations that relate observables to one another. The basic unit of the model, the equation, typically relates observables whose values are affected by the actions of multiple individuals, so the natural modularization often crosses boundaries among individuals. ABM represents the internal behavior of each individual. An Agent's behavior may depend on observables generated by other individuals, but does not directly access the representation of those individuals' behaviors, so the natural modularization follows boundaries among individuals. This difference in model structure gives ABM a key advantage in commercial supply chain management, in 2 ways:

First, in an ABM, each firm has its own Agents. The internal behaviors of Agents are not required to be visible to the rest of the system. Firms can maintain proprietary information about their internal operations. Groups of firms can conduct joint modeling exercises (MarketPlace) while keeping their Agents on their own computers, maintaining whatever controls are needed. Construction of EBM require disclosure of relationships that a firm maintains on observables so that equations can be formulated and evaluated. Distributed execution of EBM is not impossible, but does not naturally respect boundaries among the individuals (why several public e-MarketPlaces failed to survive).

Second, in many cases, simulation of a system is part of a larger project whose desired outcome is a control scheme that more or less automatically regulates behavior of the entire system. Agents correspond 1-to-1 with individuals (firms or divisions of firms) in the system being modeled, and their behaviors are analogs of the real behaviors.

These two characteristics make Agents a natural locus for the application of adaptive techniques that can modify their behaviors as Agents execute, so as to control emergent behavior of the overall system. Migration from simulation model to adaptive control model is much more straightforward in ABM than EBM. One can imagine a member of adaptable business network or supply chain using its simulation Agent as the basis for an automated control Agent that handles routine interactions with trading partners. It is less likely that such a firm would submit aspects of its operation to an external "equation manager" that maintains specified relationships among observables from several firms.

ABMs support direct experimentation. Managers playing "what-if" games with the model can think directly in terms of familiar business processes, rather than having to translate them into equations relating observables. ABMs are easier to translate back into practice. One purpose of "what-if" experiments is to identify improved business practices that may be implemented. If the model is expressed and modified directly in terms of behaviors, implementation of its recommendations is a matter of transcribing the modified behaviors of Agents into task descriptions for the underlying physical entities in the real world.

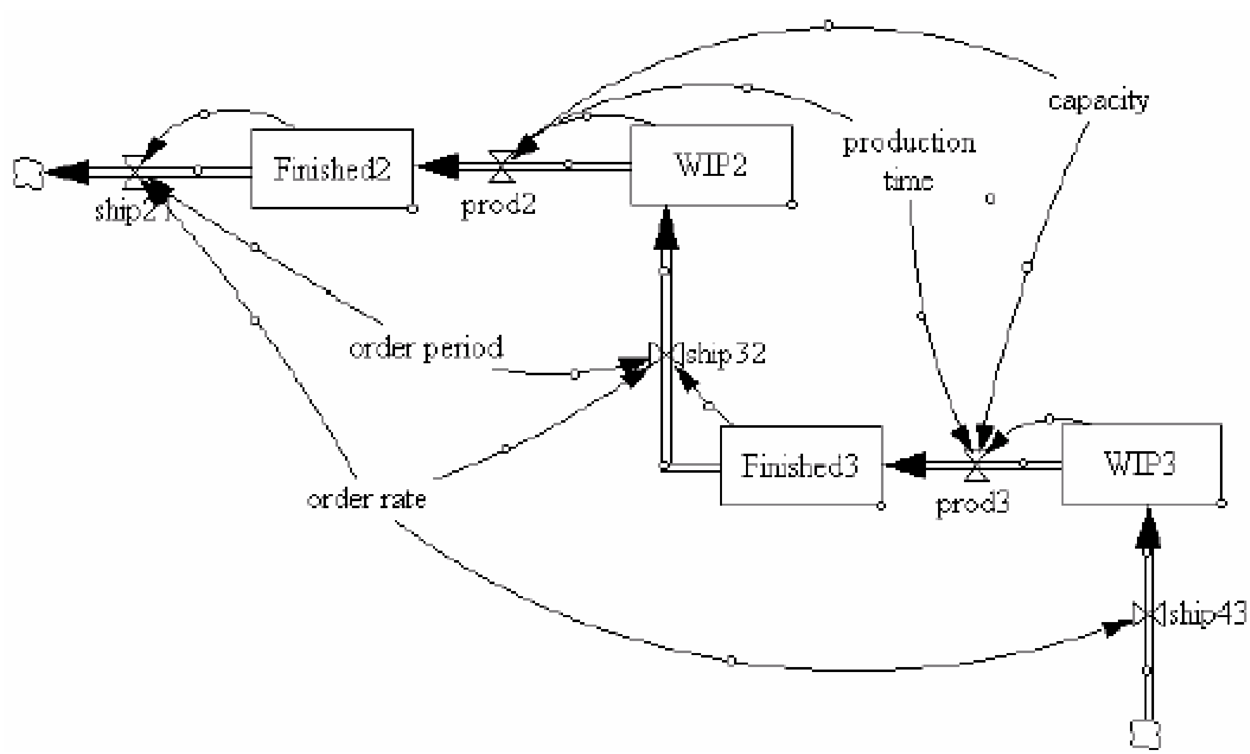
Disadvantages of EBM result largely from use of averages of critical system variables over time and space. EBM assumes homogeneity among individuals but individuals in real systems are heterogeneous. When the dynamics are non-linear, local variations from the averages can lead to significant deviations in overall system behavior. In business applications driven by 'if-then' decisions, non-linearity is the rule.

Because ABMs are inherently local, it is natural to let each Agent monitor the value of system variables locally, without averaging over time and space and thus without losing the local idiosyncrasies that can determine overall system behavior.

Both approaches simulate the system by constructing a model and executing it on a computer. The differences are in the form of the model and how it is executed. In Agent-based modeling (ABM), the model consists of a set of Agents that encapsulate the behaviors of the various individuals that make up the system and execution consists of emulating these behaviors, which is essentially dynamic. In equation-based modeling (EBM), the model is a set of equations (pre-determined static) and execution consists of evaluating them. Thus "simulation" is a general (umbrella) term that applies to both methods, which are distinguished as Agent-based emulation and equation-based evaluation.

The approach to system design and supply chain management with Agents in the software landscape is at odds with the centralized top-down tradition in current systems. The question usually arises in terms of the contrast between local and global optimization. Decision-makers fear that by turning control of a system over to locally autonomous Agents without a central decision-making body, they will lose value that could have been captured by an integrated (enterprise) global approach. The benefits of the Agent-based architecture approach versus the centralized approach are conditional, not absolute. In a stable environment, a centralized approach can be optimized to out-perform initial efforts of an opportunistic distributed system of Agents. If the distributed system has appropriate learning capabilities, it will eventually become as efficient. Market conditions are marked by rapid and unpredictable change, not stability. Thus, the appropriate comparison is not between local and global optima but between static versus adaptable systems that can respond to uncertainty.

Following the pioneering work on system dynamics by Jay Forrester of MIT, virtually all simulation work to date, principally on supply chains, integrates a set of ordinary differential equations (ODE) over time. These models are graphically represented, using a notation that suggests a series of tanks connected by pipes with valves. The illustration below offers an example of material flow in a supply chain.



The *rectangular boxes* (Finished2, WIP2, Finished3, WIP3) are "levels" or variables whose assignments change over time. In this particular model, they represent four critical inventory levels (sites 2 and 3, in this model), a work-in-process (WIP) inventory and a finished goods inventory for each site. The *arrows with valve symbols* (shaped like hour-glasses: ship21; prod2; ship32; prod3; ship43) are flows between the levels they connect. The valves are "rates" or variables that determine the rate of flow. For example, the rate "prod2" is the rate at which site 2 converts work-in-process to finished goods. The *cloud shapes* at the upper-left and lower-right corners represent external sources and sinks. In this case, upper-left corner represents end consumer (site 1), lower-right represents the supplier (site 4). The *legends* associated with neither a box nor a valve (order rate, order period, production time, capacity) are auxiliary variables. *Single-bodied arrows* show the dependency of rates on other variables (both levels and auxiliaries). The exact mathematical form of dependency is not shown in the graphical representation, but is encoded in the rate. For example, the arrows show that the rate "prod2" depends on the level "WIP2" and the auxiliary's production time and capacity.

This illustration models the components involved in the interplay between site capacity and order rate. When the order rate exceeds the site capacity, it demonstrates oscillations that, by extrapolation, may be representative of the fluctuations that characterize the supply chain. This system dynamics model shows the same periodicities as the Agent-based model but it does not show many of the effects that we observe in ABM and in real supply networks (memory effect of back orders, transition effects, amplification of order variation). Such effects require explicit representation of levels and flows for orders as well as parts. In particular, they require a model of production planning and inventory control (PPIC) in the system dynamics formalism (not easy to produce). System dynamics models of this nature are widely used in studying organizational dynamics, business processes, environmental and ecological systems, policy implications and a wide range of similar domains. In principle, ABM can be applied to all of these domains, often in a way that seems more natural.

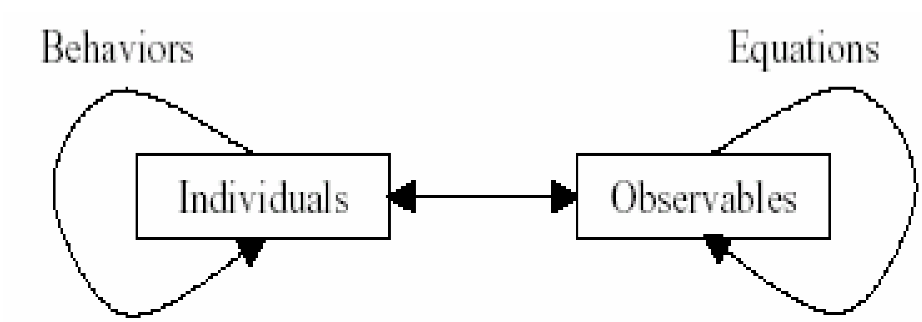
2.1 Differences in Design Principles between ABM and EBM

ABM and EBM share some common concerns but differ in two ways: the fundamental relationships among entities that they model and the level at which they focus their attention. Both approaches recognize that the world includes 2 kinds of entities: individuals and observables, each with a temporal aspect. Individuals are bounded active regions of the domain. In some domains, the boundaries that set individuals apart are physical, as when we are studying ants or bees or people. In other domains, the boundaries may be more abstract, as in the case of the supply chain model, each representing a business firm. In any case, the boundaries are such that those who are interested in the domain recognize the individuals as distinct from one another. They are "active regions" because those interested in the domain conceive of individuals as entities that possess behaviors. Individuals "do things" as time passes.

Observables are measurable characteristics of interest. They may be associated either with separate individuals (velocity of gas particles in a box) or with the collection of individuals as a whole (pressure in the box). In general, the values of these observables change over time. In both kinds of models, these observables are represented as variables that take on assignments. Each of these sets of entities invites us to articulate the relationships that unify it and show how those relationships predict the behavior of the overall system through time. The first fundamental difference between ABM and EBM is in the relationships on which one focuses attention.

EBM begins with a set of equations that express relationships among observables. The evaluation of these equations produces the evolution of the observables over time. These equations may be algebraic or they may capture variability over time (ordinary differential equations, as used in system dynamics) or over time and space (partial differential equations). The modeler may recognize that these relationships result from the interlocking behaviors of the individuals but those behaviors have no explicit representation in EBM.

ABM begins not with equations that relate observables to one another but with behaviors through which individuals interact with one another. These behaviors may involve multiple individuals directly (for example, foxes eating rabbits) or indirectly through a shared environment (for example, horses and cows competing for grass). The modeler pays close attention to the observables as the model runs and may value a parsimonious account of the relations among those observables. However, such an account is the result of the modeling and simulation activity, not its starting point. The modeler begins by representing the behaviors of each individual, then turns them loose to interact. Direct relationships among the observables are an output of the process, not its input. The illustration below summarizes the critical relationships:



- Individuals are characterized, separately or in aggregate, by observables and affect the values of these observables by their actions.
- Observables are related to one another by equations.
- Individuals interact with one another through their behaviors.

A second fundamental difference between ABM and EBM is the level at which the model focuses. A system is made up of a set of interacting individuals. Some of the observables of interest may be defined only at the system level (for example, pressure of an enclosed gas), while others may be expressed either at the individual level or as an aggregate at the system level (for example, location of an organism vs the density of organisms per unit space of habitat). EBM tends to make extensive use of system level observables, since it is easier to formulate parsimonious closed-form equations using such quantities. In contrast, the natural tendency in ABM is to define Agent behaviors in terms of observables accessible to the individual Agent, which leads away from reliance on system-level information. In other words, the evolution of system-level observables does emerge from ABM but the modeler is not as likely to use these observables explicitly to drive the model's dynamics as in EBM.

These two distinctions are tendencies, not hard and fast rules. The two approaches can be combined within an individual Agent in an ABM. Behavior decisions may be driven by the evaluation of equations over particular observables and one could implement an Agent with global view whose task is to access system-level observables to make them visible to local Agents, thus driving an ABM with system level information. Furthermore, while Agents can embody arbitrary computational processes, some equation-based systems (based on PDE's, but not the simple ODE's used in system dynamics) are also computationally complete. The decision between the two approaches, or a mix, must be made case by case, on the basis of practical considerations.

2.2 ABM versus EBM: Practical Considerations

The difference in representational focus between ABM and EBM has consequences for how models are modularized. EBMs represent the system as a set of equations that relate observables to one another. The basic unit of the model, the equation, typically relates observables whose values are affected by the actions of multiple individuals, so the natural modularization often crosses boundaries among individuals.

ABM represents the internal behavior of each individual. An Agent's behavior may depend on observables generated by other individuals, but does not directly access the representation of those individuals' behaviors, so the natural modularization follows boundaries among individuals. This fundamental difference in model structure gives ABM a key advantage in commercial applications such as public portals hosting e-marketplaces or adaptable supply network modeling, in two ways:

First, in an ABM, each firm has its own Agents. An Agent's internal behaviors are not required to be visible to the rest of the system, so firms can *maintain proprietary information* about their internal operations. Groups of firms can conduct joint modeling exercises (eMarkets) while keeping their individual Agents on their own computers, maintaining whatever controls are needed. EBMs require disclosure of the relationships that each firm maintains on observables so that the equations can be formulated and evaluated. Distributed execution of EBM is not impossible but does not naturally respect commercially important boundaries (why eMarkets failed to take off).

Second, in many cases, simulation of a system is part of a larger project whose desired outcome is a control scheme that more or less automatically regulates the behavior of the entire system. Agents correspond one-to-one with the individuals (firms or divisions of firms) in the system being modeled. Agent behaviors are analogs of the real behaviors. These two characteristics make Agents a natural locus for the application of adaptive techniques that can modify their behaviors as the Agents execute, so as to control the emergent behavior of the overall system.

Migration from simulation model to adaptive control model is much straightforward in ABM than in EBM. One can imagine a member of adaptable business network or supply chain using its simulation Agent as the basis for an automated control Agent that handles routine interactions with trading partners. It is much less likely that such a firm would submit aspects of its operation to an external "equation manager" that maintains specified relationships among observables from several firms.

2.3 ABM versus EBM: Compare and Contrast

Agent-based Model	Equation-based Model
Better suited to domains where the natural unit of decomposition is the individual rather than the observable or equation and where physical distribution of computation across multiple processors is desirable.	Better suited to domains where the natural unit of decomposition is the observable or equation rather than the individual.
Naturally represents the process as a set of behaviors, which may include features difficult to represent as rates and levels, such as step-by-step processes and conditional decisions.	Represents the process being analyzed as a set of flow rates and levels. ODEs are well-suited to represent purely physical processes. However, business processes are dominated by non-linear, discrete decision-making.
Easier to construct. Certain behaviors are difficult to translate into rate-and-level formalism. PPIC algorithms are an example.	Includes "black boxes" for specific entities (such as conveyors or ovens) whose behavior is difficult to represent in a pure rate-and-level system.
Distinguishes physical space from interaction space.[27] Permits the definition of arbitrary topologies for Agent interactions.	ODE methods, such as system dynamics, have no intrinsic model of space. PDE's provide a parsimonious model of physical space but not of interaction space.
Validated at the system level and at the individual level, since the behaviors encoded for each Agent can be compared with local observations on the actual behavior of the domain individuals.[28]	Only validated at the system level by comparing model output with real system behavior.

[27] In many applications, physical space helps define which individuals can interact with one another. Decades ago, customer-supplier relationships were dominated by physical space, leading to the development of regional industries, such as the automotive industry in southeast Michigan (USA) and Bavaria (Germany). Advances in telecommunications and transportation along with globalization now enable companies that are physically separate from one another to interact relatively easily, so that automotive suppliers in Michigan are experiencing unprecedented levels of competition with suppliers based in Mexico or the Pacific rim. Such examples show that physical space is an increasingly poor surrogate for interaction space in global commerce.

[28] ABM's are easier to translate back into practice. One purpose of "what-if" experiments is to identify improved business practices. If the model is expressed and modified directly in terms of behaviors, implementation of recommendations is a matter of transcribing the modified behaviors of Agents into task descriptions for the underlying physical entities in the real world. In many domains, ABM gives more realistic results than EBM, for manageable levels of representational detail. The qualification about level of detail is important. Since PDE's are computationally complete, in principle, one can construct a set of PDE's that completely mimics the behavior of any ABM (producing the same results). However, the PDE model may be much too complex for reasonable manipulation and comprehension. EBM's (like system dynamics) based on simpler formalisms than PDE's may yield less realistic results regardless of the level of detail in the representation. For example, the dynamics of traffic networks achieved more realistic results from traffic models that emulate the behaviors of individual drivers and vehicles, compared with the previous generation of models that simulate traffic as flow of a fluid through a network. The latter bears similarities to the flow-and-stock approach to supply chain simulation.

Applications driven by "if-then" decisions are non-linear. ABMs are inherently local and hence allows each Agent to monitor the value of system variables locally, without averaging over time and space and thus without losing the local idiosyncrasies that determine and/or influence overall system behavior.	Use of averages of critical system variables over time and space. Assumption of homogeneity when real systems are generally heterogeneous. Local variations from averages may lead to significant deviations in overall system behavior under conditions which are typical of non-linear dynamics.[29]

3.0 Characteristics of Agent-Oriented Programming

The history of software is one of increasing localization and encapsulation. Originally, the basic unit of software was the complete program. Arbitrary jumps of control made it difficult to manipulate any unit other than the individual line of code and the entire program. Data often occupied the same deck of cards as the program and the entire deck (code and program) was the responsibility of programmer, who thus determined the behavior of the complete ("monolithic") program before it began execution.

The "structured" programming movement designed programs from smaller packages of code, such as structured loops and subroutines, with a high degree of local integrity. Though a subroutine's code was encapsulated, its state had to be supplied externally through arguments and it gained control only when externally invoked by a call (for example, remote procedure call).

The next generation was "object-oriented" programming, which localized not only a segment of code but also the variables manipulated by that code. Originally, objects were passive and gained control only when some external entity sent them a message. Objects are visible, have specific locations and may contain things.

---

[29] The assumption of homogeneity in EBM based on system dynamics may find a parallel with the assumption of homoskedasticity in forecasting models based on classical linear regression models. In both instances real world situations are unlikely to be represented. Variations in local conditions that system dynamics may fail to capture resonates with the concept of volatility of data that traditional regression models ignore by using "sample data" analogous to "averages of critical system variables" in EBM. Whereas ABM offers closer to real-world representation over EBM by encapsulating individual behaviors, the time series technique of GARCH (generalized autoregressive conditional heteroskedasticity) may be potentially useful to improve forecasting accuracy by taking into consideration the volatility of real-time data from operational systems (for example, supply chain) that is increasingly available in high volume due to gradual adoption of automatic identification technologies, such as, use of radio frequency identification (RFID). For a discussion on potential of GARCH in forecasting: Advances in Supply Chain Management and Decision Support Systems: Potential to Improve Forecasting Accuracy by Shoumen Datta *et al* (http://esd.mit.edu/WPS/esd-wp-2006-11.pdf).

Agent architectures gives each object its own thread of control and its internal goals, thus localizing not only code and data, but also *invocation*. Such an "active object with initiative" is the most basic manifestation of Agent software, sometimes called an "Autonomous Agent" to emphasize that it does not need external invocation. It is also referred to as "responsible agent" to emphasize that it watches out for its own set of internal responsibilities.

100 agents, each with 10 behaviors, require programming of 1000 individual behaviors, yet provide a behavior space of 10^{100} which is a number far larger than total number of elementary particles in the universe! Integration and maintenance costs, traditionally two of the largest expenses in software development, are greatly reduced if one compares the cost of programming for 1000 entities that deliver the functional equivalent of 10^{100} entities.

With code, state and control all localized within the Agent, little or no integration is required to launch an application. Agents can notice things, can carry out actions, can go places (environment) and can learn (therefore, adapt and respond to environment). In the ultimate Agent vision, the application developer simply identifies the specific mixture of Agents required or desired for each of the processes in the final application. The Agents organize themselves and the self-organizing emergent behavior of multi-Agent systems executes the most appropriate and productive response required for optimal functionality. Thus, the appeal of Agent architectures depends on the ability of populations of Agents (Multi-Agent Systems) to self-organize and respond dynamically to changing circumstances without top-down control from a system operator.

Multi-Agent Systems (MAS) are likely to share 4 properties and such systems also employ 3 mechanisms. These features are analytic (not synthetic) and represent properties or mechanisms common to existing complex adaptive systems or multi-Agent systems. The four properties are:

[1] Aggregation

MAS are not constructed monolithically but consist of smaller Agents which may be themselves aggregates of still smaller units (Agents). The behavior of the aggregate is often distinct from the individual behaviors of the parts.

[2] Non-linearity

The behavior of MAS is not linear and their interactions are, thus, not simply additive.

[3] Flows

MAS are characterized by flows of various substances through networks of Agents. These flows exhibit two important effects: multiplication (in which one change produces a chain of others) and recycling (feedback loops).

[4] Diversity

The Agents in a MAS could differ from one another.

The three mechanisms are:

[1] Tagging

Agents in a MAS are able to recognize and differentiate among one another (share information).

[2] Internal Models

The internal structure of an Agent in a MAS enables it to anticipate changes in its environment (share information).

[3] Building Blocks

Agent's internal model is made up of small, reusable modules, enabling it to capture a rich set of alternatives with a limited representational vocabulary. This is where the impact of ontologies is most profound (share information).

The paradigm shift in software medium fostered by Agent technology requires new ways for people to think about decentralized systems. The following "guiding heuristics" may be used to learn to understand these systems:

[1] Positive feedback is not always negative. Sometimes it leads to destructive oscillations but in other cases it is critical to self-organization (entropy).

[2] Randomness can help create order by providing diversity among Agents.

[3] A flock is not a big bird. The behavior of an aggregate system is not the same as the individual behaviors of the lower-level units out of which it is constructed.

[4] A traffic jam is not just a collection of cars. Decentralized systems generate emergent objects that are distinct from any of the individual parts.

[5] The hills are alive. Environment is an active process that impacts the behavior of the system, not just a passive communication channel between Agents.

[6] Emergent behavior is distributed across the components rather than localized in any single one (decentralize).

[7] Persistent disequilibria. A useful system must balance stability with constant change, otherwise it will degenerate (entropy).

[8] Change changes itself. In dealing with changing circumstances, complex systems change and over time the rules of change undergo change (share information).

[9] Think small. Keep Agents small, in mass, time and space / scope.

4.0 Bio-Inspired Systems Influence the Development and Design Principles of Agents

Agents-based software architecture may be viewed as a reductionist approach to deal with equation-based software complexity that continues to increase exponentially. Part of the answer lies in the fact that teamwork of social insects is decentralized. An ant's individual solution is primitive but collectively it results in efficient solutions to complex problems. The "active" key is the deposition of pheromone along their trails in the communal travel of ants. Pheromones are a medley of cyclical biological macromolecules affecting olfactory and neuro-sensory pathways to elicit diverse responses in insects, humans, yeasts. Some pheromones may belong to arachidonic acid derivatives generically referred to as prostaglandins and influence a diverse variety of responses in humans.

Advances in data routing emerged from study of ant-based algorithms, based on concept of *swarm intelligence*. To build better data networks, "artificial ants" created as software, travel through programs depositing "artificial pheromones" while in pursuit of route optimisation. Bending the rules of this ant behavior is a primary step. Bending the rules involve endowing the "soft-ants" with memory (parameters, variables, formulas) and enabling them to retrace "good routes" and mark them with extra "artificial pheromones" that then may be extrapolated to routing, marketing and event management strategies.

Ant-based algorithms are also being borrowed to solve a classic puzzle known as the traveling salesman problem. It can be also related to systems, goods, raw materials, parcels, packets of data. As the number of cities involved increases in the traveling salesman problem, the difficulty of the problem increases exponentially. In one model, artificial pheromone deposited along longer routes evaporate, leaving the links to the greatest number of shortest routes most densely covered with artificial pheromone. When soft-ants are sent out again, they rely on "memory" tables storing information about amount of pheromone on each link. Repeating these trips result in progressively shorter overall trips.

The latter may be particularly relevant to unpredictability in data flow for supply chain volatility or internet data traffic. Because soft-ants are constantly exploring different routes, many alternatives surface if a particular "good" route goes out of commission. The processing ability in this model is based on numbers. If there are 5 ants in a "colony" then this complex processing fails but with 10,000 or 100,000 "soft-ants" the results are impressive.

In complex distributed systems, such as, supply chain network and event management, where routing and sequence is critical, current solutions are offered from people aiming to optimize interactions with a centralized mind-set. Swarm intelligence based decentralized control underlies the paradigm shift for optimisation in future adaptable business networks. These models, as key principles in Agents technology, are able to find very good solutions (but may not be the perfect optimisation) in a reasonably short time. Not provably optimal but that which is "very good" is often required for real-world applications in real-time.

Ants construct networks of paths that connect their nests with available food sources. These networks form minimum spanning trees, minimizing the energy ants expend in bringing food into the nest. Graph theory defines a number of algorithms for computing minimum spanning trees, but ants do not use conventional algorithms! Instead, this globally optimal structure emerges from the simple actions of individual ants. Each ant that forages for food has the same basic program, consisting of five rules that fire whenever their conditions are satisfied:

[1] Avoid obstacles. Whatever an ant does, it will not aimlessly push against a wall.

[2] Wander randomly in the general direction of nearby pheromones. If it does not sense pheromones, an ant executes Brownian motion, choosing each step from a uniform distribution over possible directions. If it senses pheromones, the ant continues to wander randomly but the distribution from which it selects its direction is weighted to favor the direction of the pheromone scent.

[3] If the ant is holding food, it drops pheromone at a constant rate as it walks. In simplest simulations, the ant continues to move randomly. In others, it follows a beacon (distinctive pheromone at the nest) that leads in the general direction of home. Both approaches yield same global behavior. The homing beacon generates paths sooner but continued random wandering works in the emulation as well.

[4] If the ant finds itself at food and is not holding any, it picks up the food.

[5] If the ant finds itself at the nest and is carrying food, it drops the food.

Brownian motion eventually brings the ant arbitrarily close to every point in the plane. As long as the separation between nest and food is small enough compared with the range of the ant, a wandering ant will eventually find food if there is any and (even without a beacon) a food-carrying ant will eventually find the nest again. In most cases, food is available only in some directions from the nest and ants who wander off in the wrong direction will starve or fall to predators. But, as long as there is food close enough to the nest and as long as there are enough ants to survey the terrain, the food will be found.

Only food-carrying ants drop pheromone and because ants can carry food only after picking it up at a food source, all pheromone paths lead to a food source. Because pheromones evaporate, paths to depleted food sources disappear, as do paths laid down by food-carrying ants that never reach home. Paths that touch the nest are easily found by out-bound ants. As long as they lead to food, they will be reinforced by those ants once they pick up food. The initial path will not be straight but the tendency of ants to wander even in the presence of pheromones will generate short-cuts across initial meanders. Because pheromone paths have some breadth, they tend to merge together into a trace that becomes straighter, the more it is used. The character of resulting network as a minimal spanning tree is not intuitively obvious from individual behaviors, but does emerge from the emulation.

An ant hill houses different kinds of things, including larvae, eggs, cocoons and food. The ant colony keeps these entities sorted by kind. For example, when an egg hatches, the larva does not stay with other eggs but is moved to the area for larvae. Computer science has developed many algorithms for sorting things, but ants in the ant hill are not executing a sorting algorithm! Individual ant algorithms that yields system-level sorting behavior contains some behaviors similar to those in the path-planning problem:

[1] Wander randomly around the nest.

[2] Sense nearby objects and maintain a short memory (about ten steps) of what has been seen.

[3] If an ant is not carrying anything when it encounters an object, decide stochastically whether or not to pick up the object. The probability of picking up an object decreases if the ant has recently encountered similar objects.

[4] If an ant is carrying something, at each time step decide stochastically whether or not to drop it. The probability of dropping a carried object increases if the ant has recently encountered similar items in the environment.

As in path planning, Brownian walk eventually brings the wandering ants to examine all objects in the nest. Even a random scattering of different items in the nest will yield local concentrations of similar items that stimulate ants to drop other similar items. As concentrations grow, they tend to retain current members and attract new ones. The stochastic nature of the pick and drop behaviors enables multiple concentrations to merge, since ants occasionally pick up items from one existing concentration and transport them to another. The put-down constant $\boldsymbol{k^-}$ must be stronger than the pick-up constant $\boldsymbol{k^+}$ or else clusters will dissolve faster than they form. Typically, $\boldsymbol{k^+}$ is about 1 and $\boldsymbol{k^-}$ is about 3. The length of short-term memory and the length of the ant's step in each time period determine the radius within which the ant compares objects. If the memory is too long, the ant sees the entire nest as a single location and sorting will not take place. The limited short-term memory of ants ensures that ants forget.

The ant's ability to "forget" is a boon to real-world adaptable business networks if impregnated with software Agents. In traditional equation-based planning algorithms, demand forecasting is based on a weighted-average of past consumption data. If there was an anomaly, for example, spike of sales 20 weeks ago that was a variant, the planning algorithm continues to consider that value because equation-based modeling cannot "forget" facts, although the weight will decrease over time. The forecasting engine, therefore, continues to reflect the effect in its subsequent forecast for many weeks unless the parameters are manually changed. Multiplicity of such events along the various stages of the supply chain lead to aberrant forecasting and may contribute to the fluctuations that manifests as the Bullwhip Effect.

Ant-based algorithms, utilizing the concept of swarm intelligence, enables Agents to "forget" since Agent behavior is modeled on natural systems behavior rather than equation-based models of current software. The ability to learn and to forget, enables Agents to be adaptive. Such adaptive Agents in combination with improved analytical engines may enhance planning and forecasting accuracy aided by real-time data from Agents collecting data from RFID tags. The latter may result in reduction of inventory. Reduced inventory reduces working capital charges and may reduce production waste. These reductions improve return on assets as the manufacturing cash cycle gets shorter.

4.1 Task Differentiation

Mature *Polistes* wasps in a nest are divided in 3 groups: a single Chief, a group of Foragers who hunt for food and a group of Nurses who care for the brood. These varied roles are filled by genetically identical wasps. The relative proportion of Foragers and Nurses varies with the abundance of food and the size of the brood. The nest has resource manager or management and no wasp (not even the Chief) computes what this proportion should be.

Each wasp maintains a *Force* parameter that determines its mobility and a *Foraging Threshold* that determines how likely the wasp is to go seek for food. The brood maintains a third parameter, *Demand,* which stimulates the Foragers. Wasp behaviors involve interactions of these parameters.

[1] When two wasps meet, they engage in a face-off. The winner is chosen stochastically. Wasp with higher force has a higher probability of winning but the wasp with lower force will occasionally win. A quantum of force is transferred from the loser to the winner wasp.

[2] When the brood receives food, it reduces demand.

[3] A wasp near the brood, determine probabilistically whether or not to forage. Successful stimulation reduces foraging threshold by a certain constant*,* the learning coefficient, while failure to forage increases foraging threshold by a constant, the forgetting coefficient.

Confrontation among wasps shifts force from one to another and represents a form of communication. Foraging reduces brood's demand and thus brood's stimulation on nearby wasps. Stimulation reduces wasps' thresholds and triggers foraging.

When a community of "soft-wasps" executes these behaviors over a period of time, the population stabilizes into 3 groups, corresponding to the division observed in the natural insects. A group with high force and low threshold corresponds to the Foragers, who have both the strength to move about and the sensitivity to the brood to respond to their needs. A second group with low force and low threshold corresponds to the Nurses, who also are attentive to the brood but, lacking force, cannot move about and must remain near the brood. Finally, a single wasp with high Force and high threshold corresponds to the Chief. The Chief does not command or control the others but grounds the force and threshold scales (by wandering around the nest and facing off against the other wasps) and balances these variables across the population.

4.2 Local and Global Optimisation

Flocks of birds stay together, coordinate turns, avoid collisions with obstacles and each other. Schools of fish exhibit similar coordinated behavior. Humans address similar problems, in air-traffic control and convoys of ships but conventional solutions depend on communication and central coordination structures. Most sophisticated human coordination cannot handle density of coordinated entities of fishes or birds. Each bird or fish follows only 3 simple rules to achieve such coordination:

[1] Maintain a specified minimum separation from the nearest object (other birds, fish)

[2] Match velocity (magnitude and direction) to nearby objects

[3] Stay close to the center

The flock or school is a self-constraining structure in which each entity's individual actions (local) simultaneously respond to and change the overall structure of the flock (global). Although each bird or fish senses only the movements of its nearest peers, its responses to these movements propagate to others, so that the system as a whole exhibits global coordination. Natural systems behavior shares common principles of self-organization. As we learn to recognize and understand these principles, we construct artificial systems that emulate the desirable behavior we observe in these natural systems. These are key to Agent engineering principles.

4.3 Bio-Inspired Design Principles of Agents: Small

Naturally occurring adaptive systems have parts that are small compared with the entire system, in mass, time and space. Tropical termites construct mounds that can exceed five meters in height and ten tons in mass. These multi-story structures store food, house the brood and protect the population from predators. The existence of some of these structures has been documented for over 350 years, which is as long as they have been accessible to the European compulsion for chronological records. In spite of the complexity, durability and effectiveness of these structures, no termite serves the role of a chief engineer to plan the structure and manage its construction. Termites draw on the pheromone mechanism as follows:

[1] Metabolize bodily waste, which contains pheromones. Excretory waste is the material from which termite mounds are constructed.

[2] Wander randomly but prefer the direction of the strongest local pheromone concentration.

[3] At each time step, decide stochastically whether to deposit current load of waste.

The probability of making a deposit increases with the local pheromone density and the amount of waste that the termite is currently carrying. A full termite will drop its waste even if there is no other nearby deposit and a termite that senses a very high local concentration of pheromones will deposit whatever waste it is carrying, even if it is a relatively small amount. The probabilistic algorithm leads to the generation of scattered initial deposits. These deposits attract termites that wander close enough to smell them and increase the probability that these visitors will make reinforcing deposits. Because pheromones decay over time, the most recent deposits at the center of the pile are the strongest and thus the piles tend to grow upward rather than outward, forming columns. When two columns grow near one another, the scent of each attracts termites visiting the other, thus pulling subsequent deposits into the shape of an arch. A similar dynamic leads to the formation of floors joining multiple arches. When one floor is complete, the cycle repeats to construct the next. Each termite is an almost negligible part of the entire termite hill. As a result, the behavior of the whole is stable under variations in the performance of any single member. Collective dynamics dominate. This and similar examples suggest implementing artificial systems with large numbers of Agents, each small in comparison with the whole system.

The motivation for this principle derives not from the theory of multi-Agent systems, but from the experience of software engineers. Difficulty of designing, implementing and launching computer-based systems increases exponentially with the size of the system. Individual Agents are easier to construct and understand than large monolithic systems. Thus, the impact of the failure of any single Agent will be minimal. In addition, a large population of Agents gives the system a richer overall space of possible behaviors, thus providing for a wider scope of emergent behavior. Roughly, the number of agents is a multiplicative factor in determining implementation effort, but an exponent in determining the size of the overall system state space (100 agents with 10 behaviors each is equal to a resulting state space of 10^{100}).

Keeping agents small in mass often means favoring specialized Agents over more general ones, using appropriate aggregation techniques. For example, a complete manufacturing cell would be extremely complex as a single Agent, but can be developed as a community of Agents for individual mechanisms (one for the fixture, one for the tool, one for the load-unload mechanism, one for the gauging station). Multi-Agent systems supports such aggregation, since an environment (and its associated Agents) can become a higher-level Agent by defining its inputs and outputs to another environment.

Assignment of a separate Agent to each machine, each part, each tool, in a manufacturing enterprise provides Agents that are light weight than traditional shop-floor software systems. The automation and industrial controls market offers increasing support for Agents as small as the sensor/actuator level, including small controllers of traditional design as well as more novel architectures combining computation, control and networking.

Naturally occurring Agent systems can "forget" because pheromones evaporate and obsolete paths leading to depleted food sources disappear rather than misleading members of the colony. The mechanism of forgetting is an important supplement to the emphasis in conventional artificial intelligence (AI) systems on mechanisms for learning. In a discrete-event system, forgetting can be as complex as learning since both represent discrete state transitions. In a time-based system, forgetting can take place "automatically" through the attenuation of a state variable that is not explicitly reinforced.

Pheromones suggest that a time-based environment can support a "forgetting" function for discrete-event agents. In one model, each segment of a switched conveyor system is a separate Agent that seeks to maintain its population of a given part type within certain limits. If the population rises above the upper limit, the segment seeks to spill excess parts to some neighbouring segment. If the population falls below the minimum, the segment sends requests out to neighbouring segments. The probability that a segment will spill a part to a given neighbour is a state variable for the spilling segment. It is increased each time the neighbour requests a part and decreased each time the segment spills a part to the neighbour. Because the information regarding a neighbour's interest in a part is maintained as a real number rather than symbolically, obsolete behaviors are forgotten automatically as the value is modified to reflect more recent behavior. The rate at which the probability changes in response to interactions with neighbouring segments is a tuning parameter.

Participants in natural systems usually sense their immediate (local) vicinity (birds, fishes). In spite of this restriction, they can generate effects that extend far beyond their own limits (global), such as networks of ant paths or termite mounds. The exposition of the principle, "control from the bottom up" recognizes the superiority of many local interactions over a few global ones. Increasing bandwidth makes it easy to connect Agents directly with every other Agent but it may be beneficial to follow natural examples and engineer Agents that limit recipients of their messages. These limitations need not be geographical but natural examples suggest that effective systems will restrict communications in some way. Where feasible and appropriate, Agents should define the audience that needs to receive the message, at least by subject-based addressing rather than broadcasting information.

A suggestive analogy comes from the history of physics. Newton's classic equation for gravitational force usefully *describes* the effects of gravity but does not *explain* the underlying mechanism. Massive bodies do not figure out how hard to pull on one another by communicating their masses and measuring the distance that separates them. Einstein's theory of general relativity replaces this impossible notion of action-at-a-distance with a model of local interaction in which masses warp space in their immediate vicinity and respond to the local geometry of the space in which they find themselves. Could localization of Agent interactions follow in the intellectual traditions of physics?

4.4 Bio-Inspired Design Principles of Agents: Distributed

How ants sort their nests yields an interesting and related lesson. An ant whose short-term memory is too long "sees" the entire nest at once and is unable to sort. In simple terms, giving an Agent access to too much of the world may lead to sensory overload, reduced ability to discriminate and lower performance. In terms of the <Agents, Environment, Coupling> model, if Agents are small compared to the environment, their State will have fewer elements than the environment's State. Their Input and Output subsets of State will be even smaller. The accuracy of the model of the environment presented by an Agent's Input will begin to decrease once the Agent's scope of sensing is so large that the portion of the environment accessible to it contains more information than its Input can model.

In this model, only environment has a state space large and rich enough to represent the entire system. This observation does not mean that we should neglect engineering the environment. The <Agents, Environment, Coupling> model emphasizes that we must pay attention to both components of the system. However, the rich information interface that computers support is available only through the Agents and there are important reasons not to make one Agent large and complicated enough to control the entire system:

[1] A central Agent is a single point of failure that makes the system vulnerable to accident.

[2] Under normal operating conditions, it can easily become a performance bottleneck.

[3] Even if adequately scaled for current operation, it provides a boundary beyond which the system cannot be expanded.

[4] It tends to attract functionality and code as the system develops, pulling the design away from the benefits of Agents and in time becoming a large software artifact that begins to resemble equation-based system or EBM.

Centralization can sometimes creep in when designers confuse a class of Agents with individual Agents. For example, one might be tempted to represent a bank of paint booths as "the paint Agent" because they all do the same thing. Certainly, one would develop a single class (in an object-oriented sense of the word) for paint-booth Agents but each paint booth should be a separate instantiation of that class.

4.5 Bio-Inspired Design Principles of Agents: Diversity

The difference in size between an individual Agent's State and that of the environment not only favours small Agents and decentralized control but also encourages Agent diversity. The environment's State contains information concerning both opportunities that Agents should exploit and threats that they should avoid. The more of the environment's State the Agents can sense and modify, the better they can exploit those opportunities and avoid threats. Any single Agent can model and manipulate only a small portion of the environment and a population of completely identical Agents will not perform any better since they will still cover only the same subset of the environment's state. A population of diverse Agents will cover more of the environment's state and thus provide better performance.

While diversity is not the same as population size, the two are correlated. In a physical environment, it is impossible for two Agents to occupy the same place at the same time. Thus two otherwise identical Agents will be at different locations in the environment. They will differ in that element of their respective States that records their location. This simple but crucial element of diversity enables them to monitor different elements of the environment's state and collectively be more robust than a single Agent. The important observation is that the advantage of the larger population lies not merely in numbers but in the diversity that results from physical exclusion laws.

Natural populations often have a "critical size" that is much larger than the simple breeding pair that a standard accounting paradigm might justify. If the population falls below this level, the colony dies out. Once an ant finds food, it generates a pheromone trail to guide other ants to it, but the critical initial discovery depends on having enough ants wandering around that some will stumble across food, wherever it may be. Unused alternatives (for example, unsuccessful scout ants) are insurance for unforeseen change, not waste.

The example of similar Agents at different locations illustrates that diversity is not the same as incompatibility. The diversity in location among ants is able to benefit the ant society as a whole because in many other respects the ants are interchangeable. They all like the same kind of food and they all sense the same pheromones. Their diversity of location would be of no value if they were not similar enough to one another to share the benefits that diversity offers. Diversity can be mathematically quantified by formalizing these insights.

Naturally occurring random behavior can be a good model on which to base ways to achieve the diversity a population of Agents needs in order to adapt. Randomized agents attack a problem in a Monte Carlo fashion that does not require a detailed advance model of the domain and lend themselves to much simpler coordination than approaches that require an explicit reason for every decision.

Compared with conventional systems, natural Agent-based systems may seem wasteful. They allocate far more resources to a task than a global analysis may require (redundancy). For example, an ant's ignorance of where food may be found requires a large army of scouts, some of whom wander far from both nest and food and either starves or are taken by predators.

Redundancy supports diversity in two ways. First, the diversity in location among the ants enhances the colony's chances for finding food and thus surviving. Second, the small size of one ant in comparison with the colony means that several individuals can perish without endangering the entire community. These two insights reflect distinction among homogeneity, diversity and incompatibility. Diverse Agents are not homogeneous and can monitor an environment that is much more complex than any single Agent. Because they are not incompatible, the community as a whole can tolerate the loss of any one individual, drawing on the overlapping capabilities of others.

Fish and bird examples show how repulsion among Agents can maintain diversity of location. The same technique could be applied to other aspects of the relation among Agents. In a machine job shop, a new job can be processed more rapidly if the necessary tooling is already on the machine. Machine Agents, representing otherwise identical machine tools in such a shop, might use a repulsive force with respect to their tooling characteristics to see to it that the currently unused machines have diverse tooling, thus increasing the chances that a new job will find a machine ready to run it.

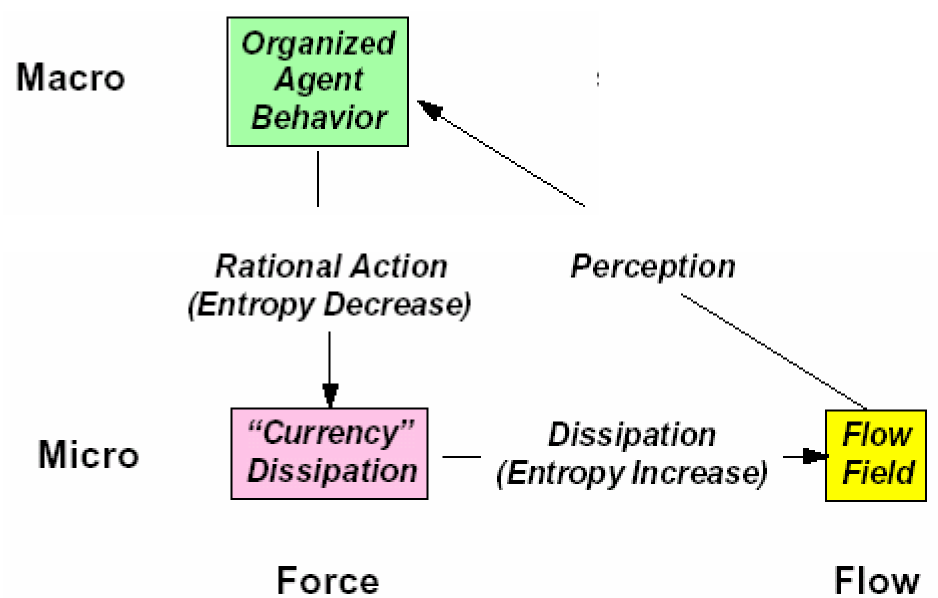
Similarly, natural redundancy and its effect on diversity have useful real-world Agent applications. Sometimes lack of redundancy in the application domain leads to centralisation. If every product in a plant must pass through a single paint booth, it is all too easy to give the Agent representing that booth strong central characteristics. For example, with multiple paint booths, booths might bid for products, but with only one, it calls the shots unilaterally. In such a case, it is sometimes possible to reduce this effect by imagining that there are two or more such Agents and letting the "real" one regularly win run-time competitions. Only the "real" Agent would ever win a bid for work because the others would always report that their machines are off-line. If there is ever a need for new machines, they can be added without modifying the overall design.

Redundant agents can step in for one another because Agents may not communicate directly with one another, but only mediate, by making changes in the environment, that are subsequently sensed by other Agents. Any Agent that senses these changes can respond. Failure of an individual Agent does not bring the system down as long as there are other Agents sufficiently similar to sense the same environmental changes and respond appropriately. Redundancy in production capability may be supported by using a negotiation protocol to identify potential suppliers for the inputs to a given unit process. For example, a unit process that needs a milling machine, issues a request indicating the required machine class in the subject, and through subject-based addressing, all machines of that class, that are on the network, will receive the request. Through a negotiation, the unit process discriminates among available machines on the basis of criteria such as operating cost and availability and selects a platform on which to execute. If one machine breaks down or is fully loaded, this approach permits another to take its place, as long as the physical shop has redundant capabilities.

4.6 Bio-Inspired Design Principles of Agents: Dissipation

The Second Law of Thermodynamics observes that closed systems progressively may become more disordered over time. It is not obvious that a large collection of Agents will organize itself to do useful things. The Second Law warns that the result of such an architecture may be disorder. Natural agent-based systems do organize themselves with striking efficiency. A common explanation is that a system can become more organized if energy is added to it from the outside (for example, by the metabolism of the food). The addition of energy is necessary for self-organization, but hardly sufficient. Gasoline in construction equipment can erect a building but the same gasoline in a terrorist's bomb can destroy it.

In natural systems, Agents can organize themselves at the macro level because their actions are coupled to a dissipative or disorganizing process at a micro level. The system can reduce entropy at the macro level by generating more than enough entropy at the micro level to pay its debt to the Second Law. Thus, it provides an entropy leak to drain disorder away from the macro level (where useful work is done) to the micro level (where it won't interfere with the system's function). In one model of this leakage, micro-level dissipation in the environment generates a flow field that the Agents can perceive and reinforce, and to which they can orient themselves. Insect colonies leak entropy by depositing pheromones whose molecules evaporate and spread through the environment under Brownian motion, generating entropy. The resulting flow produces a field that the insects can perceive and to which they orient themselves in making further pheromone deposits.



The above illustration **(Macro Organization through Micro Dissipation)** generalizes this example in terms of the three fundamental processes: micro dissipation, macro perception of the micro flow field and macro reinforcement of the micro dissipative mechanism.

One example of this model is the movement of currency in a market economy. Money benefits its holders only when they spend it. As it spreads from purchasers to buyers, entrepreneurs perceive the resulting flow field and orient themselves to it, resulting in the self-organization of structures such as financial supply chains and geographic economic centers.

Artificial Agent communities will be more robust and better able to self-organize if they are designed to include a dissipative mechanism (entropy leak) such as a currency model. This mechanism should have three characteristics:

[1] It must flow, either among Agents or through the environment, thus, setting up a gradient field.

[2] Agents must be able to perceive this field and orient themselves to it.

[3] Agent actions must reinforce the field (positive feedback).

4.7 Bio-Inspired Design Principles of Agents: Learn and Share Information

Natural systems exchange information among members of the population at three levels: species, individual and society. Sexual reproduction exchanges information from one generation of a species to another by passing on successful characteristics in the form of chromosomes to offspring. Individual organisms can also pass on skills post-embryonically. For example, young black bears learn to rob food hung from trees by watching older bears. The society as a whole can learn even if individual members do not, as in the development of pheromone paths in insect colony. In each case, a community reduces the need for expensive search by finding ways to cache and share accumulated knowledge.

In the <Agents, Environment, Coupling> model, species and individuals learn by modifications to Agents' State and Process components, while societies learn by modifying the environment's State. Agents can often use similar mechanisms. While learning in a single Agent can require sophisticated techniques, methods for learning across generations, such as genetic or evolutionary programming, have proven their maturity in numerous applications. Changing a community's structure to enable its members to respond better to a changed environment is straightforward for Agent architectures. In addition, Agents architecture can support genetic modification of various aspects of Agent behavior.

4.8 Bio-Inspired Design Principles of Agents: Plan and Execute Concurrently

Traditional systems alternate planning and execution. A firm develops a schedule each night that optimizes its manufacturing the next day. Unfortunately, changes in the real world tend to invalidate advance plans. Engineers in industries as diverse as auto, semiconductors, aerospace and agricultural equipment agree that a daily schedule is obsolete less than an hour after the day begins.

The problem is a natural consequence of a system in which the environment as well as the Agents has a Process that can autonomously modify the environment State with which Agents interact. Natural systems do not plan in advance but adjust their operations on a time scale comparable to that in which their environment changes. The coherence of their behavior over time is maintained by the dynamics of their interactions, not imposed by an external plan or schedule. To achieve the robustness exemplified by these natural systems, Agents should seek to avoid the "plan *then* execute" mode of operation and instead respond dynamically to changes in the environment.

Consider an example of concurrent planning and execution where the actual time at which a job will execute may not be known until the job starts. The resource does not schedule a newly-arrived job at a fixed point in time but estimates probabilistically the job's impact on its utilization over time, based on information from the customer about the acceptable delivery times. The width of the window within which the job can be executed is incrementally reduced over time, as needed, to add other jobs to the resource's list of tasks. If the resource is heavily loaded, the jobs organize themselves into a linear sequence but if it is lightly loaded, the actual order in which jobs are executed is decided at the moment the resource becomes available, depending on the circumstances that exist at that time.

5.0 Adoption of Agents in Decision Support Systems

Market conditions are marked by rapid and unpredictable change, not stability. Therefore, one should evaluate the competing options in a particular case theoretically, strategically, tactically, and practically.

Theoretically, there are decentralized mechanisms that can achieve global coordination. For example, economists have long studied how local decisions can yield globally reasonable effects. Recently these insights have been applied to a number of domains that were not traditionally considered as economic, such as network management, manufacturing scheduling and pollution control.

Strategically, managers must weigh the value of a system that is robust under continual change against one that can achieve a theoretical optimum in a steady-state that may never be realized. A company that anticipates a stable environment may well choose centralized optimisation. One that also incorporates Agent-based decision support software does so because it cannot afford to be taken by surprise.

Tactically, the life-cycle software costs are lower for Agent-based systems than for centralized enterprise software because Agents can be modified and maintained individually at a fraction of the cost of opening up a complex enterprise software system. In systems that must be modified frequently, losses due to sub-optimal performance can recovered in reduced system maintenance expenses.

Practically, Agent-based systems that follow these principles have been piloted or deployed in regular operations. The Agents in these systems regularly reflect the principles outlined here rather than those of centralized systems. Growing acceptance of Agents technology in competitive business environments may be evidence of the benefit they bring to their adopters.

APPENDIX 6: Common Observations and Analysis

Charlie and the Chocolate Factory in Pursuit of Interoperability

Introduction

Some of the suggestions in this article are not new, at least, not anymore, since my original article [1] was posted on the web during August 2003. This article draws on common observations from the public domain and offers rambling analysis that may be suggestive of imperfect clues for future generations of intelligent decision systems that may evolve through the combinatorial convergence of technologies, concepts and ideas to catalyse innovation. I hasten to add that nothing new was invented by the author. The tools, technologies, concepts and ideas are the discoveries of others based on centuries of visionary research. This article simply assembles them in the context of the observations and/or ***processes*** used in business, industry, government, defense, healthcare, security, logistics, services, finance, supply chain, customs operations and other pursuits where interoperability [2] may be necessary.

More innovations can and shall arise from borrowing and combining ideas rather than from isolated inventions. Better decisions (including the profitable ones) stems from an informed convergence of processes that culminate in an action or transaction that leads to an improved outcome. It is erroneously assumed that informed convergence is directly related to and dependent on or proportional to data (assuming available data is accurate). Analytical introspection suggests that data is not equivalent to information and even if information may be extracted from data it does not imply that the information is valuable or may trigger actionable decisions of consequence. One example of how real-time data may be valuable is presented by the GE VeriWise system (Appendix A).

Inability to generate value from data may be rooted in the fact that the analytical tools (software, algorithms) are simply inadequate, unable to adapt to changes or are simplistic representations of complex real-world scenarios. The cumulative inefficiencies stemming from lack of insight with respect to analytical handling of data, information, transaction and decision may be traced back to the ambiguity or inability to ask the right questions at the right time. Such inabilities are pervasive and cannot be ameliorated merely through the investment in technology (for example, the acquisition of real-time data must be based on process innovation or optimization that calls for use of such data). For the logistics sector alone, more than US$3 trillion have been spent in 2004 and this represents almost 5% of the global GDP. Given the enormity of this spending, inefficiencies in logistics network, estimated at 20% or more, indicates losses (due to inefficiency) of US$600 billion. The latter is a glaring example representing an opportunity for organisations to improve decision making. However, businesses cannot grow by "saving" money, hence the prudence to pursue strategies that may stimulate growth through innovation.

Innovation in the domain of dynamic interdependence between data, information and process may be one plausible future direction as to how a broad spectrum of interoperable systems may contribute to improve profitability, aided or unaided by humans. It is this theme that may often surface in this article. Hopefully, this article may even succeed to indicate how pivotal interoperability may be, for any operation, in the context of globalisation.

Can we connect bits to atoms?

Information (bits) about goods (atoms) is therapeutic to consumer product businesses and their friends (retailers, distributors, suppliers, manufacturers). Therefore, connecting bits to atoms, in a systemic sense, may be the holy grail for all those involved in producing and consuming physical goods. But few may learn how to become an expert in "bit dribbling" or ways in which to use the bits to gain value from atoms. Certain characteristics of goods (for example, colour of iPod, temperature of milk and percentage of titanium in alloy) are important to select sectors of consumers (for example, personal shopper, Tesco, Boeing, respectively) who may choose a variety of channels to gain to such information (customer relationship). eTailing is one example of bits connected to atoms. The internet serves increasingly as the medium or channel of choice for consumers to seek or exchange such information.

Therefore, the "now-classical" model (barely a decade old) of Amazon to sell books without any physical inventory is undergoing an evolution where once nearly-warehouse-less Amazon may no longer claim near-zero inventory. Convergence of brick-and-mortar stalwarts on the dot com channel is evident from increasing number of online visitors to tyrannosaurus establishments such as Wal-Mart. As shown below, **illustration 1** (modified from The Economist, 3rd December 2005) offers interesting clues for future developments. Let us explore a few examples:

[a] Shopzilla and comparison-shopping

[b] Amazon and its foresight to collaborate

[c] Apple and the inevitable communication convergence

[d] Why American Greetings and Target are indicators sweeping demographic changes

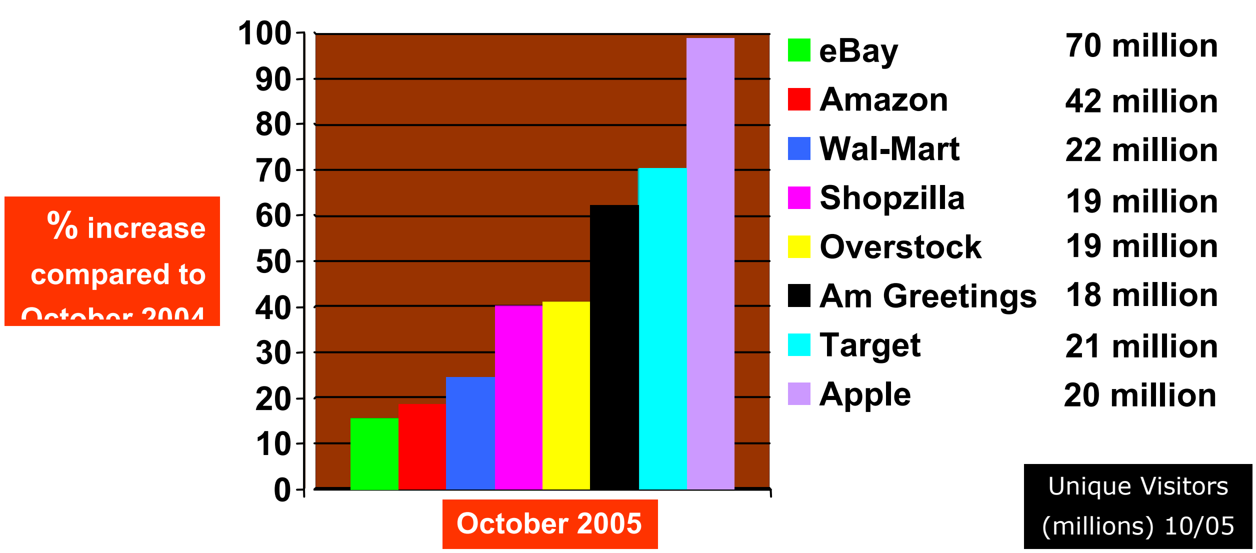


Illustration 1: Connecting Bits to Atoms: A Model of the Physical World ?

[a] Shopzilla and comparison-shopping

Buying the same brand from stores located in two different parts in the same city may offer a considerable price differential. But the savings may diminish if time and travel expenses are taken into consideration. This age old concept is now facilitated by the internet and anyone can take advantage of e-comparison provided one has made a basic investment to access such services. If H&M were to disclose its database of items and prices to a 3rd party (such as Shopzilla in US and/or Ciao in EU) then the threat of such competition may catalyse M&S to do the same. The growth of such comparison shopping may make the end-user (consumer) the beneficiary. The operating scale of Shopzilla is still modest although its ability to attract visitors is appreciable (40% increase when comparing number of unique visitors for October 2004 vs October 2005, see illustration 1).

Consider that Arrow brand shirts, a respectable US mid-market brand, are sold for €50 in Dublin's Brown Thomas and for €45 in Helsinki's Stockmann. But, comparison-shoppers also find No-Name Outfit in Dhaka (Bangladesh) selling the identical Arrow SKU for €15 and advertises that the shirts are authentic Arrow brand being sourced from the manufacturers of Arrow (who are in Dhaka). Globalization and its discontent cannot be more aggravating for shirt-makers! Should Arrow continue to outsource its manufacturing to Dhaka? Should WTO-like organisations empower Arrow to prosecute No-Name Outfit and its partners in Dhaka? Will the tide of ethical globalization disable Arrow to procure shirts at €5 from the East and sell for €50 in the West without profit sharing with its suppliers?

Suddenly comparison-shopping and its intended transparency meets the conundrum of global supply chain or the dilemma that your Christmas shopping can be completed online, from China. Ciao Target! Ciao, Ciao Wal-Mart! However, in the short term, Shopzilla or Ciao's potential for success may be minimally affected by the forces of ethical globalization. Rather, the possibilities for comparison-shopping are limited by the chasm in interoperability of systems that must be accessed by Shopzilla or Ciao to offer value to online comparison-shoppers. If H&M and M&S are the only entrants in this transparency battle then Shopzilla or Ciao remains unattractive to serious bargain hunters given the proliferation of discount stores and sales (Yellow Bag in Stockmann, Helsinki or January Sales in USA). The number of online enthusiasts may fail to offer sufficient economies of scale for survival of comparison-shopping unless systems interoperability can be addressed to gain credible transparency of inventory across a critical mass of stores and multiple brands with minimum effort but without compromising security of transactions.

Technology sympathizers and analysts may offer a grossly different perspective celebrating the 19 million unique visitors to Shopzilla.com during October 2005. In keeping with what is in vogue, this will be touted as the reason why radio frequency identification (RFID) tags may help consumers to save through comparison shopping services. Several assumptions must be made in promoting this line of thought but that is nothing new given that the entire hype curve of RFID usage catalysed in part by MIT's Auto ID Center, made blatantly irresponsible assumptions (for example by Price Waterhouse Cooper Consulting white paper "Focus on the Supply Chain: Applying Auto-ID within the Distribution Center" in 2002) and predicted a fall in RFID tag prices to under 5 cents by 2005. In reality, it is not the cost of technology (for example, RFID tags) that matters but the ***value*** that it delivers to its users. Peering through the lens of history, it appears that lack of imagination fueled by short term corporate gains has plagued the adequacy of value extraction from several technologies and RFID is not an exception. The immense potential for judicious use of automatic identification tags (in various forms) as a data acquisition tool remains to be realized but use of item level RFID tags for comparison-shopping on Shopzilla may not be the "killer application" or strategy.

[b] Amazon and its foresight to collaborate

Mr Jeff Bezos and Mr Jeff Wilke at Amazon headquarters may be less than ecstatic with less than 20% increase in traffic over the same period a year ago despite the 42 million unique visitors to Amazon.com during October 2005. For growth of Amazon in line with the imagination of its senior management, a bit more than paranoia is required. I shall borrow from Mr Andy Grove's "only the paranoid survives" and add a string of adjectives to suggest that only the "insightful collaborative paranoid" shall triumph by exercising their imagination to adapt, continuously.

Hence, Amazon can no longer be classified merely as an online retailer despite the fact that its business model is the subject of global mimicry. To the millions of end-users its brand shall continue to grow as the online store for anything they wish to buy (with peace of mind). To a vast number of small and medium businesses who aspires to be an Amazon but cannot, it serves as a channel to sell their wares as individual businesses operating under the credibility of the Amazon Mall (that may also offer comparison shopping). In this respect, eBay is a carbon copy of the Amazon Mall except for its pricing strategy. The astute use of centuries-old auction pricing sets Pierre Omidyar apart as the innovator *sine qua non*.

Amazon "mall" is essentially an internet-catalysed Sumerian Bazaar. A plethora of such markets existed 7000 years ago along the banks of the Indus River in the cities of Mohenjodaro and Harappa. However, the complexity of the interactions have undergone a radical change and increasingly calls for interoperability. Collaborative transparency may describe one aspect of Amazon Anywhere. It banks on innovative use of standards (and its brand credibility) that are emerging to catalyse the transition of the syntactic web to the semantic web [2]. In view of the evolution of semantic grid web services [1], as a first step, Amazon is offering programmers virtually unlimited access to the foundation of Amazon's business: its product database. Developers can grab product data, reformat it, add related services and use it to attract niche visitors to their own sites. These parallel Amazons may have added marketing features for niche customers which may be an otherwise expensive proposition if Amazon wanted to control its content and reach those segments. Imagine the innumerable variations necessary to be a global online retailer in several languages catering to multiple interests. By opening up its product database, Amazon is probably the first to practice the nearly-one-to-one marketing that may be feasible through the internet. For this access, Amazon demands that visitors to satellite sites complete purchases through Amazon (site owners receive a commission). Exposing the world's largest product database — along with editorial content and personalization functions — is a counterintuitive business strategy but one which may distinguish Amazon as the pioneer in collaborative innovation.

Amazon is an example of how foresight, investment in new tools and use of standards may converge to generate software that acts as a vehicle for interoperability for thousands of developers who may be using varying mark-up languages leading to a cacophony of impotent proprietary systems. Amazon's web server software mitigates this diversity by creating (API) interfaces that could retrieve product data and reformat it for select devices (PDA). Amazon's success is rooted in its human capital, that is, software engineers attuned to the emerging semantic web and W3C standards (XML, SOAP, OWL). Amazon, therefore, is a set of independent parts, including the database, shared interfaces for access and repackaging data for the site-specific layout preferred by a developer or site owner.

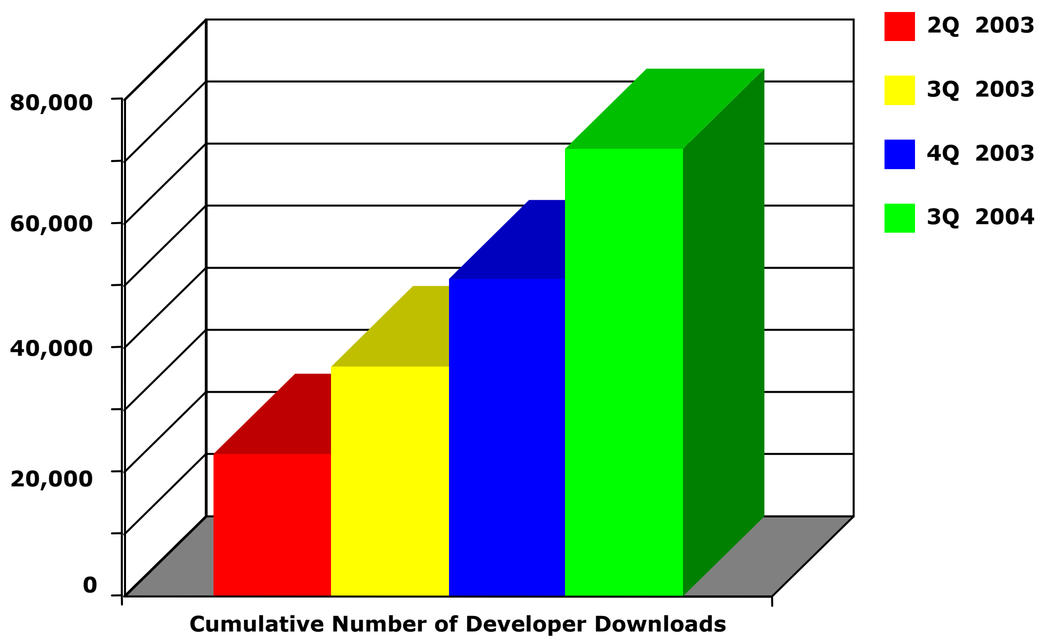


Illustration 2: Profit from Amazon: Give Away the Store (from MIT Technology Review, January 2005)

[c] Apple and the inevitable communication convergence

20 million visitors to Apple's site during October 2005 is not a signal for revival of Mac buying craze, if ever there was one. Doubling of unique visitors since October 2004 suggest that the lure of the Apple may be in the flavour of the iPod *du jour*. Hence, the interface innovator (Apple) is blurring the compartmentalization along industry lines in ways that earlier device manufacturers (Sony, Phillips) failed to penetrate. Neither did Apple stop at providing the receptacles for music, video and movies (iPod) but has organized services (iTunes, iMovies) that shall continue to prod Mr Jobs to make frequent visits to the bank, online, of course. In other words, it is a rehash of the old system where you get the telephone for free but pay for services or the reason why Xerox is eager to offer you a discount on the purchase of a photocopier only to sell you low cost products with higher profit margins (ink cartridge, paper).

The serendipitous innovator in Mr Steve Jobs has extracted the service model through iTunes and may be poised to compete with the likes of Nokia or NTT DoCoMo. Skype-like features may soon be a standard in IPv6 enabled iPod with built-in software defined radio (SDR), 802.11b/a/g (WiFi), 802.15.4 (ZigBee) and 802.16a (WiMax) features but aesthetically engineered to expose the crème de la crème of human-machine interface that is central to Apple's innovation. Wave your iPod.femto at Macy's to compare the price of the "collezione" charcoal grey cashmere scarf you saw at Tie Rack, buy petrol and pay for groceries at Tesco or pause "Last Tango in Paris" if Mum is Skype-in mode. The iPod is imagination beyond the obvious (for example, NTT DoCoMo service for RFID tagged objects).

The scenario above is as incredible as the time you found a 12 megapixel digital camera under your Christmas Tree last year (movie screen resolution is achieved with only 8 megapixels). Exponential growth of megapixels after CCD commoditization made that possible and feasible. Pervasive cell broadband with WiMax (802.16) and the inevitable merger of GSM with 3G may aid infrastructure interoperability. Location awareness (GPS, RTLS, P2P), identity of goods (UWB, RFID), multilingual speech and biomedical monitoring (nano-sensors) may deliver greater value to users if enterprise-wide ontologies facilitate contextual understanding. Object to object (O2O) communication is inevitable with the adoption of Internet Protocol version 6 (IPv6) and increasing bandwidth between components (multi-core microprocessors and memory) as well as precipitous drop in processing cost (for example, Sony's PSP3 priced around $400 offers a 8-core processor co-developed with IBM that operates at 2 teraflops).

The iPod of the near future may hold more than 40 petabytes (40,000,000 gigabytes). But 10,000 GB is sufficient for 20 million books in the US Library of Congress. With the remaining storage capacity why not download all of the nearly 500,000 movies currently available! It will take more than 100 years to watch all the movies if one started viewing the minute one was born and never slept for the next 100 years. If that is not your definition of "living" yet you have iPod-esque power in your palm. What are the choices? The ability to communicate with objects through the internet of things is one front runner. Internet protocol version 4 (IPv4 or TCP/IP circa 1980) that assigns IP addresses to each object (for example, a laptop) is limited by its current structure to "name" only about 4.3 billion objects. The next generation IPv5 (Internet Stream Protocol or ST2 from 1990s) was skipped to arrive at IPv6. IPv6 partly ameliorates this deficiency by offering 3.4×10^{38} unique addresses. In other words, every woman, man and child in the world (~ 6.5 billion) may claim more than 10^{28} unique IP addresses for their personal objects of use (shoe, toothbrush, towel, night light switch). Thus, the iPod of the future may be your personal global control center with the ability to communicate with your Mom, pay bills through the online bank and open the garage door or start cooking your pizza in the microwave oven. The irony is that it is not inconceivable that this is possible but it is truly incredible that these scenarios represent a march of reasonable convergence of innovation at hand. To see a world in a grain of sand and hold infinity in the palm of your hand is not only the innovation of iPod in action but poetry (of William Blake) in motion. Locomotion may not be far from the iPod (**illustration 3**) as it evolves as an universal mobility platform and the "brain-key" of your automobiles.

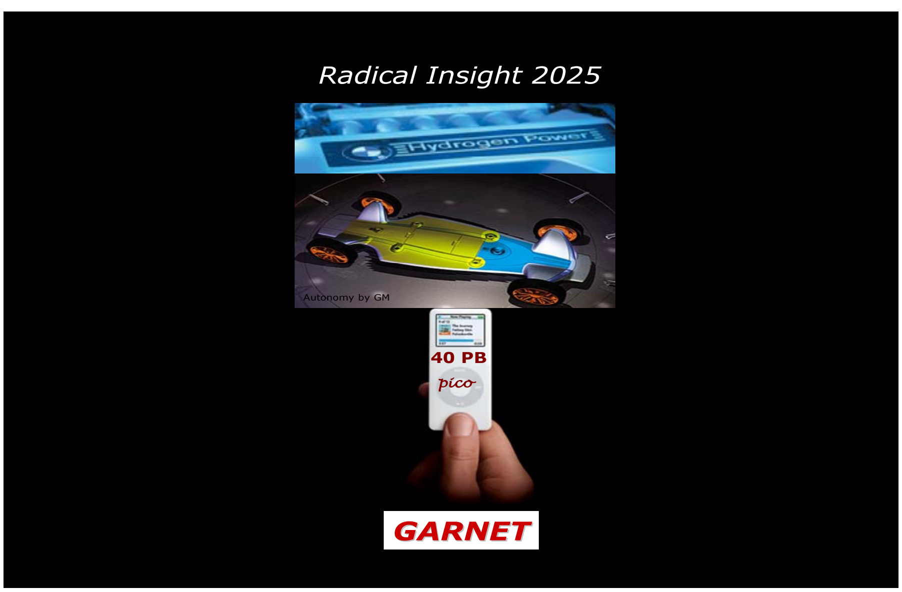
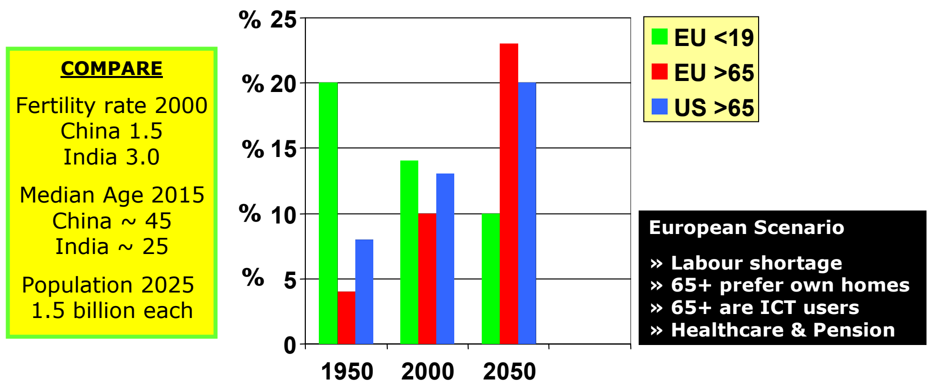


Illustration 3: Radical Insight ?

[d] Why American Greetings and Target are indicators of demographic change

With nearly equal number of unique visitors clicking on Target and Wal-Mart during October 2005 (illustration 1), it is interesting to note that the number represents a 70% increase for Target and 20% increase for Wal-Mart. The visitors to Target may prefer better quality or wish to support enterprises with progressive corporate policies over the sole criterion of rock-bottom prices. Target shoppers may be clustered in geographies with higher broadband penetration (urban or metropolitan zones) compared to Wal-Mart locations in non-urban or rural markets (USA).



Europe	Population (millions)	Under 19 year (millions)	Over 65 year (millions)	Over 65 year % population
1950	350	70	14	4.0
2000	450	60	40	8.9
2050	400	40	90	22.5

Illustration 4: Is Age an Asset ?

The demographic spread of the point-and-click shoppers are rapidly changing as is the profile of video gamers. Semi-retired septuagenarians are increasingly the partners of teenagers and grand-children in networked video games. The aging baby-boomers are mentally agile, physically active and pro-ICT. It is this change that shall create new markets for products and services (for example, books on iPod) in a manner to be driven by indicators outlined in **illustration 4**. Healthcare is a prominently application domain [2] and it appears that some countries are leading the charge with Japan as a front-runner.

The suggestion [1, 2] to use nanosensors in healthcare "sense and then respond" mechanism may have found an outlet through Synclayer, a Nagoya based cable TV & LAN integrator. According to the Economist of 3 December 2005, it has developed a means for the elderly living in their own homes to use a device that takes basic medical measurements (such as, blood pressure) and transmit them to a local health database. Synclayer also makes a sensor that can be placed, for example, on the refrigerator door, to send a message when ever the door is opened.

The latter is similar to an Intel initiative (in Oregon) which uses strategically placed RFID tags on kitchen cabinets and other areas to monitor eating or other activities of Alzheimer's patients who prefer to stay in their own homes. In Japan, with over 25,000 people over the age of 100 and another 30 million (25% of population) waiting in the wings to join the over 65 club by 2015, the "grey market" offers attractive business opportunities, such as, for Synclayer. Importing "gaijin" nurses may not be a long term solution in an immigration averse political climate. However, in 2004, the Government of Japan issued 80,000 visas for "entertainment" purposes to young Filipinas.

Suitable high tech response to the needs of Japanese elderly are evolving from the research on anthropomorphic robots developed for service. Already precursors are available in the market, such as the Snuggling Ifbot and Primo Puel, an interactive doll, which has become an unexpected hit with elderly single women although it was designed for boy-friend-less young girls. Equally appealing are the common sense solutions pioneered by the appliance maker Zojirushi, a brand trusted by millions of Japanese for rice cookers and electric kettles. iPot, developed by Zojirushi in collaboration with NTT DoCoMo and Fujitsu, has a wireless device which transmits a message to a NTT server each time the water-dispensing button is pressed. Then, twice daily, the usage record is sent to designated mobile phone or email address of family, friend or health care provider.

Elderly in Japan and in other affluent nations are likely to live healthier lives and may spend more on services than on goods. The average Japanese in their 60's has net assets of ¥21 million (USD 200,000). Although US citizens save far less, if the current demographic trends continue, the average US citizen will be twice as rich as a French or German in 20 years, according to Jean-Philippe Cotis (Chief Economist, Organization for Economic Cooperation and Development, 07 February 2006). The affluent elderly are likely to seek lifestyle services and healthcare in a manner that can be only provided through interoperable collaborative service innovation. The implication for e-business or any business is that making targeted products and offering suitable collage of services for an ever expanding 'niche' is one mechanism for businesses to grow profitably by taking advantage of interoperability.

Is auto "mobile" platform an innovation down the toilet ?

In 2004, losses of €2 billion ($2.5billion) in cars for Fiat (passenger cars account for under half the company's €49 billion in sales) dragged the whole Fiat Group into a €1.6 billion loss. In 2005, the car division of Fiat may report an operating loss of around €360 million (but the Fiat Group is expected to make a pre-tax loss of about €129 million buoyed by profits from its truck-tractor divisions). In October 2005, Sergio Marchionne, the Canadian-Italian who is the new chief executive of Fiat Group, reported a 70% fall in the group's third quarter loss (compared with a year earlier). With $2 billion from GM (penalty paid to Fiat in 2005 to scrap the GM-Fiat alliance) and the conversion of €3 billion-worth of maturing bank loans into equity, Mr Marchionne is trying to rebuild Fiat's passenger automobile business, panel by panel. Instead of closing plants, Mr Marchionne is taking advantage of the surplus workers parked in state-run unemployment schemes to engineer an investment of €10 billion over the next 4 years to bring out 20 new Fiat passenger car models, reports The Economist (3rd December 2005). Risk pooling and collaboration are the drivers. Future Alfa Romeo's (Alfa 159) will be made from the same basic platform, to reap economies of scale. Maserati will be repositioned at the top end of the Alfa range in the spirit of risk pooling to share components with the Alfa platform. The spirit of collaboration is in high gear at Fiat with three alliances with PSA Peugeot, a new venture with Ford to build cars in Poland, a licensing deal with Suzuki and talks of cooperation with China's Shanghai Automotive. Fiat may introduce even more changes in Europe's fast lane by building cars in collaboration with India's Tata not only in India but also extending the Fiat-Tata collaboration to build automobiles in South-East Asia as well as Europe.

Should we be impressed with this story from The Economist of 3rd December 2005 outlining the traditional risk-averse re-engineering of Fiat's passenger car division through Method Marchionne? It may be what the strategic gurus at Harvard Business School may pontificate or the bean counters or McKinsey may prescribe. It may be worth a reminder that according to Clayton Christensen of Harvard Business School, McKinsey "is able to crank out high-quality work year after year because its core capabilities are rooted in its processes and values rather than in its resources (vision). These capabilities of McKinsey also constitute its disabilities. The rigorously analytical, data-driven processes that help it create value for its clients in existing, relatively stable markets render it much less capable in technology markets" (page 168-169 in *The Innovators Dilemma*). Critics might argue that getting out of the 'red' is more urgent than exploring new vistas for Fiat. The critics may be right, too, in the short term. But will that be enough to fuel the flame for potential car buyers to choose Fiat because the company is back in the black?

Progress of classical automobile engineering (double wishbone suspension, antilock brakes) is slow and increasingly stirs less enthusiasm from buyers. Supply chain optimization and modular engineering doesn't influence Joe Fox. Electronics on the other hand is a high "clockspeed" subset of the automobile industry with remarkably impressive marketing impact. Thus we have witnessed the evolution of NorthStar and similar extensions of GPS in addition to NeverLost or other road navigation aids. Voice-dialing is standard in a few luxury sedans and tyre pressure may be monitored and projected on a dashboard screen. GE's VeriWise system successfully tracks equipment and vehicles with the aid of onboard monitors, GPS, RFID and other technologies to deliver value (Appendix A). Music in the car is pre-recorded in some format (CD) or radio-station controlled. Although Norway first implemented use of RFID tags for toll collection (in 1987), not much has changed in the use of transponders for such purposes.

Technology appears to be a driver for automobile sales. A multitude of vendors are creating products or services and diagnostics, with limited interoperability. It is divorced from the use of a common platform that can aggregate such services for the consumer. You cannot Skype-out or have access to email (if you need to). If you are 65 and recovering from a myocardial infarction, perhaps your physician would like to keep an eye on you while you are driving. Your onboard diagnostic LCD panel may alert you to decreasing level of fuel but can it guide you to a nearby petrol station? Your authorized Lexus dealer can call you to alert you for service but has no clue whether you have driven 3000 miles or not since your last oil change.

The innovation in consumer vehicles and automobiles that may bundle dispersed services is likely to be catalyzed, as a first step, simply by the introduction of a platform. It could be as unimaginative as the equivalent of a laptop PC in the car with the mouse and menu driven features that PC users already know how to use. If I have music downloaded on my iPod and I wish to hear it on the speakers, the USB port of the iPod plugs into the PC, rather a separate car-iPod gizmo. If the pressure gauge reading is alarming, then an auto-Agent locates the nearby service station and offers direction. You can Skye-out if you are trapped in a ditch or Google the license plate of the car in your next lane to get some coordinates of the glorious being driving the convertible Mini Cooper (juvenile but fun)!

Are you impressed with my unimaginative suggestion to get a PC to aggregate auto products and services? Is that innovative? Note: innovation may also arise from using common things. Think different! Think disruptive! To think about the next innovation in the automobile industry let us take a lesson from bathroom fixtures by TOTO. What TOTO has done for the everyday toilet offers a template for consideration by the auto industry. It is true that some of the features are available in commercial systems such as GE's VeriWise but it still seems very far from my car!

Illustration 5: Innovation Down the Toilet ?

The porcelain toilet bowl (top right hand corner in **illustration 5**) may not have changed over the centuries but TOTO of Japan has certainly created an intrigue for its use. The dashboard (bottom left) for the toilet bowl may be mistaken for a modern car control panel with its shiny knobs, buttons, switches, USB port, LEDs, radio, LCD panel, digital clock and several other functions inscribed in Japanese. Could I read my e-newspaper on the LCD panel while on the throne? It offers a whole new dimension to reading in the toilet, doesn't it? However, I had to request the cheerful help of the Duty Manager at the Sheraton Taipei in Taiwan to enable me to operate the controls.

TOTO has created an aggregate platform that not only flushes the toilet on command but offers the potential to accomplish a slew of chores and can act as a conduit for services. The automobile in the affluent world is emerging as an extension of our home or office. We already know how to turn on the dishwasher from the car or read the electronic grocery list [1] on the refrigerator before arriving in the grocery store pick-up bay on the way to home. What mobility needs is not a separate set of rules and multiple handlers (point of contacts) but an "organizer" that acts as a central clearinghouse. The automobile industry may need a platform aggregator for service providers to converge and product manufacturers to integrate. Auto manufacturers can upgrade the fuel injection algorithms through software downloads to the PC platform or you can insert the memory stick in your PC and hit "install" to get the job done. Send your blood pressure reading to the physician without getting out of the car. Evolution of the e-business model in the automobile industry for services, diagnostics, maintenance and new product marketing is in need of innovation and we need to look no further than the toilet to be inspired to innovate through convergence. The success of TOTO toilets has improved its brand recognition to the extent that the soon to open luxury hotel (Mandarin Oriental) in New York City is advertising TOTO toilets as a part of its claim to luxury. TOTO may be a "how to" guide for creating new markets for old products and ordinary services coupled to trigger innovation.

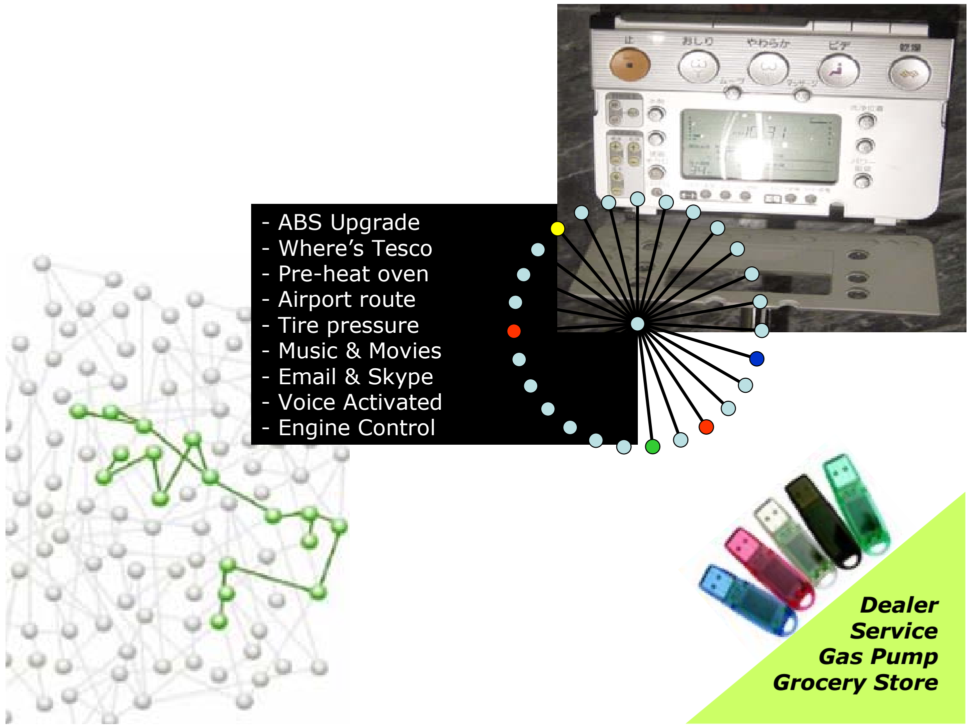


Illustration 6: Paradigm Shift in Interoperability ?

Is interoperability a catalyst for change or is change a pre-requisite for interoperability ?

Interoperability may help the onset of pervasive computing with significant benefits for the commercial sector by reducing information asymmetry. The $600 billion savings opportunity cryptic in the logistics operation mentioned earlier, is one example. However, software peddlers eager to sell products are at the heart of the cycle of problems that lead to further inefficiencies through lack of interoperability stemming from proprietary systems. While the latter is a strategy for software vendors to lock-in customers, the inability of users to successfully challenge such efforts are not primarily technological, but sociological. In the narrowest sense, if we consider innovation to promote interoperability, it is possible and feasible. But the movement necessary for adopting such innovation will be minimal unless innovation is connected to change (hence, our cultural heritage). Without change, innovation belongs to the problem set rather than the solution set.

Critics may cite the scholastic research of William Easterly in the *Elusive Quest for Growth* and vociferously argue that social or cultural heritage are lesser impediments than economic incentives. As evidence, critics may choose to point out the quantum leap of the open-source world wide web browser demonstrated in 1990 by Tim Berners-Lee of MIT (while still at CERN, Geneva) and compare it to the pathetic rate of progress of Linux (first entirely GNU operating system) released by Linus Torvald in 1991. The world wide web consortium (W3C) standards (HTTP, HTML, XML, RDF, OWL) provides a ***platform*** for vendors to build for-profit services and hence the incentive for adoption of world wide web consortium (W3C) collection of standards.

In sharp contrast, Linux OS is a "free" operating system if one can use it but the marketing avenue for Linux OS is feeble given that the sale of Linux OS is not aimed at creating MicroLinuxSoft. In addition, the services that most users expect to use (word processing, database, spreadsheet, powerpoint) are stunted in their development because developers have almost no incentive (except altruism) to create products that cannot be used unless the platform on which the product is based (Linux OS) has reached sufficient penetration to create market demand. Thus, it reasons to forward the opinion that economic incentives are the true catalysts for growth from innovation.

Economic incentives for interoperability are necessary to stimulate the innovative forces latent in ubiquitous computing. We have talked about ubiquitous computing for decades but the conventional reason for its sluggish growth may be readily attributed to the "dead weight of old technology" and archaic forms of educational thought as well as the resistance to change of (human) habits. However, it is more than likely that lack of proper economic incentives has paralyzed the progress of ubiquitous computing. Ubiquity of computing cannot reach pervasive status unless systems are interoperable enough to exchange, understand, compute and distribute information. Such architecture calls for standards of interoperability that are globally acknowledged. For the past 15 years, the open-source Linux operating system has languished while the Win-Tel duopoly operates as the dominant monopoly.

It is remarkable that the new emerging economies (but not the Western nations) are pushing the envelope through their practise of some forms of ubiquitous computing, albeit, limited in scope, in select areas and between pre-organized users, using select open source platforms that are largely incompatible with the mainstream. Leaving aside the socio-economic parameters, in general, widespread practice of ubiquitous computing requires ubiquitous programming. Ubiquitous programming is of necessity popular programming. It is disquieting that the populace in industrialized nations is not yet with the program. While ICT affinity among the elderly is gaining, it remains a fact that a significant percentage of people above a certain age rely on their children to program their VCRs while arm-chair computer science policy wonks continue to pontificate about the magic bullet that will cause people to script correct programs without having to think. Instead, we need policies to enable how to teach people (students) how to think, provide tangible building blocks (for example, for children, Lego MindStorm) to create *ad hoc* programs and learn how to ask the right questions to better utilize the benefits of cheap processing power of microprocessors.

The role or question of change repeatedly alluded to (above) is difficult to gauge, initiate, manage and measure. But change happens. Several observations discussed here provide some evidence. For example, economies that were teetering on the verge of hyperinflation only 15 years ago are now preparing to adopt the Euro. A decade ago, simply getting a truck to Poland may have involved a weeklong wait at the congested and corrupt border crossing. But that view of Eastern Europe ("it is like Africa but closer") is changing. Supply chain adaptability is a driver for increasing interest in Eastern Europe despite lower cost of operation in China, India or South-East Asia. Zara is changing time to market through its "fast fashion" trend. eBay's purchase of Skype for $2.6 billion bodes well for its Scandinavian founders who used programmers from Tallinn, Estonia rather than California's Silicon Valley or Bangalore, India. Skype employs 130 highly paid under-thirties mostly (four fifths) from Estonia. On the other side of Tallinn is Elcoteq, makers of mobile phone handsets for Nokia and other behemoths. Elcoteq's middle-age workforce of about 3000 performs repetitive, semi-skilled tasks, is modestly paid and entirely local. However, increasing prosperity may force changes. By 2010, it may be necessary for Elcoteq to switch production to Russia to reap similar cost benefits unless political risks and corrupt practices continue to threaten businesses in Russia.

Hence, lack of change in bureaucracy is the single most overwhelming deterrent for locating operations in Eastern Europe despite its crucial geographic proximity to affluent European nations. For example, Sciant, a Bulgarian software company, is opening offices in Vietnam plagued by administrative burdens in Bulgaria. Sciant employs three full-time employees necessary to deal with arcane record-keeping requirements for the state and getting the right stamps on the right paper for the sleepy customs service which takes two days or longer to release an incoming shipment. “I have a friend whose things were stuck in customs for four months” comments Steve Keil, chief executive of Sciant in Bulgaria. How can interoperability help without changes in the legacy of bureaucracy?

Thus, process inefficiency appears to be a severe detriment in the path of progress. Process is also at the heart of technological change or for technological changes to produce the necessary benefits. The recent interest in RFID provides a suitable case to explore. It is clear that the benefit from RFID is still far from systemic because process innovation necessary for RFID deployment is still immature. Whether RFID is a disruptive technology or not is still subject to debate but it bears reminding that RFID technology has been in existence since 1944.

RFID proponents may wish to learn about general process technologies (GPT) from the case study of the electric dynamo as discussed by Paul David and Gavin Wright in *The Economic Future in Historical Perspective* (Oxford University Press, 2003). It appears that Clayton Christensen’s use of the term “disruptive technology” may be a modern version of GPT.

Paul David and Gavin Wright reveals in their scholastic paper that the slow pace of electrification prior to the 1920’s was attributable largely to the lack of profitability of replacing still serviceable manufacturing plants adapted to the regime of mechanical power derived from water and steam. Prior to 1920, the group drive system of within-plant power transmission caused electric motors to turn separate shafting sections, so that each motor drove related groups of machines. Primary electric motors (the new technology) were merely added to the existing stock of equipment. It is this latter *modus operandi* that we see with RFID. Tags merely attached to objects in the name of improvement. With the favorable investment climate of the 1920’s, firms had the opportunity to switch from group drive to unit drive transmission, where individual electric motors were used to run machines and tools. Advantages of the unit drive extended well beyond savings in fuel and energy efficiency. They made possible single-story, linear factories with reconfigured machine placement, permitting flow of materials through the plant that was rapid and reliable. Rearrangement of the factory contributed to cost savings in materials handling operations, serializing machines and thereby reducing or eliminating back-tracking. Thus, process changes in the entire operation of the factory occurred as a result of electrification. The ability to harness accurate data through automatic identification holds the potential for sweeping changes (reminiscent of the age of electrification) but such changes call for convergence of process and technology with interoperable systems to reap the benefits from data.

Can standards drive interoperability ?

It is difficult to institutionalize change and thus the quest to bypass the quagmire in order to pursue interoperability. Ironically, a standard is impotent without adoption and resistance to change delays adoption of standards. More than a century later we still have to switch between 110 Volts and 220 Volts or frequently seek electrical adaptors when travelling between countries even on the same continent. There may never be any one standard for anything.

It is not inconceivable that in some cases standards may not be the standard solution (Appendix B). At the least, standardization is not the panacea that one may believe although it serves a fruitful reductionist approach. It reduces the chaos to an acceptable number of choices and offers a handful of mechanisms that may, through mapping functions, ensure 'connectivity' between standards or interoperability between a group of standards. In other words, it is better to be a part of an ecosystem of standards and optimize collaborative processes or systems interoperability based on such an ecosystem. It may be an useful exercise to appreciate the process of emergence of various standards through the lens of history. The excerpts below are from Tom Gibbs (Intel) address at GCI.

In 1855, Henry Bessemer established the metallurgical process which allowed the manufacture of high grade steel. Rapid improvements in rail technology were possible with steel but several forces held back the introduction of the railroad (analogous to the Luddites who opposed introduction of technology in the textile industry). The railroad dissenters came from a number of public avenues including canal owners who had only recently finished enormous investments to develop the Erie Canal in New York State (which arguably led to the creation of New York harbor as the port to the world). Canal owners of the day were quoted in the Boston Globe in the late 1820's that "there would as likely be a rail road to the moon as one that would link Boston to New York."

Finally, railroads were made possible by the use of two standards: rail track gauge and time zones. The evolution of rail transportation was largely funded by individual entrepreneurs and they attempted to compete by winning share with unique gauge which locked the other company out, in much the same manner that software vendors use proprietary practices, today. The issue with time zones stems from the reluctance of the general population to adopt a standard time citing the need to optimize local agriculture to the position of the sun.

Great Britain was the first country to adopt one standard time. William Hyde Wollaston (1766-1828) suggested the idea and it was popularized by Abraham Follett Osler (1808-1903). The first railway to adopt London time was the Great Western Railway in November 1840. On 22 September 1847, Railway Clearing House, an industry standards body, recommended that GMT be adopted at all stations. By 1855 the vast majority of public clocks in Great Britain were set to GMT. The last major holdout was the legal system, which stubbornly stuck to local time for many years, leading to oddities like polls opening at 08:13 and closing at 16:13. The legal system finally switched to GMT when the Statutes (Definition of Time) Act took effect through the Royal Assent on 2nd August 1880.

Standard time zones were instituted in US and Canada by the railroads on 18 November 1883. However, Detroit kept local time until 1900 when the City Council decreed that clocks should be put back twenty-eight minutes to Central Standard Time. Half the city obeyed, half refused. After considerable debate, Central time was adopted by city vote in 1905. Standard time zones were established in the US by the Standard Time Act of 19 March 1918. It is interesting to note that standardization of time in specific time zones in US and UK required 30 years or more for adoption. A study conducted by Norman Poire also reveals (see illustration 7, below) that it takes about 30 years for new technologies to be adopted. Why? One reason may be that it takes an active generation to retire before the younger generation can adopt the advances from and reap the benefits of innovation.

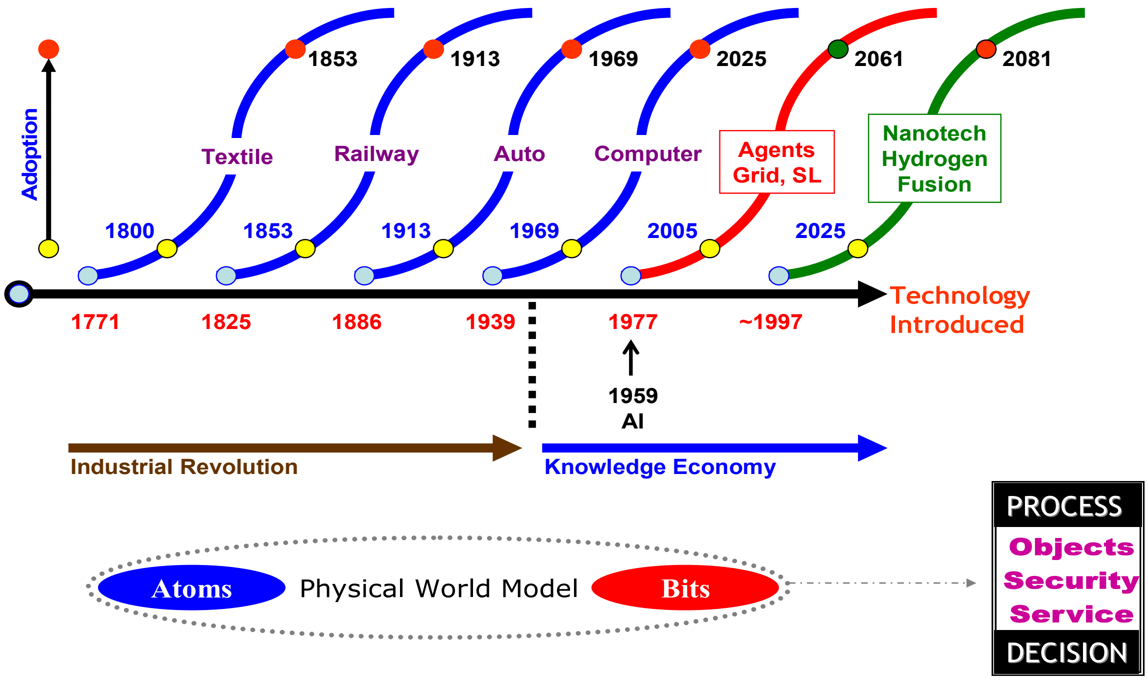


Illustration 7: Why does it take three decades for adoption?

Concluding Comments

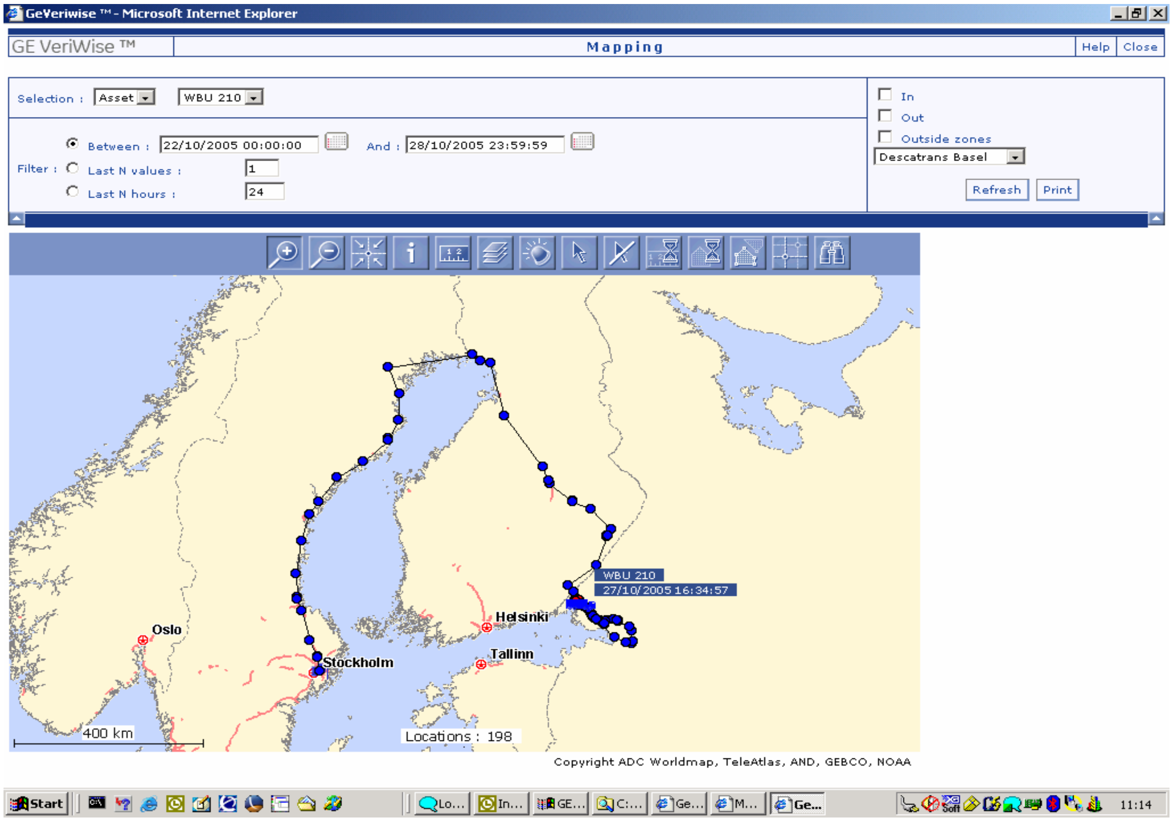
Scores of PC's linked together can outperform a giant mainframe, argued German Chancellor Angela Merkel in her first speech to the Bundestag on 30th November 2005. That was the image this former physicist chose to describe a fragmented programme that promises many little steps in the general direction of reform, rather than one big leap. Innovation is a collective process that stems from confluence of ideas, concepts, tools and technologies rather than one Big Bang. Distributed computing, alluded to by Chancellor Merkel, is a rudimentary extension of our primitive understanding of the memory and visual system of Octopus that leads to high maneuverability of its arms and the capacity of its peripheral nervous system to perceive and process chemical and tactile information. Coordinated propagation of the bend and neuronal activity is achieved by local feedback from propreioceptors in its muscles. The stomata in plant cells also offer an advanced system for distributed computing. The ability of ants to improve routing algorithms is well documented. In other words, the Octopus, stomata and ants are coordinated in their local and global optimization. Decision systems find it difficult to achieve such harmony and perpetuate inefficiencies. Hence, decision systems and humans involved in their design might find it stimulating to explore biological systems in search of innovation and interoperability which is key to global businesses as well as Charlie's Chocolate Factory.

References

1. **Datta, S.** Adopting Decisions, Optimizing Facts and Predicting Figures http://supplychain.mit.edu/shoumen.htm and http://esd.mit.edu/WPS/esd-wp-2006-11.pdf
2. **Datta, S.** Advances in SCM Decision Support Systems: Semantic Interoperability between Systems. Working Paper. Engineering Systems Division, MIT http://esd.mit.edu/WPS/esd-wp-2006-10.pdf

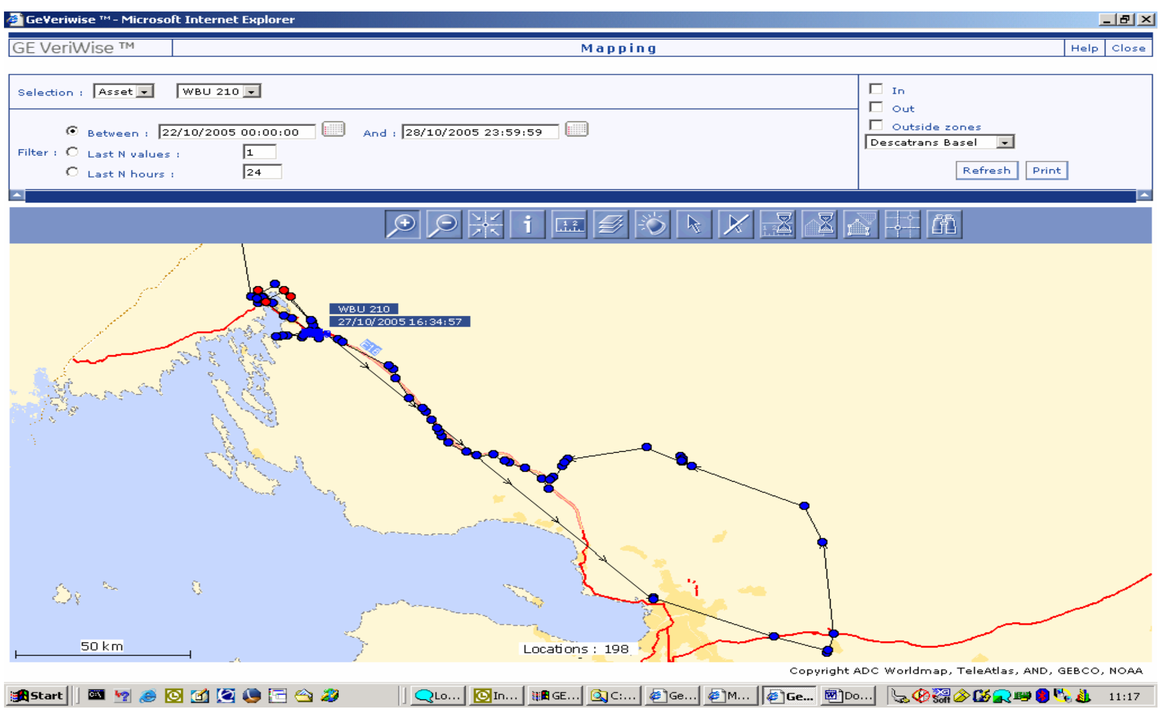
Appendix A: **GE VeriWise Real-Time Tracking Data for Recovery of Stolen Trailer**

Trailer Movement: Stockholm to St Petersburg via Finland (Data from Kimmo Lindqvist, GE)



GE VeriWise Real-Time Tracking Data

Trailer Movement: St Petersburg, Russia back to Finland (Data from Kimmo Lindqvist, GE)



Courtesy of Mr Kimmo Lindqvist, Mr Willem Duijf (GE Europe) and Dr Joseph Salvo (GE Global Research, NY)

GE VeriWise TRAILER RECOVERY LOG

08.10 Info about stolen trailer received.
Trailer stolen from Drivhjulsvägen, Stockholm at 17.00 22nd October 2005.
Trailer went through northern Sweden to Finland.
Crossed border between Finland and Russia at about 14.00 24th October 2005.
Trailer location 31 km east of St. Petersburg at this moment.

08.15 Contact with Russian police in St. Petersburg.

08.20 Contact with Matti Löfman in TIP Finland who knows Russian.
Matti will contact his contact at Finnish police crime division.

08.35 Guest login created on GE VeriWise portal for police.
08.35 Trailer still at same location.
08.50 Received info of what happened from J. Malmquist (trailer owner)
09.20 Polling not working, Spacechecker to fix it in 5 minutes.
09.30 Info to W. Duijf (GE) polling should be ok
09.47 Poll: trailer still moving north, now 39 km east of St. Petersburg.

09.53 Poll: no reply from trailer

09.57 Matti on his way to the Finnish police

10.05 Poll: no reply from trailer

10.07 Got contact with Finnish border at Nuijamaa (trailer crossed the border).
Finnish border response: cannot control all traffic (how can we know before it comes to the border).
Obtained phone numbers (to call if trailers need to be stopped at border).

10.10 Polling not working. Willem Duijf checked with Spacechecker.

10.14 Poll: no reply from trailer

10.19 Poll: no reply from trailer

10.20 Tracking mode set up by Spacechecker and report every 2 min

10.23 Matti at Finnish police: info about trailer with chassis number and description of vehicle

10.28 Poll: no reply from trailer

10.35 Trailer report, standing still untethered on battery mode at 42 km north of St. Petersburg

10.39 Poll: Still at same location

10.43 Contact with Matti (filing criminal records with Finnish police to send info to Russia

10.49 Poll: no reply from trailer
10.50 Given address from where trailer was stolen in Sweden

11.05 Matti informed me to contact the Swedish police regarding Schengen border controls

11.07 Information given to Swedish police

11.11 Poll: no reply from trailer

11.12 Russian police urged to take action. Details sent from Finnish police

11.29 Poll: no reply from trailer

11.35 Finnish police requests urgent action from Russian central militia in Moscow.
11.44 Finnish police live with Russian police through GE VeriWise 'POLICE' page.

12.18 Tracking mode running reports every 3rd minute since 11.14 visible.
Last report from 11.38, trailer still on same location.

12.20 Trailer moving again. New location 50 km north of St. Petersburg.
Tracking mode fully operational through Spacechecker, update every minute.

12.24 Trailer soon on the highway E18 between Finnish border and St. Petersburg

12.35 M. Schnatmeier confirms modem configured over air to send reports, even on battery mode

12.35 Trailer close to E18 stopped

12.36 Gave the same info to Matti

12.42 Trailer on battery mode, going south towards St. Petersburg 5 km from latest position

12.45 Trailer on same position again as 12.29, 12.35 (GPS glitch?)

12.58 Trailer driving northward at or beside E18

13.03 Gave info to Finnish border control to be ready if trailer tries to cross the border to Finland.
I am in contact with Finnish border by mobile phone.

13.10 Trailer still on its way towards Finnish border

13.15 Willem Duijf has created a zone alarm for Finnish border

13.18 Finnish border control tracks trailer on GE VeriWise 'POLICE' page

13.35 Matti requests info: time when trailer crossed Sweden to Finland at Nuijamaa
No cameras at Nuijamaa border but they have cameras on the border between Sweden and Finland at Tornio.

13.38 Matti received time stamp info. Possibility that trailer and truck has been caught on camera.

13.39 Trailer closing in on the Finnish border (50 km from Finnish border)

13.50 Trailer seems to be heading towards Nuijamaa: border control alerted

14.02 Trailer seems to take the same route out of Russia as it did when it came to Russia

14.20 Trailer stops outside Vyborg

14.35 Trailer still on same location (untethered, on battery mode)

15.03 Russian police has located trailer.
Trailer is standing without license plates. Russian police investigating.
Finnish border control informed. We have recovered the trailer with help from GE VeriWise.

-----Original Message-----
From: Lindqvist, Kimmo
Sent: Monday, February 13, 2006 2:44 AM
To: shoumen@mit.edu

After that day, I found out that the police in Finland found the picture from the border between Haparanda / Tornio (Sweden/Finland) of the stolen trailer. Finnish police sent a formal request to the Russian police to investigate how it was possible for a stolen trailer without documents to cross the border. The Russian border control did not find any papers of that trailer crossing the border. Puzzle! Did Russian border control “allow” stolen trailers/vehicles without documents into Russia?

Appendix B: **Standards may not be the Standard Solution**

Interoperability in Global Track & Trace: How to use UWB and UWB+NB with SDR as an LPS Solution

Preface

The thinking in this note about track & trace technologies may bring to light the multiplicity of so-called standards available or introduced in this narrow space. Problems encountered due to globalisation result from the unlimited movement of objects (inanimate and animate objects such as humans) in domains where the standards stapled to the objects are not practised, accepted, adopted, implemented or enforced. Systems between domains severely lack communication skills due to lack of interoperability. We have an embarrassment of riches with respect to data while we continue in abject poverty of information due to lack of interoperability in this systems age. The latter has, erroneously, promoted even more calls for standards and even larger consortiums are being formed to muscle global adoption (acceptance). The success of this approach is open to question by the failure of Wal*Mart-esque efforts to usher in the global visibility that automatic identification was touted to deliver. The lesson from the age of introduction of the electric dynamo is ignored by the current drive to pursue that holy grail of one standard or the lowest common denominator. It is clear that global leverage may be used to promote interoperability between select few partially adopted standards in a way that foreign systems can interface seamlessly with another through translational mechanisms (obvious from human language). It is this theme that is referred to as being standards agnostic yet standards compliant. The example in this note draws from the recent hype of RFID but aims to show how technologies at hand are available to extend the debate beyond identification. The use of technology must be viewed as a tool to add value to a process or enable decisionable information.

Automated Identification with Global Interoperability: An Incomplete View and Some Imperfect Ideas

Most RFID types (125KHz, 13.56MHz, 433MHz, 915MHz) possess a spatial capacity of 1 kbpspm^2 (IEEE). Spatial capacity focuses not only on bit rates for data transfer but on bit rates available in confined spaces (grocery stores) defined by short transmission ranges. Measured in bits per second per square meter, spatial capacity is a gauge of "data intensity" that is analogous to lumens per square meter that determines the illumination intensity of a light source. Growing demand for greater wireless data capacity and crowding of regulated radio frequency (approved ISM spectra) will increasingly favour systems (spectrum) that offers appreciable bit rates and will function despite noise, multipath interference and corruption when concentrated in smaller physical areas (stores, warehouses). Will spatial capacity limitations clog the ‘interrogation’ system when item level tagging becomes a reality? Some are exploring BlueTooth with spatial capacity of 30 kbpspm^2 while asset management may use 802.11a protocol (5.15-5.35 GHz) with spatial capacity of 55 kbpspm^2 (spare parts in an air force base or inventory in Lockheed Martin).

Quite a few companies are exploring ultra wideband since its appearance on the scene in 1962. UWB spans several gigahertz of spectrum at very low power levels below the noise floor of existing signaling environment. The spatial capacity of UWB is 1000 kbpspm^2 or 1000-fold more than 802.11b. Conventional narrow band technology (802.11b, BlueTooth, 802.11a) rely on a base "carrier" wave that is modulated to embody a coded bit stream. Carrier waves are modified to incorporate digital data through amplitude, frequency or phase modulation. These mechanisms are, therefore, susceptible to interference and the coded bit stream (for example, electronic product code or EPC) could be decoded/intercepted (defense/security). UWB wireless technology uses no underlying carrier wave but modulate individual pulses either as bipolar or amplitude or pulse-position modulation (sends identical pulses but alters transmission timing). UWB offers pulse time of 300 picoseconds and covers a broad bandwidth extending to several gigahertz. UWB operates in picosecond bursts, hence, power requirements are far lower (200 mW) compared to 802.11b (500 mW) or 802.11a (2000 mW). Equally staggering is the data rate for UWB (0.1 – 1.0 $gbps^2$) when compared to 802.11b (0.006 $gbps^2$). Sony and Intel are leading this research for wireless transmission of data, video, networked games, toys and appliances (some of which are already in stores). Robotic vacuum cleaners and lawn mowers can clean the living room or manicured garden without touching the sofa or grazing by the rose bush. Universal appeal for UWB is latent in its capability to offer a global standard. Without FCC-like country-specific restrictions, an old technology like UWB still remains virgin for many possible applications and may be the only global wireless communication medium that may claim, someday, a truly global standard.

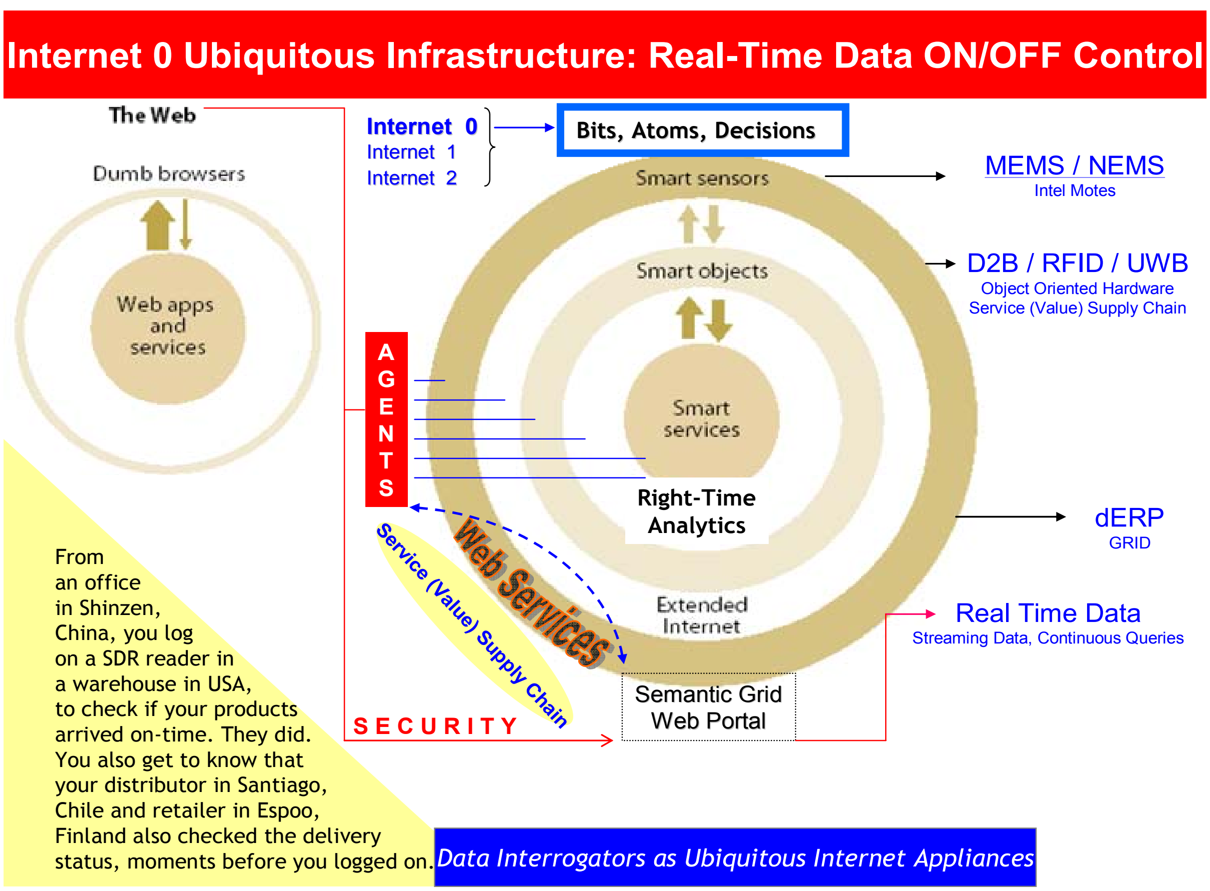
After September 11, UWB transmitters were mounted on robots for search missions at the World Trade Center since UWB is less hindered by metal (Coke cans or turbine spare parts) or concrete (buildings and warehouses). On 14 Feb 2002, the FCC gave qualified approval to use UWB in the range >960 MHz, 3.1-10.6 GHz and 22-29 GHz.

UWB-RFID active transponders are not prohibitive (cost) while transmitters may become cheaper than 802.11b RFID readers because they do not need many analog components to fix, send and receive specific frequencies. The combination of UWB plus narrowband technology to produce a passive UWB transponder may be a reality (by combining UWB communication with narrowband RFID tag). Combining a narrowband receiver and a wideband transmitter in the tag optimizes collecting RF energy on the receive channel combined with ultra low power on transmit channel. At the MAC layer, optimized conflict resolution algorithms allow multiple tags to communicate efficiently and effectively with the reader. Due this algorithm the channel is used very efficiently (OFDM), resulting in an increased effective bandwidth that allows more tags to communicate with the reader. Similar use of OFDM to enhance fixed frequency RFID readers is not known, to the best of my knowledge. Thus, utilization of narrow band downlink and wide band uplink communication enables wholly (passive) or partially battery-less tag designs to be manufactured at low cost. UWB communication is resilient to selective RF absorption, since the data can be recovered by the reader by relying on the message content in the 'not-absorbed' frequency bands. Due to the very broad frequency content of the transmitted UWB impulse, it is extremely resilient to path fading and enables readers to determine location of tags. Thus, not only can UWB tags identify but it can also locate (for example, movement of objects in a warehouse as well as the storage organization). Tags can be programmed to have 64, 96, or 128 bits (EPC, GTIN and other UPC-EAN numbering schemes) and are rewriteable. Unlike traditional passive RFID tags, passive UWB tags use the accumulated power to transmit UWB impulses to the reader.

Despite the clear advantages of passive UWB RFID tags, in general, the dispute in the field stems from claims that UWB transmission may interfere with spectrum used by cell phones and air traffic control. FCC is investigating but it is poised to open up even more of spectrum for UWB commercial applications. Without the burden of license fees for spectrum usage, the commercial floodgates for UWB usage may be unstoppable much to the chagrin of the telecom industry. MSSI is charting new territories and PulseLink has shown that SDR readers work with UWB chips.

The current thinking to use 'readers' specific to one or more RF modes may not be a sustainable approach for the infrastructure necessary for object identification to become pervasive. In 2004, heterodyne readers that can read MHF (13.56MHz) and UHF (902-956MHz) tags cost about USD5000 (ThingMagic). Consider commonly used frequencies, RFID vs UWB, passive vs active, many standards (EPC, GTIN, GTAG) and regional regulations (RF spectrum, emitted radiated power). Taken in combination, it spawns several types of transponders and to read the tags we will need a variety of readers. Multi-frequency tags may not stem the problem. Current readers cannot read UWB tags. According to the current model, businesses dealing with objects from global partners, therefore, must possess infrastructure (several types of readers) compatible to read a plethora of tags. Current reader vendors, present hype and lack of foresight, thus, may deliver a debilitating blow to the real benefit of object identification and sharing of data to improve processes (supply chain, readiness). Readers must be ubiquitous as a civil engineering infrastructure similar to electrical outlets, evolving to form the internet of devices (Interdev). Such a framework may make pervasive data acquisition and sharing a real possibility when viewed with the idea that open source software should also become a part of the IT infrastructure. Control, security, updates and hardware improvements are delivered via this ubiquitous software infrastructure.

It is this scenario (see illustration below) where the reader in the warehouse is always 'on' but the ability to read certain objects (or not) is controlled through the software layer by the authorized user and the authorizations allowed by the principal user. The 'views' of the contents of the warehouse is limited to goods that the user can 'read' by virtue of the preamble that must be exchanged and validated between the reader and tag (similar to current architecture that is embedded in EPC specifications and can be adopted elsewhere).



Software defined radio (SDR) is probably the simplest solution at hand to deliver this ubiquitous infrastructure in a manner that will remain transponder hardware agnostic with all modulations effected through the SDR OS. This view, that of, using SDR hardware (in some form) as ubiquitous RFID interrogators (in your refrigerator or in a warehouse) is the proposal based on the current understanding of SWR (software radio) and confirmation from its inventor (Vanu Bose). In 1991, the term "software defined radio" was coined to describe radio devices implemented in software and running on generic hardware. Because SDR is intrinsically linked to the future of global mobile telephony[30], an area of convergence between SWR infrastructure for real-time data and delivery of real-time data as a service, may evolve as a business for telecom providers to serve small and medium enterprises. In 2002, a relevant service model was explored by NTT (www.ntt-east.co.jp/tmmall/rf.html).

Convergence in our ability to communicate with objects through the internet of things is one front runner in the identification services business. Internet protocol version 4 (IPv4) that assigns IP addresses to each object (for example, a laptop) is limited by its current structure to "name" about 4.3 billion objects. The next generation IPv6 protocol ameliorates this deficiency by its ability to offer 3.4×10^{38} unique addresses. In other words, every woman, man and child in the world (~6.5 billion) can claim more than 10^{28} unique IP addresses for their personal objects of use (shoe, toothbrush, towel, night light switch). It might be interesting to note for those who wish to track and trace objects, humans, animals and processes that the 128-bit alphanumeric structure in some of the standards format (for example, EPC or electronic product code) may be synchronized through some of interoperability and/or translational mechanism with the 128-bit format proposed for IPv6.

Taken together with the issues of biometrics, it begins to point toward true global transparency, if necessary, to better align supply with demand for global commerce or determine non-obvious relationships to ensure security. It is difficult to conceive that heterogeneity and complexity of the 21st century may be served by conforming to one or a few standards. Processes that will stand the test of time and operations that will be sustainable may thrive on being standards compliant through mapping functions while rigorously maintaining its principle of being standards agnostic.

[30] Typically, software-defined radio (SDR) converts incoming analog radio signals to digital signals and then uses software (hence, software-defined) to sort through the frequency bands. Asad Abidi and his team at UCLA, modified conventional SDR by using an electronic component called wideband anti-aliasing (WAA). This enables access to a wide range of the spectrum (analog signals 800-5000 MHz) but can emphasize one preferred or designated band to be converted to a digital signal, using milliwatts of power. Because mobile devices are increasingly bundling more wireless capabilities (WiFi, GPS, RFID tag readers), these devices require multiple chips (one for cell phone, one for WiFi, one for GPS). SDR+WAA devices can tune into a wide range of frequencies with one chip yet selectively access any wireless service as well as open a car door or read a RFID tag, with the mobile telephone. However, such an "universal chip" would not only receive but also transmit across a wide range of frequencies. The latter is work-in-progress. Reference: 800Mhz to 5GHz Software-Defined Radio Receiver in 90nm CMOS by Bagheri, R. *et al* (2006)

APPENDIX 7: **WEB SERVICES**

Contributed by Dr Shang-Tae Yee, GM Research, General Motors (Detroit, Michigan)

Semantics are expected to add considerable value to web service composition. Examples of service composition include travel booking, car broker service and loan searching services. Web service composition consolidates appropriate web services in a manner that fulfills user request given by a high-level description. Applications for business-to-business (B2B), electronic commerce (e-commerce) and enterprise application integration (EAI) may be the primary beneficiary of web services. Currently, these services are accomplished using application programming interfaces (API) that are developed and deployed on-demand. When web pages change, the API must be modified accordingly [i]. This approach is ad-hoc, time-consuming, error-prone, requires low-level programming and expect users to possess skills to invoke and map web services to each other [ii].

A wide spectrum of embedded and mobile computers may usher an ubiquitous computing environment that may accelerate the demand for automatic web service composition and interoperability because users do not want to spend time searching multiple web service resources to obtain information. Automatic web service composition involves the automatic discovery, reasoning, selection and interoperability of related web services. Automated reasoning and selection filters out services based on evaluation criteria and then, chooses a composition plan. Interoperability ensures seamless exchange of information between services.

Automatic web service composition requires descriptive clarity. Web service description language (WSDL) has been widely used for defining web services. However, WSDL describes Web services from a syntactic point of view and does not support semantic description of services. Ontology is essential to provide 'semantic context-based' services. Semantic web services necessitate ontology standards, such as resource description framework (RDF) and DARPA agent markup language (DAML). RDF expresses classes, properties, ranges and documentation for resources. DAML was built on RDF and represents further relationships and/or properties. Two solution approaches were presented [iii] by Srivastava and Koehler: [a] using WSDL and business process execution language for web service (BPEL4WS) and [b] semantic web solution using RDF/DAML and Golog (logic programming language). In this space, agent technology can play an important role for coordinating the composition process from discovery, through reasoning, selection and interoperability by taking into account user constraints and preferences. In order to make a "meaningful" composition plan, web service composition process involves exchanging messages between services in which a message has one or more types of parameters. When a sender sends a message, its parameter types should be compatible to those received by a receiver. When multiple services are involved in this composing process, all the messages exchanged need to pass compatibility check for their parameter types after which associations between services are formed based on mutual relationships. It is necessary to build a structure, say a template, to systematically represent the relationships and in addition, to determine the value and soundness when a new service is added. The templates are stored for reuse (stored templates). Each time a new service is defined and discovered, its template is stored in a repository and updated. Also, the stored templates are used as a foundation to develop a knowledge base for capturing user requests.

A composition template may be developed to an agent template consisting of different types of agents (decision agents, information-gathering agent). A decision agent binds a web service(s) with customer requirements to support decision-making. Information-gathering agent connects to the appropriate web service(s) using reasoning and filtering logic. For a specific business, an agent-based composition template may be built within a semantic web service-enabled multi-agent architecture that captures user requirements, services, location, reasoning, filtering, binding, decision-making and delivery. A composition template proposed by Medjahed *et al* contains the general structure within a web service and is modeled by a directed graph consisting of a set of services and relationships between services (for example, the car broker template that relates credit history and driving record). Semantic web may potentially accelerate interoperability because it improves context-dependent communications. In a supply chain environment, an original equipment manufacturer (OEM) conducts multiple data translations to exchange information with suppliers because business partners use different data formats. An OEM uses numerous non-interoperable IT tools while suppliers use multiple customer-specific inventory software. Lack of interoperability calls for investment in development of new adapters and connectors. Inconsistency in nomenclature, terminologies, contents and proprietary protocols defined differently by each business type further amplifies these problems resulting in an intractable cacophony of different set of ICT tools used for quite similar purposes but in vain.

The Automotive Industry Action Group (AIAG) has been conducting a pilot project [iv] with the (US) National Institute of Standards and Technology (NIST) that intends to test automotive inventory visibility and interoperability (IV&I) scenarios using an interoperability test-bed, called ATHENA. It is a model-driven interoperability tool set based on business process modelling, ontology and web service data exchange. Electronic kanban (eKanban) is used for exchange of business data (OEM and supplier). eKanban material replenishment protocol consists of four signals: consumed, authorized, shipped and received. These signals allow suppliers visibility of the customers' inventory level in order to determine replenishment needs and initiate material shipment [v]. A reference ontology module was developed to capture the common terms and names used in material replenishment processes. When an OEM wants to send XML data to a supplier, it is converted to RDF format and open application group (OAG) business object documents (BOD) are generated after binding the RDF data with the reference ontology module. Then the BOD is sent to the supplier through a secure web service messaging protocol.

---

[i] McIlraith, S., Son, T., and Zeng, H. (2001) Semantic Web Services. IEEE INTELLIGENT SYSTEMS

[ii] Medjahed, B., Bouguettaya, A., and Elmagarmid, A. (2003) Composing Web services on the Semantic Web. The VLDB Journal

[iii] Srivastava, B. and Koehler, J. (2006) Web Service Composition - Current Solutions and Open Problems.

[iv] Automotive Industry Action Group (2005) Advanced Technologies for Interoperability of Heterogeneous Enterprise Networks and their Applications.

[v] For an operational environment, an agent can model a business entity performing a particular business function, e.g., material shipment specifically as process function for loading a truck. When every business entity has a linkage with its agent, a distributed information processing and decision-making structure can evolve. It may coordinate and make decisions in a collaborative manner. Agents can receive RFID or other real-time data and "publish" to other agents. Thus, a multi-agent system may provide information from a specific business function to a back-end system or feed other multi-agent systems deployed in different business sectors or layers. Multi-agent technology is a potentially powerful tool for real-time information exchange and decision-making in operational environments.

Disclaimer

Observations mentioned in Appendix 6 are drawn from sources in the public domain including The Economist of 3rd December 2005. This article is an analysis of the "art of the possible" based on trends or common observations. It attempts to be several "things" to serve several purposes. The attempt may fall short on many counts and it is the fault of the author alone. The article is deliberately simplified in its exploration of incomplete ideas and plagued with unfounded extrapolations from these observations. All errors of content or coherence are due to the author. This section is geared for general readers interested in global interoperability issues. It may be unsatisfactory for experts and practitioners. However, the amalgam of ideas may spark new thinking. This is a casual exploration to catalyse convergence and innovation to bolster interoperability among common elements. The author has freely used several sources of information to 'connect the dots' and offer clues. This is not an original research or marketing study and neither is due to the author. Opinions and comments expressed here are attributable solely to the author and do not represent the views of MIT as an institution or any individual or their organizations. For academics, there may be nothing 'new' in this article. The synthesis of ideas from a variety of sources, when presented in confluence, as suggested in this article, may reveal why rapid transformation of decision support systems is necessary and the increasing role of interoperability. Please email to shoumen@mit.edu or write to Dr. Shoumen Palit Austin Datta, Research Scientist, Engineering Systems Division, Department of Civil and Environmental Engineering and Research Director and Co-Founder, MIT Forum for Supply Chain Innovation, School of Engineering, Rm 1-179, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (Phone +001.617.452.3211).