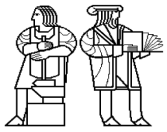

Massachusetts Institute of Technology
Engineering Systems Division

Working Paper Series

ESD-WP-2007-17

UNIFIED THEORY OF RELATIVISTIC IDENTIFICATION OF INFORMATION IN A SYSTEMS AGE:

Proposed Convergence of Unique Identification with Syntax and Semantics through Internet Protocol version 6

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
shoumen@mit.edu

March 2007

Unified Theory of Relativistic Identification of Information in a Systems Age: Proposed Convergence of Unique Identification with Syntax and Semantics through Internet Protocol version 6

Shoumen Palit Austin Datta

Research Scientist, Engineering Systems Division, Department of Civil and Environmental Engineering and Co-Founder and Research Director, MIT Forum for Supply Chain Innovation, School of Engineering, Massachusetts Institute of Technology, Building 1, Room 1-179, Cambridge, Massachusetts 02139, USA

shoumen@mit.edu

Abstract. This paper proposes to utilize internet protocol version six (IPv6) to uniquely identify not only things (objects) but also processes, relationships (syntax, semantics) and interfaces (sensors). Convergence of *identification with information* using the 128-bit IPv6 structure offers $3.4x10^{38}$ unique instances. It is not necessary that all instances must be connected to the internet or routed or transmitted simply because an IP addressing scheme is suggested. This is a *structure* for identification which (1) may improve revenue potential from data routing (P2P packet tracking) for telecommunication industries, (2) potential use in healthcare and in biomedical sciences, (3) scope of use in the semantic web structure by transitioning URIs used in RDF, (4) applications involving thousands of mobile *ad hoc* sensors (MANET) that demand dynamic adaptive auto-reconfiguration. This paper offers clues for innovation based on a confluence of ideas that may augment systems interoperability necessary for operational transparency in a global economy.

Keywords: Interoperability, Data, Information, IPv6, Semantics, Syntax, Security, MANET, Sensors, Healthcare, Biomedical Ontology, Decision Systems, RFID, EPC, Logistics, Adaptive Value Networks, Supply Chain, P2P, ZigBee, WiFi, WiMax, WiTriCity

1 Introduction

Data (bits) from unique identification of objects or things (atoms) are often helpful to the decision making process. Decisions, however, are often based on information that takes into account multiple factors. Physical objects and their unique identification may be one of many factors, as is the internet of things, from the perspective of a systems approach. Real-world decisions are often based on collective information gathered from multiple sources (or systems) that includes data (bits) about "things" (atoms) and processes associated with "things" which may be used in combination with a higher level domain that may eventually trigger a decision or execute an action, aided or unaided by a human. Currently, we do not have a globally unique mechanism to identify *information* derived from data originating from things (objects) *and* processes. Unique identification of information, hence, is an open question.

Information, to be of value, must be ***relative*** to the context of the process. In general, contextual information is of greater relevance in the decision making process or in decision systems. In this paper, I shall refer to such information as *decisionable information*.

Since information is key, one who holds information can use data, for profit, as a pay-per-use or pay-per-access service. Hence, unique identification of data has gained considerable momentum. Transmission of data is essentially the domain of routers. Routing data is an ubiquitous and essential function (real-time data, supply chain, emergency, medical results, networked entertainment, video-on-demand, energy-use optimization, any data) performed by products in boxes (for example, routers). Some corporations are likely to explore new revenue streams simply from use of raw data, for example, data-as-a-service or how to profit from data routing by providing access to data (pay-per-use hosted services). The next higher level for potentially higher revenue originates from processed data or information-as-a-service, in addition to physical data routing.

The transformation of data to information is made even more difficult by the inability of systems and software to comprehend or understand. Advances in systems interoperability [1], adoption of sophisticated analytical techniques [2] for forecasting

and risk analysis [4] and growth of the semantic web [4] infrastructure may stimulate in-network processing of data to boost the information-as-a-service business model by making sense of data and information relative to each other. These are new sources of revenue emerging from a function that is massive in scale but poorly regulated due to inadequate ability to document and charge for individual events and instances. The identification and identity structure necessary for a scale so massive calls for a system that is able to uniquely identify and assign identity to objects, process and decision layers. This paper explores the use of the structure and alphanumeric format [5] in Internet Protocol version 6 (IPv6), for this task.

2 Format of Internet Protocol version 6 (IPv6)

This paper advocates the most obvious distinguishing feature of IPv6 due to its use of much larger number of unique addresses. The size of an address in IPv6 is 128 bits, which is four times the larger than an IPv4 address. A 32-bit address space allows for 2^{32} or 4,294,967,296 unique addresses. IPv6 uses the 128-bit address space allowing for 2^{128} or 340,282,366,920,938,463,463,374,607,431,768,211,456 (3.4×10^{38}) unique addresses. The relatively large size of the IPv6 address was designed to be subdivided into hierarchical routing domains that reflect the topology of the modern internet. The use of 128 bits allows for multiple levels of hierarchy and flexibility in designing hierarchical addressing and routing that is currently lacking in the IPv4-based internet (as well as 64-bit and 96-bit versions of the electronic product code or EPC).

IPv4 addresses are represented in the dotted-decimal format. This 32-bit address is divided along 8-bit boundaries. Each set of 8 bits is converted to its decimal equivalent and separated by periods. For IPv6, the 128-bit address is divided along 16-bit boundaries. Each 16-bit block is converted to a 4-digit hexadecimal, separated by colons (colon-hexadecimal).

The following is an IPv6 address in binary form:

0010000111011010000000001101001100000000000000000010111100111011
0000001010101010000000001111111111111110001010001001110001011010

The 128-bit address is divided along 16-bit boundaries:

0010000111011010	0000000011010011
0000000000000000	0010111100111011
0000001010101010	0000000011111111
1111111000101000	1001110001011010

Each 16-bit block in an IPv6 address, converted to hexadecimal and delimited with colons (colon hexadecimal format) 21DA:00D3:0000:2F3B:02AA:00FF:FE28:9C5A. IPv6 representation can be simplified by removing the leading zeros within each 16-bit block but each block must have at least one digit. The representative address is: 21DA:D3:0:2F3B:2AA:FF:FE28:9C5A (with leading zero suppression).

3 Opportunity Spaces

3.1 Network Services

In-network processing of data to generate information may emerge as a value-added differentiator in the data-as-a-service model. The latter may be possible with grid computing by standardizing the use of operating system and system architecture, for example, Globus Tool Kit and Open Grid Services Architecture [6] which represents evolving architecture facilitating web services (Web 2.0, Web 3.0, Web X.0) with tools, such as, SOA (service oriented architecture) over transmission medium “mesh” (including FTTH, FTTB, ATM, SONET, WiFi, WiMax, ZigBee, PLC, UWB, GPRS,

GPS, WiTriCity[1]). Grid based 'in-network' processing functions may be pervasive with increasing diffusion of software as infrastructure, which may, in turn, offer systems the functionality to harvest and route distributed data from various sources to analytical engines running various hosted applications, analytics, predictive tools, forecasting algorithms, event management alerts, scheduling and planning updates. Output from online analytical processing (OLAP) engines may be made available to businesses on a pay-per-use and/or pay-per–access basis, with costs reflecting real-time versus near-real-time service delivery.

Hence, service may "mature" to provide "answers" and not only numbers (raw data). In this scenario the network is the business. With vast quantities of data (instances of data) that need accounting (if you are the service provider), it becomes critical to find a robust, globally feasible and easily adoptable *modus operandi* to contextualize and "number numbers" that does more than deliver data. Distributed data from multiple sources is dynamic, often sporadic and volatile yet must have unique identification, usually, to be valuable. The key to profitable service is to deliver information of value that can be uniquely identified, for example, in buckets or data cubes [7]. The ability to count and account for the identity of every instance of buckets on *your network* [2] catalyses the profits if one can bill (charge) for every instance a new bucket is created.

---

[1] Electrical power without wires http://web.mit.edu/newsoffice/2007/wireless-0607.html

[2] Expanded routing and addressing capabilities that offer greater ability to control the path of traffic in order to direct the transmission of packets on the service provider's network (hence billing for service or use) is a business driver for networks seeking new revenue streams. IPv6 offers this functionality by improving the scalability of multicast routing and by introducing a new type of address called "anycast" address. When used as a part of a route sequence, anycast permits a node to select which of the several internet service providers (ISP) is wants to carry its traffic (source selected policy). This is implemented by configuring anycast addresses to identify the set of routers assigned to the ISP (one anycast address per ISP). These anycast addresses can be used as intermediate addresses in an IPv6 routing header to direct a packet to be delivered via a particular provider or sequence of providers (alliances between ISPs). These new routing extensions in IPv6 are powerful tools for provider selection and mobility. If users are enabled to change routing, then it will increase competition between service providers because choices may be guided by cost factors and quality of service. (http://playground.sun.com/pub/ipng/html/INET-IPng-Paper.html).

It could be argued that distributed data identification, of the type referred above, may be achieved using distributed hash tables [8] commonly used in peer-to-peer (P2P) systems (eg. Napster, Gnutella, Freenet, BitTorrent). However, the implications of using variants of consistent hashing to map key nodes may introduce the same type of nomenclature issues (disagreement) as with the date/format/EPC type keys. Node keys (identifier or ID) are linked to the keyspace between two keys, in other words, a node with ID ***i*** owns all the keys in the keyspace. The benefit of consistent hashing in distributed hash tables are due to the fact that alterations in the keyspace, i.e., addition or removal of nodes, changes one associated/adjacent key, leaving all other nodes unperturbed (in traditional hash tables changes require the entire keyspace to be re-mapped).

In the context of this idea, the change of nodes in the keyspace (addition or removal) may be viewed as independent events, for example “buckets” containing inventory data from a third-party supplier that is updated at regular intervals. In the pay-per-access business model, every instance or event of uploading or updating the data cube “bucket” with inventory data (for example hourly data after completion of batches of work-in-progress or WIP data) should incur a fee, payable by the user to the host network provider, as a fee for pay-per-access. It is unclear if event detection and link to pay-per-access for this type of itemization is possible through use of DHT.

In the same vein, the manufacturer (OEM) may periodically “ping” the bucket to determine inventory (supply) data. Each time the manufacturer “looks” in the bucket, there should be a mechanism to capture this event and the manufacturer (user) pays a fee to the service provider (network carrying the data and/or data cubes) for the service type: pay-per-use/access. The observation (inventory of specific item) could impact higher level WIP (work-in-progress) functions. For example, if 3 components are required for sub-assembly of a part then updating one SKU in the BOM (bill of materials) may not reflect a change of status of the *sub-assembly inventory* (this is no longer merely raw data, but *information*) because all 3 parts are required in a certain volume (ratio) to ensure that the sub-assembly can proceed and will produce the next higher level component, in question.

This is *contextualized* numbering of information where inventory of sub-assembly of component – *is the information* – dependent on inventory – *which is itemized raw data* – of individual components. While DHT is an advanced mapping tool that reduces bandwidth requirement to rapidly connect nodes in a vast network topology, it is unclear if DHT may be as useful as IPv6 to assign unique identification to every change in, or access to, data and information, that may profit from being *accounted*.

Hence, identification requires a unique way to *innumerate* packets or the data and the data holders (data cubes). Application of granular "innumeration" of data packets is in demand and likely to offer significant value for internet service providers (ISP) who are usually loathe to carry P2P (peer-to-peer) traffic on their networks because the P2P *modus operandi* (leaf to leaf) bypasses the current accounting practices of tracking packet flow (hence, revenue generation) which generally operates on the "trunk" of the system (charges for operations or content delivery that use bandwidth on the trunk of the ISP). Due to proliferation of P2P users who are taking advantage of the networks, ISPs are seeking tools to track and trace P2P traffic through the leaf nodes in order to charge P2P users for transactions (video, music, data) based on volume rather than the current flat-fee revenue model used by the telecommunication industries and ISPs because they do not yet have a granular mechanism to track how much data (packets) an user is downloading or uploading. A satisfactory revenue model based on tracking P2P usage in units of packets may further bolster revenue from legitimate P2P users and help reduce ISP bandwidth congestion by distributing payload to leaf nodes or mesh networks (away from trunk lines). To help with these accounting problems in P2P mesh networks, in one approach[3] files are reassembled in 'slices' where a slice consists of the *n*th bit of every block. It is open to exploration whether in this context the numbering of files (the *n*th bit of every block) with an unique IPv6 id may offer business value through granularity of accountability.

---

3 Hui Zhang of Carnegie-Mellon University has founded Rinera Networks and Paul Francis of Cornell University has developed a system called Chunkyspread to pratially address these problems (MIT Technology Review, March-April 2007).

3.2 Event Management

In event management and related supply chain operations, database tables with unique identification for each data cube may lend itself to re-use as data holders. Each event (that is, every time data or a data cube or a data holder is used) is uniquely numbered (companies subscribe to number domains) and itemized for pay-per-access service charges based on instances. Event data update and/or management may utilize online analytical and/or transaction processing (OLAP/OLTP) type systems for customer relationship and billing purposes, respectively. This proposal to use IPv6 may address the central dilemma: how do you number numbers? How does one number events? Merely "counting" the number of numbers is not enough. This will run into octillions of instances of data and even more instances of contextualized data that is of value (relativistic or decisionable information).

Consider problems faced by retailers if they double-count or re-count. Then imagine if a system-wide data duplication goes unnoticed and if services (finance, insurance) are duplicated or irregular. Individual pieces of data are often sterile as decisionable information but the collective analysis of data in the context of the process or transaction (where it is used) is far more critical and valuable as decisionable information. Object data identity using other formats, such as the electronic product code (EPC) or GUID (global unique id), are static identification formats that remain oblivious of context (process). Although EPC and other formats such as the UCR (unique consignment reference) offers unique id for things or objects, they lack structure to offer *contextual information* that may have unique identification.

The problem with "unique id" is addressed, grossly inadequately, by EPC Global and other related bodies (GS1) through the EPC-IS interface standard. Entities stored in EPC-IS are events, contextualized through "what/where/when/why/how" combination of parameters. Unique numbering for events rely on generating a combined key by using, for example, EPC+location+time. Pre-agreement of time, location coordinates and formats are core assumptions when creating such a key. It promotes the fallacy that such agreements are universal and universally accepted (explore how many

permutations and combinations are possible with different formats of times and dates used on different continents).

Therefore, this problem may benefit from innovation and further exploration of "numbering of numbers" to include EPC type information (if necessary) in a sub-layer of a multi-layer data and information aggregation model to provide unique *identity* for information. One practical example of data "layers" may be viewed as a substitute or replacement for EPC IS where a unique serial number (bottle of Aspirin with EPC) is combined with attributes (not designed for EPC association) such as [a] who checked the packaging, [b] where was it packed, [c] where was it manufactured, [d] what is the expiration date, [e] when was it shipped, [f] when was it received, [g] where was it shipped from, [h] where was it received. "Who" denotes a role and individuals with unique IPv6 id may be linked to role based authorization for accessibility purposes. Location, transportation vehicles, physical spaces (distribution center, retail store) may have unique IPv6 type id, too. The *combined* higher level information linked with the serial number of the bottle of aspirin (with its unique IPv6 id) with the "who-what-where-when" parameters is as simple as a "name" (John Smith) in a relational database linked to mailing address, phone number, date of birth, biometric data, social security number. If the "name" forms the skin of an IPv6 type crawl [9], it may extract the *related* data to generate information about John Smith. Alternatively, if the zip code or post code serves as the skin in a Web X.0 "search and discover" function, it may help track and trace all who are John Smith in Cambridge (02139 or CB2 1HQ). Thus, ***layers*** of unique IPv6 id representing object data and semantic data (who-what-when-where) collectively generates ***information*** when combined. The "simple sensed" semantic data (local distribution centre ***is a*** regional distribution centre) may not be mixed up with EPC type id. It may be worth the upheaval to forge a new direction by associating semantic data (such as, a retail store, that does not serve as an actuator) with IPv6 type identification to facilitate future search engines and "intelligent" agent-based systems to help track and trace not only the movement of objects but the linked data and processes in relation to information.

In this respect, it may be rather unfortunate that EPC has deviated from its insightful 128-bit structure that was proposed by Sanjay Sarma and Dan Engels (MIT, 1998). It may have successfully aligned EPC with IPv6 evolution. EPC 128-bit was a format for data (id) transmission but without routing yet it had the potential for convergence or structural amalgam with routing if translated to IPv6 format with substantial unique identification provision and transmission-routing capabilities for any RF mode (RFID, UWB, sensors) that can be plugged directly into the internet. Currently, sensor data cannot be uploaded directly to the internet because individual sensors cannot be assigned IP addresses due to limitations of the IPv4 unique addressing system. This proposal works with sensors (interfaces) because individual sensors can now have unique IPv6 address. However, the fundamental question is how to extract value from data and how can providing data as a network service evolve a robust revenue model?

Thus, data as a service model begs to find a mechanism that can number numbers in a rational manner to generate the higher order contextual and relativistic information. This mechanism must exist within the current framework of TCP/IP because it is aligned with the internet architecture. Hence, the suggestion to use the 128-bit IPv6, a scheme that is already at hand and being deployed, gradually but globally. This may be the “one-size-fits-all” hyper-id. IPv6 hyper-id may be a lucrative accounting tool in event recognition network delivered services to enable billing, for example, for high volume entertainment events, such as, iTunes from Apple. Management of iTunes billing, tracking and tracing, may be overwhelmed by sheer volume. It may also experience constraints not only for billing services but also up-selling/cross-selling types of customer relationship management (CRM) and marketing due to its inability to track and trace combinations of parameters that identify customer (nodes) choices.

3.3 Relativistic Information

In a similar vein, it may be necessary to identify people and process, in combination. How do we uniquely identify people and their linked or related context (process), for example in healthcare or in eGovernment functions (such as pension, voting).

EPC is object based identification. Think of IPv6 as a solution that is pre-agreed for global adoption. Consider contextual relevance of data and how to "number the numbers" but especially how to route numbers (data) which may be identical but with different identities. For example, a blood glucose result of 120 mg/dl may be identical for multiple individuals (same number, 120, with same units, mg/dl) but with different identities at the informational level because it belongs to different individuals. This may seem an esoteric brain teaser to some but *au contraire*. This is a fundamental information infrastructure issue that requires serious attention. It may have critical impact, for example, in a healthcare scenario where in-network data processing may be valuable with concomitant development of semantic tools for discovery and hosted analytics over semantic grid (Web X.0) services. However, in-network processing is granular but only for the application in question (*per se* it does not lead to higher order aggregation or network topology). Connecting data from, for example, from an emergency scenario (heart rate) with patient history may be viewed as a higher layer "abstraction" where not only the data (heart rate in bpm; corrected for age) but the information (heart rate and patient history) is identifiable and accessible through a "hyper-id" accessible via web services with linked information that is unique and uniquely identifiable. Thus, healthcare is a fertile ground for such applications. This suggestion may work with Agent-integrated security systems (to guarantee data security, confidentiality and privacy[4]) and monitoring systems to scan for errors (double counting, double dosage) as well as duplicate address detection (DAD)

---

4 Service providers can now offer secure services by using two integrated options in IPv6. One mechanism, termed "IPv6 Authentication Header" is an extension header which provides authentication and integrity *without* confidentiality. This may prevent a number of network attacks. The other mechanism, "IPv6 Encapsulating Security Header" provides data integrity *and* confidentiality. Authentication, data integrity and confidentiality are core elements of IPv6 and offers an improved option mechanism over IPv4 (not limited to 40 bytes). It permits IPv6 options to be used as functions which were not practical for IPv4 (authentication and security encapsulation for security and confidentiality, respectively). In addition, the use of anycast address enables service provider selection. This feature may be used for secure services (dedicated service provider) where the anycast address may not be altered and hence the data must be routed according to a source selected policy, eg: healthcare, customs, military and finance. (http://playground.sun.com/pub/ipng/html/INET-IPng-Paper.html)

systems to prevent data binding irregularities that may have fatal consequences in healthcare. These applications may also develop to include "Medical Google" type web services for point-of-care reference where human experience or human-aided searches may be necessary to supplement diagnosis or diagnostics or automated information processing, distribution and execution systems.

The industry may not yet consider this a mature or profitable endeavour because it is erroneously concluded by some that the impact of this mechanism may not manifest soon enough. Japan and the Nordic countries have sufficient internet penetration and an increasing percentage of the population are e-savvy yet grey (over 65 year old with increasing demand on healthcare services, as one example). They may soon seek and need e-healthcare "sense *then* respond" systems (Figure 1) to stem down healthcare costs. The system must respect privacy, data identity and security. This problem is real and will require a solution if industrialized nations expect that their healthcare expenditure remain a reasonably small percentage of their gross domestic product.

Clearly a broad spectrum of applications will find it essential to "number numbers" in a manner where identity and relativistic identification lends itself to information and context including provision for quality of real-time[5] service at the right-time, all the time, plus authentication, data integrity, confidentiality and privacy.

---

5 The buzz around "real-time" data is as intense as the inability of most businesses to extract value from real-time data (see http://esd.mit.edu/WPS/esd-wp-2006-11.pdf). However, the cost of network delays in data transmission can have hazardous or even fatal effects in emergencies or accidents. Thus, IPv6 offers a new capability by introducing [a] the 24-bit Flow Label and [b] the 4-bit Priority field in the IPv6 header to enable labeling of packets belonging to particular traffic "flows" for which the sender can request special handling for real-time applications. Currently some hosts and routers do not support this function.
[a] The 24-bit Flow Label in the IPv6 header is designed to be used by a host to identify packets for non-default quality of service (QoS) or "real-time" service. A flow is a sequence of packets sent from a particular source to a particular (unicast or multicast) destination for which the source requests special handling by intervening routers. The nature of the special handling may be conveyed to the routers by a control protocol (eg: resource reservation protocol) or through embedded information in the flow packet (eg: hop-by-hop option). There may be multiple active flows from a source to a destination, uniquely identified by combination of a source address and a non-zero flow label. Packets that do not belong to or not associated with any flow traffic, carry a flow label of zero. (http://playground.sun.com/pub/ipng/html/INET-IPng-Paper.html)

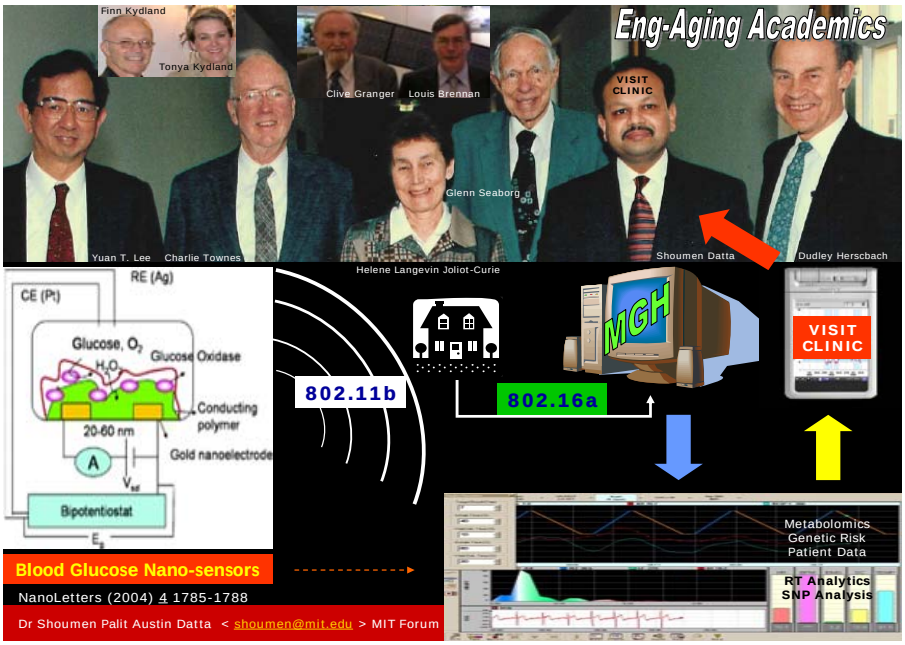

Fig. 1. Emerging eHealthcare "sense *then* respond" decision system may reduce cost of old age

3.4 Convergence with Semantic Web

The sub-title of this section is still really a question or a hypothesis, not a statement. The convergence is imminent because it is well nigh impossible to mention the word "information" without thinking about syntax, semantics, relations and other elements of interoperability, which are crucial for increasing the value of information to generate decisionable information.

---

[b] The 4-bit Priority field in the IPv6 header enables a source to identify the desired delivery priority of its packets relative to other packets from the same source. The Priority values are divided into 2 ranges: values 0 through 7 specify the priority of traffic for which the source is providing congestion control (traffic "backs off" in response to congestion). Values 8 through 15 are used to specify the priority of packets that does not back off in response to congestion (transmitted in real-time at a constant rate).

The ontological framework necessary for the semantic web uses a powerful tool, RDF or resource description framework, which works, in part, through use of URI (universal resource identifier). The latter is a higher layer abstraction that uniquely identifies a resource (even if the resource lacks a web address). These relationships must be *pre-designed* before RDF and RDF/XML can be applied.

These relationships, for example in the biomedical field, are referred to as "controlled ontological vocabulary" which must be created to plumb (search) the depths of the repertoire. This process generates even more syntax or labels to script the triples (subject-predicate-object) and introduces variability due to type of group, individual, nationality, mother tongue and a host of factors that may be difficult to standardize in an open environment (hence, routed through slew of committees for acceptance). The vast number of groups involved in creating ontologies uses a myriad of formats to specify application-specific domains. When these relation "trees" are consecutively added to higher level hierarchies (ontological soup ensues) the problem of keeping track of tags and meta data may create quagmire for which yet another standard may be sought! This problem is illustrated by the omics initiative [10] involving a handful of post genomic technologies (transcriptomics, proteomics, metabolomics). Several standardization groups are working on metadata and ontology in this sub-field.

If 'omics' may be defined as processes operating within a cell and the 'cell' subclass contains two related fields (intra-cellular and extra-cellular) and if cells are considered parts of 'tissues' and tissues are classified under 'organs' then, imagine the plethora of committees necessary to standardize and define biological molecules. Is there a feasible alternative to track and trace these ontology relationships that is amenable to machine intelligence and include use of metadata tags?

Hence, there are several deep issues involved in this simple example. Therefore, for the context of this argument, the focus is only on one sub-issue: the nature of the unique resource identifier (URI). Is there a benefit to transition the URI to a pre-agreed system that all groups may agree to use without further standardization? My naïveté (fools rush in where angels fear to tread) proposes the use of IPv6 to uniquely identify (map) URI layer from domains of pre-subscribed numbers which are assigned to classes and subclasses (illustrated in Figure 2 and Figure 3).

Table 1. An ever increasing number of standardization activities

TABLE 1. AN EVER INCREASING NUMBER OF STANDARDIZATION ACTIVITIES

Community effort and website	*Standardization activities*	*Citation in this issue*
Genomic research		
The Genomic Standards Consortium (GSC): ⟨www.genomics.ceh.ac.uk/genomecatalogue/⟩	Content (Minimal Information about a Genome Sequence: MIGS), syntax, and semantics	(Field et al., 2006; Morrison et al., 2006a)
International Nucleotide Sequence Database Collaboration (INSDC): ⟨www.insdc.org⟩	Content (INSDC Third-Party Annotation Submission Guidelines)	(Cochrane et al., 2006)
Genome Reviews: ⟨www.ebi.ac.uk/GenomeReviews/⟩	Content (review of standardization within the Genome Reviews database) and syntax	(Sterk et al., 2006)
GSC: Chair of Organelles Working Group: ⟨www.genomics.ceh.ac.uk/genomecatalogue⟩	Content (call for standardization of descriptions of organelles)	(Boore, 2006)
Post-genomic standardization		
MGED Society: ⟨www.mged.org⟩	Content (Minimal Information about a Microarray Experiment: MIAME), syntax, and semantics	(Ball and Brazma, 2006)
HUPO–Proteomics Standards Initiative (PSI) ⟨http://psidev.sourceforge.net⟩	Content (Minimal Information about a Proteomics Experiment guidelines: MIAPE), syntax, and semantics	(Taylor et al., 2006)
Experimental Standards for Proteomics	Content (call for development of standard experimental mixtures of proteins)	(Hogan et al., 2006)
Metabolomics Society–MSI (Metabolomics Standards Initiative): ⟨www.metabolomicssociety.org⟩	Content, syntax, and semantics	(Fiehn et al., 2006)
Integration activities		
Reporting Structures for Biological Investigations Working Group (RSBI): ⟨www.mged.org/Workgroups/rsbi/rsbi.html⟩	Contributions to content and semantics	(Sansone et al., 2006)
"Env" Community led by the Environmental Genomics Working Group (EGWG): ⟨http://envgen.nox.ac.uk/miame/miame_env.html⟩	Content (MIAME/Env checklist), syntax, and semantics	(Morrison et al., 2006b)
The Functional Genomics Experiment Object Model (FuGE): ⟨http://fuge.sourceforge.net/⟩	Syntax	(Jones et al., 2006)
National Center for BioMedical Ontology (cBIO): ⟨www.bioontology.org/⟩	Semantics	(Rubin et al., 2006)
Functional Genomics Investigation Ontology (FuGO): ⟨http://fugo.sourceforge.net/⟩	Semantics	(Whetzel et al., 2006)
Other initiatives		
MISFISHIE Working Group: ⟨http://mged.sourceforge.net/misfishie/⟩	Content (MI Specification for *In Situ* Hybridiation and Immunohistochemistry Experiments: MISFISHIE), syntax, and semantics	(Deutsch et al., 2006)

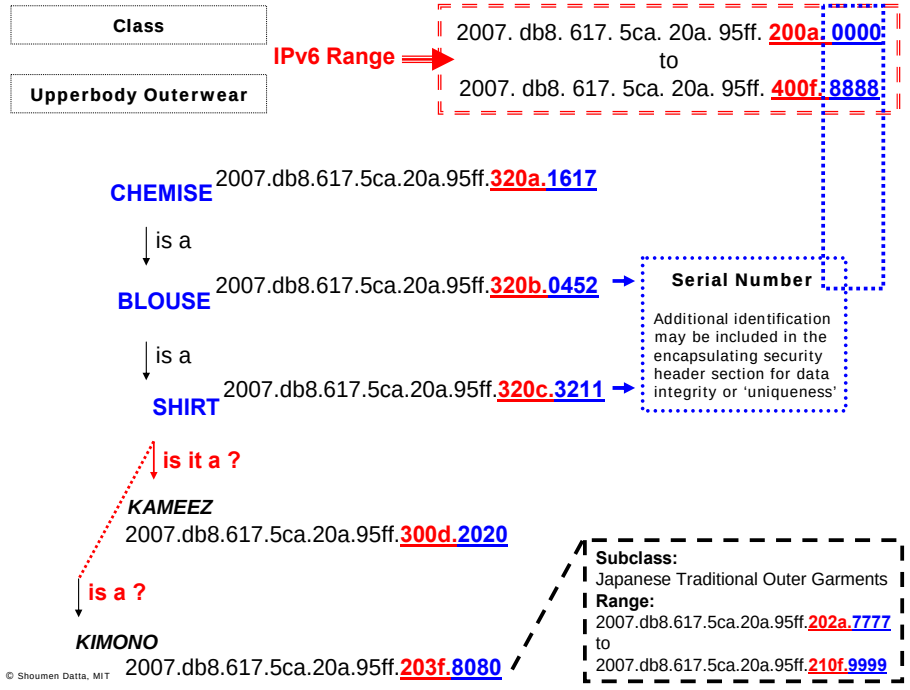

Fig. 2. Co-embedding unique id with syntax and semantics?

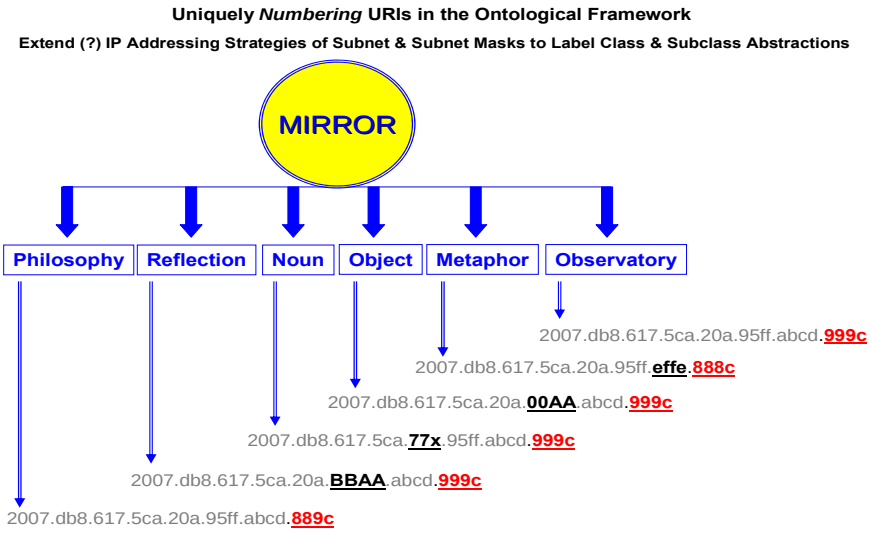

Fig. 3. Moving from descriptive or "analog" to unique or "digital" relationships?

Because IPv6 has a reasonable chance to be adopted globally, a mechanism to map URI abstractions to IPv6 structure offers synergistic convergence. It may reduce the gulf between business thinking and systems (ICT) specifications.

However, the creators of IPv6 pointed out that no matter how good a new protocol (or idea) may be, it may not matter if there isn't a practical way to transition from one to another. There may be several known reasons why the suggestion in Figure 4 (below) of transitioning URI to an IPv6 format may not be feasible. On the other hand, unknown unknowns (reasons) may exist that may make this suggestion feasible or practical when and if the unknowns emerge as knowledge.

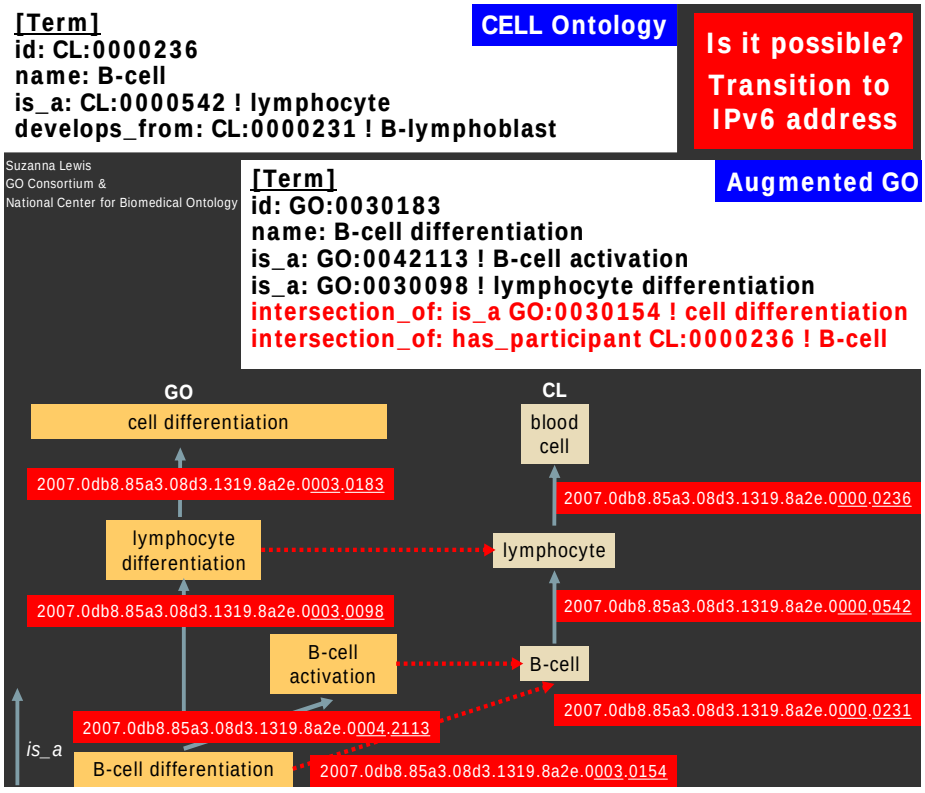

Fig. 4. Potential transition strategy?

3.5 Security

Another application area that may benefit from a vast library of detailed resources with unique identification that can be selected in a dynamic manner, is that of policy. Often, a "handbook" of policy may contain several thousand clauses and sub-clauses (assume that all these details exist as thousands of separate URIs). When a scenario surfaces (for example, US Department of Homeland Security certified Tier 2 business group authorized to import finished goods in containers, files a manifest that indicates additional inventory of sub-components) that calls for use of selected policies (to authorize a search of the container, in this case) then resources (that is, policies) that are applicable for that scenario can be selected (from a remote management location) by using the routing capability of IPv6 to choose sets of interfaces (with IPv6 addresses) from thousands of policies (from the policy domain) that may be involved in the analysis, assessment and management of security threats and risk (Figure 5).

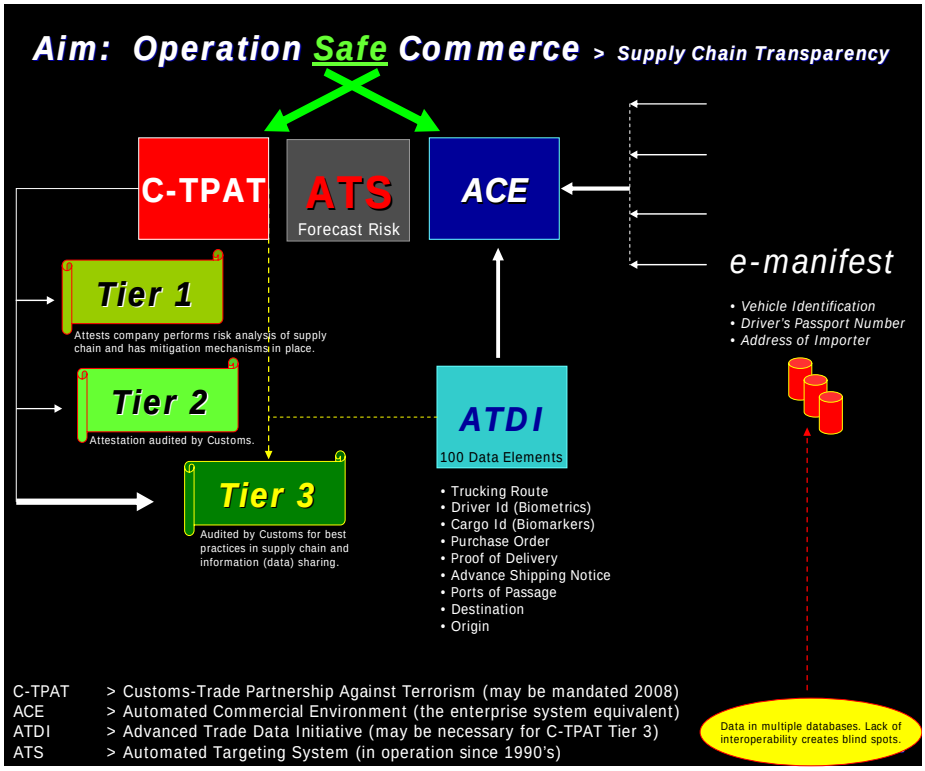

Fig. 5. Risk Management through systems approach (convergence of data, policy, information)

In a related security scenario, it may be significant to analyse the following: data acquired from sensors (data element 1) linked to a specific container (data element 2) transported on a vehicle (data element 3) driven by a credentialed driver (data element 4) who belongs to a logistics provider (data element 5) registered as a NVOCC (non-vessel operating common carrier) with a NVOCC code (data element 6). If customs and border protection (CBP) wants to anaylse this sensor data at a local port of entry for targeting purposes, it is likely to assign a key, quite similar in concept and practise to EPC-IS (see section 3.2). In this scenario the key is a field or combination of fields used to "anchor" other fields. These methods may still be in use for "look data up" using key fields which are "arbitrarily created, for example, by creating some unique record ID or other unique count to distinguish otherwise unidentifiable data" [11].

It is, therefore, amply clear to understand why such arbitrary identification does not lend itself to portability and interoperability between systems. A clue uncovered and assigned a "key" at a port of entry by the US Coast Guard (USCG) remains "unique" in that system but may not be visible or meaningful in its association within the Automated Commercial Environment (ACE) or other targeting systems where non-obvious relationship analysis (NORA) is performed (for threat assessment and risk).

If NORA points to possible need for inspection, in a related scenario, confluence of policy becomes equally important to ensure execution of action is within legal limits. The laissez-faire concept of creating keys for supposedly unique identification needs a careful review, especially in threat assessment and risk analysis. Few "red handed" instances will be revealed through obvious relationships. Non-obvious relationship analysis is increasingly crucial to connect arcane associations that must be sorted from vast amount of essential and non-essential data, uniquely identified in a manner that can be accessed by any authorized system in any geographic location and analysed repeatedly in different contexts in diverse domains to determine risk profile.

Data analysis for domestic (US) threat assessment in the transportation sector alone, could originate anywhere, in any form, in several instances, several times a minute or

hour, in the network connecting towns, cities, manufacturers and retailers, moving large volumes of goods and individuals through a system of approximately 4 million miles of roads and highways, 120,000 miles of rail road, 600,000 bridges, more than 300 tunnels and numerous sea ports, 2.2 million miles of pipeline, 500,000 train stations, 500 public airports operating more than 200,000 commercial aircrafts [12].

Table 2: Why IPv6 unique identification of information may be crucial for security and NORA

- 3.9 million miles of public roads
- 1.2 million trucking companies
- 15.5 million trucks
- 42,000 hazardous material (HAZMAT) trucks
- 10 million commercial vehicle drivers including 2.7 million HAZMAT drivers
- 2.2 million miles of hazardous liquid and natural gas pipeline
- 120,000 miles of major railroads
- 15 million daily riders on mass transit and passenger rail, nationwide
- 25,000 miles of commercial waterways
- 361 ports
- 250,000 containers per day
- 9.0 million containers through 51,000 port calls
- 11.2 million containers via Canada and Mexico
- 19,576 general aviation airports, heliports and landing strips
- 459 Federalized commercial airports
- 211,450 general aviation aircraft
- 77% of all flights are general aviation

3.6 Mobile *ad hoc* Sensor Networks (MANET)

By some estimates, the scope of unique identification using IPv6 is a staggering $6x10^{19}$ addresses per mm^2 of the surface of the Earth, based on the fact that IPv6 supports $3.4x10^{38}$ addresses and an approximation that $5x10^{14}$ m^2 is the surface area of our planet. It may be argued that the number ($6x10^{19}$ addresses per mm^2) represents a monolayer of interfaces (a node may have multiple interfaces, each with a globally unique IPv6 id) which is an unlikely scenario in internet devices, for example, an automobile or equipment such as HVAC (high voltage air conditioning) system. In a

vehicle or HVAC there may be a vast array of interfaces (nodes, sensors) packed in close proximity (multi-layer topology). Other internet devices or internet appliances (example: refrigerator, light switches) may have multi-layer topologies. Internet infrastructures, for example, buildings, may be densely packed with interfaces (for example, if every TV or light switch is an IP node, then think about the number of televisions and switches in Taipei 101, the tallest building in Taiwan).

Hypothetically, assume that the entire surface of the Earth may be organized as layers of interfaces 1 mm apart and that this hypothetical layer is 100 km in depth (consider sensors and actuators in deep sea drilling equipment and observatories on top of the Mount Everest). Even if we have such an improbable density of interfaces demanding globally unique id, the number of possible unique id is approximately $6x10^{10}$ IPv6 addresses for mm^2 of Earth's surface that is 100 km or $1x10^9$ mm deep! In other words, 60 billion unique addresses per square mm of the Earth at a depth of 1 billion mm (Figure 6). This logic fuels the speculation that internet communication with Martians and objects in Mars may be within the scope of the design of IPv6.

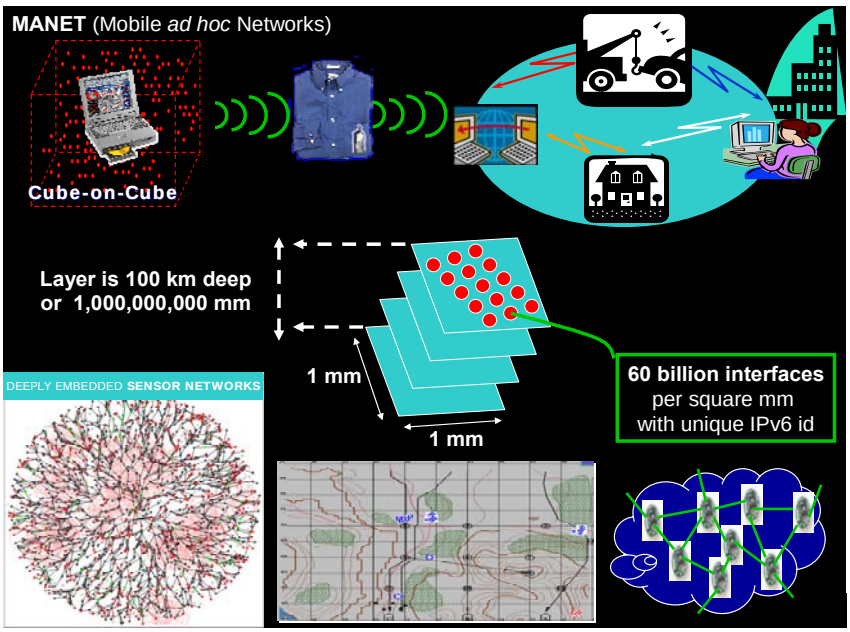

Fig. 6. Uniquely identifiable sensors

This logic supports the extrapolation that IPv6 addresses may be used for unique id in mobile *ad hoc* sensor networks (MANET) in military applications where trillions of sensor nodes may be connected to logistics decision or monitoring systems of the war-fighter (Figure 7). These systems demand dynamic adaptive routing and auto-reconfiguration yet must retain critical data links, identity, coordination and interoperability with multiple commands or legacy (ERP) systems operating in vastly separated geographic locations.

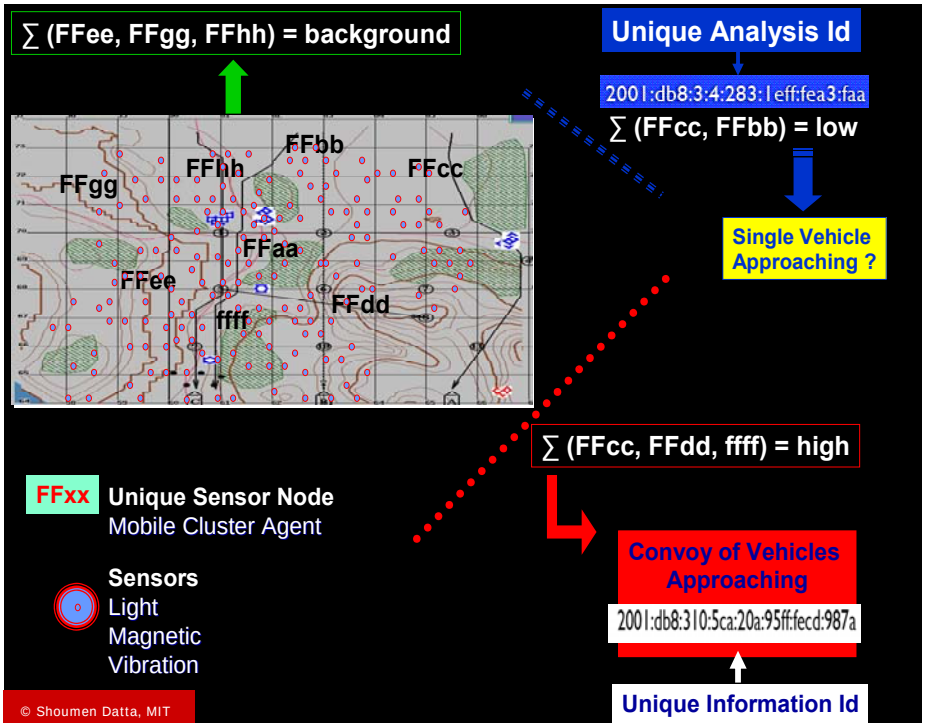

Fig. 7. Military Monitoring in Practice

Individual sensors may now be directly "plugged" in the internet using IPv6. Routing extensions in IPv6 support powerful functionalities like host mobility (route to current location) and auto- readdressing (route to new address) that may be useful for mobile *ad hoc* networks (MANET). The routing option is used by IPv6 source to list one or more intermediate nodes (in the network topology) to be "visited" on the path to the

destination of the packet. For dynamic and adaptive MANET, the management of host mobility and automatic readdressing functionalities may require convergence of IPv6 features in combination with mobile agents [13] or mobile cluster agents that can manage the "measure of distance" of routing protocols to a higher level of abstraction where granular data (packets) are aggregated (Cluster Agents) and decision or higher level information with unique IPv6 id is transmitted to one or more (unicast, anycast or multicast) superior node in the network (MANET) topology or linked to other decision systems (legacy, applications). Assignment of IPv6-based id (pre-subscribed domain) may be managed by higher order Cluster Agents. Because some of the activities may be closed loop or terminal, it is possible to reclaim underutilized assigned network id. The reclaiming process may follow a pattern of inactivity (loss of signal over extended period) and the recovery of id may be governed by elementary autonomic paradigms drawn from the AI (artificial intelligence) domain.

For consumer purposes, an example of MANET is a vehicle. Hence, transportation, logistics, supply chain and business, military, healthcare operations that need real-time data plus analysis of information may access granular data (unique data cubes) yet extract decisionable information in a manner that may be uniquely identified (traced) to its source. Track and trace identification is key to guarantee food-water safety and in customs operation where *ad hoc* assessment of threat or risk is crucial. Thus, IPv6 supported large hierarchical addresses will enable the functional growth of the internet and provide new routing capabilities that were not robust enough in IPv4.

4 Concluding Comments

The inclination to reap, with haste, the 'low hanging fruits' often derail convergences necessary for adaptive strategies [14] in business and in a broad spectrum of decision systems including logistics management [15]. Contextual numbering of numbers and contextual identification of decisionable information, using IPv6 represents, albeit only one element, but a potentially valuable confluence of identification, transmission and routing of data relative to information in systems to provide answers [16]. IPv6

may offer a plausible and simple mechanism not only to track global movement of bits (data) and atoms (physical goods) and to connect bits and atoms but also to connect ***bits to bits***. This suggestion has received some preliminary [17] support.

Acknowledgements. This idea germinated while attempting to explain to a class of international Master's students the future potential for innovative tools in systems interoperability and visibility (www.student.chalmers.se/ka/hp/hp?hp_id=3013). The author is indebted to the vibrant students in the ITR536 course for their reluctance to accept shoddy explanations of what constitutes and attributes necessary for, increasing visibility and in future, transparency, in geographically agnostic intelligent decision systems. This was during November 2006 while teaching the Master's course in Supply and Demand Chain Design and Management (ITR536) at the Chalmers University of Technology in Gothenberg (Sweden), where the author was a visiting faculty in the Division of Logistics and Transportation in the School of Technology Management and Economics. The author is grateful to Professor Kenth Lumsden of Chalmers University of Technology and Professor Gunnar Stefansson of Chalmers University of Technology and University of Iceland, for this rewarding invitation to teach and advise students at Chalmers University of Technology. Generous financial support for this period was provided by the Division of Logistics and Transportation, Chalmers University of Technology. The author's teaching duties for ITR536 were particularly invigorating due to the kind support offered by Mr Ola Hultkrantz and Mr Per Medbo of the Division of Logistics and Transportation at Chalmers University of Technology (Goteborg, Sweden). In addition, the author has gained insights from discussions about this idea with Professor Gunnar Stefansson (University of Iceland and Chalmers University of Technology), Mr Andy Mulholland (Global Chief Technology Officer, CapGemini), Dr Joseph Salvo (Director, GE Global Research), Dr Robert Ciskowski (IBM Innovation Center), Professor Rodney Brooks (Director, Computer Science and Artificial Intelligence Laboratory, MIT), Professor John Williams (Director, Auto-ID Lab, MIT), Mr Micah Samuels (Senior Manager, Amazon), Dr Richard Swan (Chief Technology Officer, T3Ci), Dr Maria Azua (Vice President, IBM), Dr Ralph Droms (Chief Engineer, Office of the Chief Technology Officer, Cisco Systems) and Professor Joseph Sussman (J R East Professor, Professor of Engineering Systems and Civil and Environmental Engineering, MIT). This idea has received some support from Charles Simonyi of IntentSoft (formerly with Microsoft Corporation), Mark Greaves of

Vulcan (formerly with DARPA, US Department of Defense) and Professor Sanjay Sarma, MIT (Member of the Board of EPC Global and Co-Founder and former Research Director of MIT Auto-ID Center). Encouragement from my colleagues at the MIT Forum for Supply Chain Innovation (Professor David Simchi-Levi, Dr William Killingsworth, Ms Janet Kerrigan and Miss Megan Gately) is gratefully acknowledged. I am thankful to Ms Janice Hall for her help with uploading and updates on the MIT ESD website while this was a working paper.

References

(1) Datta, S.: Semantic Interoperability between Systems. Working Paper. Engineering Systems Division, MIT http://esd.mit.edu/WPS/esd-wp-2006-10.pdf (2006)

(2) Datta, S., Granger, C: Potential to Improve Forecasting Accuracy. Working Paper. Engineering Systems, MIT http://esd.mit.edu/WPS/esd-wp-2006-11.pdf (2006)

(3) Datta, S., Granger, C.W.J., Barari, M., Gibbs,T.: Management of Supply Chain: an alternative modeling technique for forecasting. *Journal of Operations Research Society (in press)*

(4) World Wide Web Consortium www.w3c.org

(5) Hinden, R.M.:http://playground.sun.com/pub/ipng/html/INET-IPng-Paper.html (1995)

(6) Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)

(7) Minsky, M.: The Society of Mind. Simon and Schuster, New York (1988)

(8) Balakrishnan, H., Kaashoek, M.F., Karger, D., Morris, R., Stoica. I.: Looking up data in P2P systems. Communications of ACM (February 2003)

(9) Sarma, S.: *personal communication*

(10) Rubin, D.L.: Advancing Biomedicine through Structured Organization of Structured Knowledge" OMICS Journal of Integrative Biology 10, 185-198 (2006)

(11) Joslyn, C., Mniszewski, S.: Relational Analytical Tools: VisTool and Formal Concept Analysis. Report prepared for Advanced Knowledge Integration for Terrorist Threats (2002)

(12) Transportation Security Administration, Department of Homeland Security www.tsa.gov

(13) Maes, P., Minar, N.: Mobile Software Agents for Dynamic Routing http://nelson.www.media.mit.edu/people/nelson/research/routes-sigmobile/ (1999)

(14) Datta, S., *et al* Adaptive Value Networks in Chang, Y. (Ed.) Evolution of Supply Chain Management: Symbiosis of Adaptive Value Networks & ICT. Kluwer Academic Publishers, Amsterdam, 3-75 www.wkap.nl/prod/b/1-4020-7812-9?a=1 (2004)

(15) Datta, S.: Adapter, optimiser, prévoir - La convergence des concepts, des outils, des technologies et des normes peut-elle accélerer l'innovation? Logistique & Management 12, 3-20 (2004)

(16) Datta, S.: Answers, not Numbers: Catalysing a Solutions Approach for Distributed Decision Systems (*in preparation*)

(17) Simonyi, C.: *personal communication*