

15.9.2010

WHITE PAPER: FINLAND AND DATA RESERVES

Reijo Paajanen

Pauli Kuosmanen

Copyright © TIVIT

(ICT SHOK)

www.tivit.fi

15.9.2010

*“We're entering a new world in which data may be more important than software.”
(Tim O'Reilly)*

PURPOSE

This white paper underlines the importance of the growth of data assets and highlights the possibilities and main steps that Finland could take in order to secure its place as a leading country in the field of data reserves. This paper does not aim to go into great detail – rather outlines the major arguments and trends.

1 THE GROWTH OF DATA RESERVES

*“It is a capital mistake to theorize before one has data.”
(Arthur Conan Doyle, Sr)*

The appearance of innovations like the phone, microprocessor and computer, combined with development in tele and data communications, means that the existence of the information and knowledge society is a reality. However, this development phase is still very much in its infancy and we can expect several new development waves to come.

One important result of this development is that a new type of work – data and information work – has been created and is now present to varying degrees in almost every job which humans undertake in the developed world. When mobile phones are spreading to billions of people, soon everyone has the capability to directly create digital data and when making a simple voice call the network will record data at least for billing reasons. This way almost the every action of humans is contributing directly or indirectly to the data creation. Human society creates currently huge amount of data and this is growing exponentially.

Today processors are everywhere, memory is inexpensive and basic ICT technology is getting cheaper every day. The result is that data creation is easy and people expect there to be some kind of log, evidence or record of various tasks or actions available. What is more, we seek to store most of our documents, photos and emails and certainly want to have access to our bank account wherever we might be. Digitalization is propelling our world towards a reality in which a service is accessible by anyone, anywhere and anytime. And anything is recorded and more and more data is created.

Media is today very much initially created in digital form – with these digital media products being easy to distribute over a network. There are also plans to digitize every single book in the world. Paper is very much used as a temporary display and when its role is replaced by digital displays paper will not anymore define the limits of media products. The next phase of cinema and TV is that of 3D with high definition, of course, as standard. Such development means that people need radically bigger movie files, better quality media products and photographs – all of which again indicates just how much data storage expectations are growing.

15.9.2010

The version 5 of the CompactFlash standard will allow CF cards to hold 188 petabytes of data. To put that in perspective, we need to remember that that is equivalent to 188,000 one-terabyte drives - with sufficient capacity to contain 2.7 million hours of HD video or more than one week's worth of AT&T's daily traffic (which currently averages 18.7 petabytes). But when the cards are ready there will probably be exabytes of ultra-high-definition 3D home videos that one will want to keep track of – meaning that a paltry petabyte card will look just as pathetic as a 1GB card does today. Technology creates possibilities – and usage expectations and needs are developing almost in tandem with this development.

The following table and figure illustrate the amounts and estimations of data and its growth. During this year, 2010, human society will create clearly over one zettabyte of new digital data.

The Yottabyte Scale	
1 Petabyte 1000 Terabytes 250 000 DVDs	200 Terabytes = The contents of all books ever written in any language (Cisco)
	100 Petabytes = The amount of data produced in a minute by the new particle collider in CERN (Cisco)
	5 Exabytes = Every word ever spoken as a textual transcript (Cisco)
1 Exabyte 1000 Petabytes 250 million DVDs	21 Exabytes = Global monthly Internet traffic (Cisco, March 2010)
	100 Exabytes = A video recording of every single meeting that took place last year across the world
	500 Exabytes = The size of the World's total digital content as of May 2009 (The Guardian) = ten times higher than a stack of books stretching from Earth to Pluto
	850 Exabytes = The amount of digital data created in January - August 2010 (IDC)
1 Zettabyte 1000 Exabytes 250 billion DVDs	1.2 Zettabytes = the current total amount of global data is expected to pass this figure sometime during 2010 (International Data Corporation)
	42 Zettabytes = Everything any human has ever said, if digitized as 16 kHz 16-bit audio (Mark Liberman)
	66 Zettabytes = The amount of visual information conveyed from the eye to the brain of the whole of mankind in one year (Cisco)
1 Yottabyte 1000 Zettabytes 250 trillion DVDs	20 Yottabytes = A holographic snapshot of the earth's surface (Cisco)

Table 1. Amounts of Data.

15.9.2010

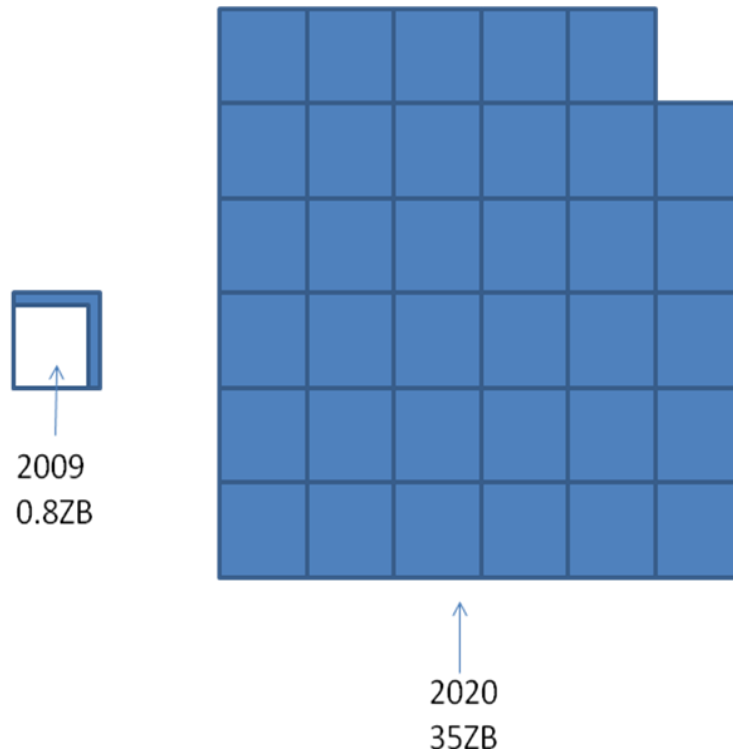


Figure 1. Increase of Data in Digital Universe (Modified from IDC, Digital University Study, May 2010).

Growth in the amount of data is obvious and society also is very dependent on data. It is time to start to speak about data reserves in the same way we describe natural resources like oil or phosphorous reserves. However, there is one big difference regarding data reserves - data reserves are growing and data is getting cheaper while other natural reserves are limited and becoming increasingly expensive as they decrease in volume.

Data reserves will be the new frontier of human society and we will see a “data reserve rush” as these reserves are mapped and as companies and countries try and maximize their share of and business opportunities from these reserves.

2 TRANSITION FROM HUMAN CREATION TO MACHINE GENERATED DATA

“Analytical software enables you to shift human resources from rote data collection to value-added customer service and support where the human touch makes a profound difference.”

(Bill Gates)

In the past, the main source of data growth was the direct human data creation - but this reality is rapidly changing. Documents have traditionally been created by typing, and photos taken one by one as someone presses the button on a camera. Now a big part of web page creation is semi-automated and devices like

15.9.2010

security cameras are always on and constantly creating data. The growth of data is no longer limited to active human creation during the working day, rather network and other services are always on – regardless of whether humans are still at work.

Computers, phones and other digital devices and sensors around us are capable of collecting, storing and transferring data automatically. These objects are not limited by the speed of typing – rather they can process and create new data much faster than humans. The quantity of these new kinds of data sources is growing as this capability spreads, irreversibly, to new product categories. A lifelong log of human individuals and also product life-cycle management will be possible when everything can be - and *is* - connected easily. The next step in this development – already in its research phase – could be lifecasting - the continual broadcast of events in a person's life through the medium of the Internet made possible by wearable technology. Such kind of development can be combined with the existing logistics systems which already are controlling the flow of physical goods and with the traffic systems which are tracking the movement of both cars and travelers. Again new data related opportunities and risks will appear.

New regulations in society demand a new kind of data collection and production of records for new kinds of tasks such as waste management and the circulation of products and materials. These requirements are typically implemented only when there are means and solutions for automation. Once more we see the pressure driving society towards automated data creation and management increasing.

Securing the availability of data also requires backups, various versions and duplicates; and this is done automatically - in addition to the creation of the metadata which is needed for automatic processing. The following email example indicates the relation between the original data and systemic data usage in a simple communication case.

A DAY IN THE LIFE OF AN EMAIL

How the digital universe feeds upon itself can be seen in the digital footprint created by a single email sent to a team of four people, an example based on an email infrastructure similar to IDC's (see Figure 7).

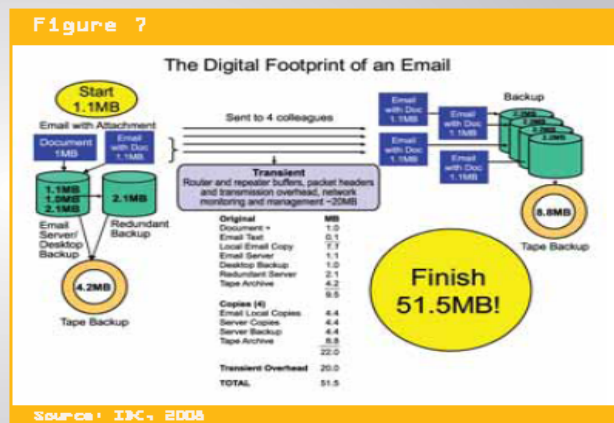
The email itself is small, but with it is a 1MB attachment. If the email is sent to four people, wouldn't that mean that there are 5 x 1.1MB involved? The original and four copies?

No, unfortunately. To begin with, there is the document itself stored on the local machine, then the email that contains the document. In this infrastructure, copies of all emails are kept on the central email server, which, in order to keep the email system up and running, includes a redundant server. Desktop files, where the original document sits, are backed up daily to a server. The servers are then periodically backed up to tape and taken offsite. Our original 1.1MB email has a footprint eight times bigger than itself.

Now add up the local and backed-up copies of the email sent to the four colleagues, and that footprint is 30 times larger than the original email.

Then there is all the temporary data created as the emails and backup systems send data back and forth across the local and wide area networks. In transmission, all manner of communications overhead is introduced: signaling data, packet addresses and headers, security codes, router caches, and management and tracking information. The estimate here is admittedly fuzzy, but it is within the order of magnitude.

There are techniques for deduplicating redundant emails and multiple copies of documents, but they aren't widely spread yet. In the meantime, a simple email can have a very long shadow.



Source: The Diverse and Exploding Digital Universe, an IDC White Paper

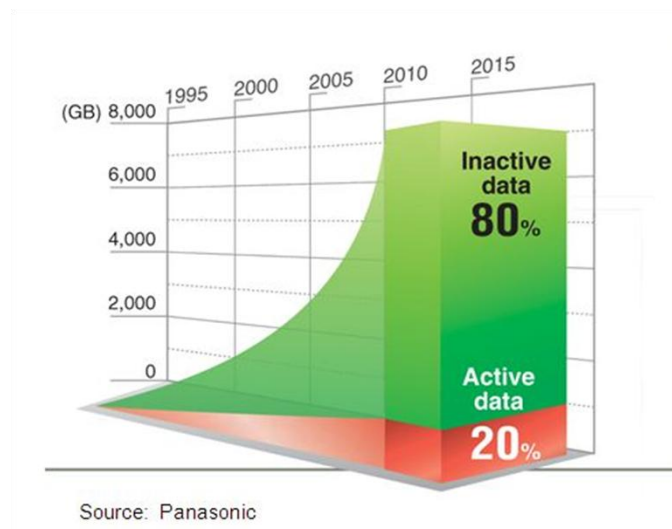
15.9.2010

Today most of the data in data reserves is created by machines and it is this portion which is growing fastest. Human processing of this data is no longer possible - instead we need more computers for control as well as better algorithms and applications for sense making. Most of the data will never be seen by a human being.

3 DATA CHAOS WITHOUT ORGANIZATION, ROLE OF ONTOLOGIES AND METADATA

*Data is not information, Information is not knowledge,
 Knowledge is not understanding, Understanding is not wisdom.
 (Cliff Stoll & Gary Schubert)*

As the example of an email (see above) indicates, nobody controls or manages the data storage from a holistic point of view. We have not agreed upon responsibilities or organized the storage work - neither is there any global data storage architecture available. Actually a big part of data is inactive and not used – a fact illustrated by the following figure:



Data needs to be organized better. Metadata is a concept that applies mainly to electronically archived or presented data and this concept is used to describe the definition, structure and administration of data files – with all contents organized in context to ensure the further use of the captured and archived data. Metadata can be created either by automated information processing or by manual work – but its sheer volume is so large that automation of metadata creation is the only productive and possible way forward.

Making sense out of data requires both ontologies, as well as new, improved methods for “communication” both between humans and machines and between machines. An ontology is a formal representation of both the knowledge by a set of concepts within a domain and also the relationships between those concepts. It is used to reason about the properties of that domain, and may also be used to describe the domain itself. Ontology engineering (or *ontology building*) consists of the methods and methodologies for

15.9.2010

building ontologies. It studies the ontology development process, the ontology life cycle, the methods and methodologies for building ontologies, and the tool suites and languages that support them. Dozens of published ontologies exist for subdomains, but no one is currently “controlling” their creation and management.

Data Governance refers to the operating discipline for managing data and information as a key enterprise asset. This operating discipline includes the organization, processes and tools to enable the establishing and ability to exercise decision rights regarding the valuation and management of data. Key aspects of data governance include decision-making authority, compliance monitoring, policies and standards, data inventories, full lifecycle management, content management, records management, preservation, data quality, data classification, data security and access, data risk management, and data valuation. (Source Nacsio: Data Governance – Managing Information as an Enterprise Asset).

As the amount of stored data expands, the consumption of energy and other natural resources also increases. Even the availability of the memory can be questioned. Due to these and other factors there will be even stronger pressure to develop much better data management theories, standards, practices, applications and services than those we have today.

4 RANGE OF PROCESSING DEVICES: FROM SENSORS TO SERVER FARMS AND CLOUD COMPUTING

*What counts is not necessarily the size of the dog in the fight; it's the size of the fight in the dog”
Dwight David Eisenhower*

Due to the continuation of effects like Moore’s law we enjoy better computing and see more and more different types of devices, gadgets and applications - which are not only connected, but also contribute to gathering and sharing of data.

Embedded devices can, for example, measure and monitor everything – though these measurements should be transmitted and processed in order for them to have value. This paradigm is called *ubiquitous computing*, *pervasive computing*, or *ambient intelligence* – with each of these terms emphasizing a slightly different aspect. Over the last two or three years, a new computing concept known as *motes* (or alternatively as *smart dust* or *wireless sensing networks*) has gained immense popularity. The objective of smart-dust devices, or *motes*, is to integrate sensor, power-supply, computing, and communications functions into a single silicon platform which can then collect, analyze and store data and then create an intelligent response. Sensors may include one or more temperature, pressure, vibration, acceleration, light, magnetic, or acoustic devices. As the mote collects sensor-data samples, it could provide the raw data for analysis elsewhere.

The smaller devices become, the more critical power consumption will be. Power consumption is a central design consideration for wireless sensor networks - whether they are powered using batteries or energy harvesters. Vital to success for either approach, however, is the need for hardware that uses power intelligently. Data processing in the sensor node is not the ideal solution for every application – though it will be for some. In addition to power consumption, other factors (such as latencies) also need to be taken into account.

15.9.2010

At the other end cloud computing (Internet based computing, where shared resources, software, and information are provided to computers and other devices on demand) does increase processing performance. The second goal of cloud computing is to provide a significant improvement in the cost-effective, elastic provisioning of IT services. As a result, two of the key characteristics of cloud computing are the further consolidation of servers into centralized data centers, and the virtualization of those servers. Because of the consolidation and virtualization of servers, the deployment of cloud computing results in additional application traffic transiting the WAN (Wide Area Network). Using clouds requires sophisticated methods to avoid extra costs which can be incurred due to factors such as additional traffic or storage needs. It should also be noted that data processing and transferring of large blocks of storage present new security related threats.

Data must move seamlessly between clouds in order to make it viable to use multiple computing clouds for a single task. An examination of hardware trends in storage systems reveals that, despite exponential growth in the capacity of media over the past decades, it is becoming increasingly unlikely that, without the arrival of some disruptive technology, this rate of growth can be sustained for single storage units such as disk drives. Instead, the high performance storage industry is applying distributed storage clustering approaches with great success. It follows that data must also move seamlessly between the increasing numbers of memory units.

The range of objects with various computing performance levels and the distribution of computing are under continuous development. When this is combined with the increased need to communicate and manage data and its location in network securely, we can define new needs for future data processing and networking architectures and protocols; what should be done on the edge and what should be done at the core of a network. A transition towards data reserves and their optimized utilization will bring up new paradigms and challenges which will require new technologies and thinking, demand new competences and open up a huge range of new business opportunities.

5 VALUE ADDED APPLICATIONS BY DATAPROCESSING

*A computer would deserve to be called intelligent
if it could deceive a human into believing that it was human.
(Alan Turing)*

Google has shown that when you have lots and lots of “stupid data” you can suddenly find valuable ways to use it. Sometimes the manual collection of free data can turn to valuable service as in Google’s case the collection of cellular base station data, which can be used in fast mobile location, or the collection of street views. The very basic map data is another example of a form of data which is freely available – with everybody (at least theoretically) able to go and collect it. Currently there are various company or social actions in use in the basic data collection.

An often-cited example of this value finding phenomenon is the detection of influenza-like illness in a population by using online web search queries, which are submitted by millions of users around the world each day. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region, with a reporting lag of about one day. Another example – again from Google - is the automatic translation from one language to another. The Google service does

15.9.2010

not apply grammatical rules, since its algorithms are based on statistical analysis between large data sets containing the same texts in different languages. The traditional rule-based analysis is replaced by simple statistical methods.

When there is meaningful seed data available, then it is easy to link more data to the same data base – this opens a long lasting avenue to both add value to the data and also develop more intelligent algorithms to extract new findings of the data base or data reserves. Clever data processing algorithms give possibilities to differentiate. Lots of additional value will be generated by those who are ready to dirty their hands getting data in, as well as in storing and processing it - so that it can be pushed out in a valuable form.

As databases grow over time, it will be more expensive to build another similar starting from scratch. In turn, if one does not have full access to data reserves, it will be impossible to develop intelligent algorithms and applications concerning reserves and thereby impossible to create new value adding services, or *spirals of success*. Access to the most meaningful data reserves of mankind will be a life-and-death issue not just for the ICT industry, but also for many countries – particularly those like Finland, which cannot secure volume access to reducing natural resources like oil.

The current open data discussion is particularly relevant because there is a need to break down those silos where data is currently stored. Adding value is about interacting with data in rich, semantically meaningful ways – going beyond a straightforward search to create the path from data to information and to knowledge. Significantly more R&D is needed in data mining, machine learning, computational, semantic and knowledge-discovery techniques in order to enable innovative ways of analyzing, organizing and retrieving information.

Data with clearly labeled semantic meaning is a tiny fraction of all the data in the world. Novel algorithms and techniques should be developed for improved extraction of structured information from unstructured/semi-structured sources, which may contain, for example, noisy data – but lots of redundancy. The greatest challenges are likely come from the needs to automate analysis, and the need to ensure the possible fusing of other relevant data. Therefore it is absolutely vital to invest in research to promote algorithmic advances.

Adding value to the service will help in the work of gathering data and this will open new possibilities to further develop service. Even stupid data can be combined with some innovative approach and turned to a valuable service. The access to data should be seen as the enabler for innovations.

5.1 Search as an example of development

Peter Norvig, the Director of research at Google has gone on record to illustrate how value can be added in the search. He points out how in 10 more years, the results we get back will be a synthesis, not just a list. For example, if one was to do a search for 'compare approaches to nuclear fusion', the major search engines would agree that a general encyclopedia article on fusion power comes first, followed by other similar articles.

A decade from now, the result will summarize the major approaches, contrast their differences, automatically translate any foreign documents into the user's language, and then rank the results by efficacy or place them in a table or chart as appropriate. If one was to then ask for 'background mathematics for fusion theory', one will get an outline for an impromptu course concentrating on the necessary complex analysis, customized both to specific applications in fusion and also to the user's level of

15.9.2010

mathematical understanding. If one is to slip up on such a course then the course would be readjusted to fit the student's needs – or, alternatively, perhaps the search engine will connect the student to a tutor or another student in a similar plight. Interaction with search engines will be an ongoing conversation; one that is integrated with the other ongoing tasks of our lives.

6 SOME SPECIAL PROPERTIES OF DATA

*A child of five would understand this. Send someone to fetch a child of five.
(Groucho Marx)*

6.1 Data ownership

Data ownership refers to both the possession of - and responsibility for - information. Ownership implies power as well as control. The control of information includes not just the ability to access, create, modify, package, derive benefit from, sell or remove data, but also the right to assign these access privileges to others. Data has intrinsic value - as well as added value as a byproduct of information processing. At the core, the degree of ownership (and by corollary, the degree of responsibility) is driven by the value that each interested party derives from the use of that information.

Today there are no good global rules for data ownership (actually ownership in the same way as ownership of physical objects is not really possible, only temporary usage rights are protected legally), which is why the first phase the data reserve rush will happen according to the adage in which “the early bird gets the worm”. When this happens societies will need to redefine ownership or usage rights related rules, regulations and legislation.

With time, better rules and standards for data usage rights, sharing and roaming will inevitably be introduced - but this development will take time because there are considerable privacy, financial and even national interests involved. It follows that the data reserves ecosystem is broad - covering the range from end users to national governments. Monopolies will not be accepted in the long run and that is why we can expect that the distributed model will win out.

6.2 Data security and privacy are key issues in Cloud Computing environments

Server farms and increasing network bandwidth make it possible for clients to utilize high quality services which reside solely on these remote central hubs. These services will often include data storage and processing. This possibility, however, forces clients to re-think pre-existing data protection schemes.

With the original Internet, which connected end hosts, the data storage and processing occurred within the secure resources of these end hosts, with the network simply providing transit. It follows that data protection could be done using privacy and security evaluations at the known intersection points of a data transaction. However, this is no longer the case when computing in the Cloud – particularly when utilizing hybrid Cloud architectures with several Cloud infrastructure providers. There multiple third party actors – the Cloud service providers – provide storage and processing resources to the consumer. Thus, additional

15.9.2010

actors will store or otherwise touch massive amounts of information – much of which the consumer might consider private, confidential, or otherwise sensitive. The same actors provide similar services to other customers using the same Cloud infrastructure and their data may reside in the same servers. Today the data can be processed in one Cloud, tomorrow perhaps in another.

6.3 Data is easily spoiled

One should remember that:

- Data privacy cannot be recovered once compromised;
- Unique data cannot be recovered once lost;
- The integrity of data cannot be restored if written improperly;
- The movement of data requires network resources and introduces latency if done on demand;
- The processing and moving data requires always energy and is prone to errors; error prevention mechanisms increase bandwidth and processing needs.

6.4 Summary

In this chapter we have mentioned a few special properties of data. The intention is not to cover everything here, rather to draw attention to the fact that the essence of data is not yet well defined – even though we are speaking about the information society. Lots of new interaction even between law makers, businesses and consumers is necessary before we will have a proper and accepted framework for data reserves. It goes without saying, however, that society cannot step back and wait until everything is ready – if we did that business would also disappear. Instead, the only route to success involves being both open and proactive and learning faster than others.

7 HORIZONTAL APPROACH AND ARCHITECTURES

*Dancing is a perpendicular expression of a horizontal desire.
(George Bernard Shaw)*

The history of computing is the success story of a general purpose technology approach, in which the rapid technology development reduces the cost and allows high volume.

In general horizontal structure has several important characteristics. These include:

- It is extremely important to work with wide range of applications and other organizations
- Standardized and open interfaces have central importance
- Interoperability is needed
- The size of the company is not particularly significant. Having a competitive core technology or know-how is more important than being big.

Cloud computing can be one step further in this direction – although current clouds are more or less proprietary and the need for intercloud technologies and standards is not yet that visible. Because society does not accept monopolies, there will be a need for data roaming – and consumers want to transfer their

15.9.2010

data from one service or cloud provider to other. This new wave of cloud evolution can not only offer new business opportunities – but also requires new development and innovative thinking.

Sensor technologies and RFID solutions will phase the same trend; full benefits will be available only with horizontal solutions and architectures which allow the interoperability of data at more advanced level than today. Also the transition from private mobile radio (PMR) to cellular is a historical evidence of the horizontalization phenomenon. Rather soon we can expect the transition from proprietary video control systems to horizontal video control networks. Open APIs have proven to be useful in many areas.

Because ICT development is so fast, a proprietary approach in the certain vertical is usually the right way to start and enter the game. However, it is very typical that somebody else has also innovated the same opportunity and that horizontalization is just a question of time. New architectures as well as new business initiatives need to be planned to ensure flexibility in the face of this obvious discontinuity. Those that are leaders in horizontalization are those that stand the most chance of becoming the long time winners.

8 THE DATA RESERVES OF THE FUTURE

*“For tomorrow belongs to the people who prepare for it today”
(An African proverb)*

As earlier already mentioned, instead of the old fashioned gold rush we will soon witness a “data reserve rush”, when various players enter the game and seek practical “ownership” of such reserves. The first phase of the “voyage of discovery” has already begun – with players like Google (Internet and media data), Facebook (social data) and Youtube (video material) successfully collecting their reserves for certain data categories. However just a few of the necessary data categories are under active, visible collection at present, with many more waiting to be discovered and transformed into spirals of success.

A huge portion of the data is now in silos: enterprise data, financial data, health data etc. Now it is important that the content of those silos is analyzed, as this will reveal what the most immediate and potential data categories for horizontalization activities actually are. Breaking down the silos provides a way to increase the shared value of data by creating an ecosystem in which data consumers can interoperate with data producers in a way that is far more powerful than currently possible – thus enabling an increased number of applications to make sense of a broader set of data. It follows, in turn, that every producer and consumer of data participating in this ecosystem increases its overall value.

The future data reserves are numerous, and include areas such as personal life log, product life cycle data, updated traffic data, access rights data etc. Actually everybody can identify some new categories. However various data categories have different value. Bearing this in mind, it is important that we start to collect data and get a foothold in those categories of data reserves which have an important role in the future.

15.9.2010

9 FINLAND AND DATA RESERVES BUSINESS

*Insanity: doing the same thing over and over again and expecting different results.
(Albert Einstein)*

Finland is a small, safe country at the very center of current western civilization. Finland is also a neutral and technically advanced country – with a good or arguably leading position in many global rankings. While Finland may geographically seem like an island, the Internet world allows us to build as many bridges as we need to serve Europe, Russia, USA or China or any other global market. Finland's strong bedrock, clean water, secured energy availability and even empty factories are also assets in the data reserves business, where safety and stability are important.

Finland has long focused on telecommunication and mobile computing and gained the current ICT position as a direct result of this. Parallel to this competence we need to swiftly build data reserves competence and business capability in order capture the meaningful part of the future data reserves business to Finland. Our strong mobile competence should be and can be used as a real strength in this game and data reserves can support the development of future mobile computing. However, the essential development in data reserves is strongly server or cloud computing driven not terminal driven.

The main focus of Finnish industry is not on the consumer business, with industry instead being focused on B2B (business to business). It is, therefore, important that this is utilized as a stepping stone when entering the data reserves business. Every strong industry vertical should seek business opportunities related to global data reserves. At the same time, however, we should also remember the horizontal approach, and cooperation between industry verticals is welcomed. Data is data in every industry. Parallel to the horizontal approach, some selected and limited topics could be created in order to go to the consumer business mode to guarantee learning of its peculiarities.

Finland has also very strong and advanced banking system and other areas, where we have excellent current strengths, which can be turned from the current vertical approach towards horizontal data reserves business and services. Our education system is one of the bests in the world and education related data management can be turned to global success story. And as a curiosity we have the world's most advanced ship tracking system on Baltic Sea, which offers example for future ship tracking data reserves and related service operating opportunities on global level. Let's use our strengths fully and be innovative in opportunity identification. The competition will be tough and the good set of niche markets should be appreciated, too.

Finland is a small country and a successful team player in EU and on global level. Distributed architectures and advanced data roaming suit well for this kind position and should be developed as new great Finnish strengths. The mobile competence background is also a very good starting point here as the best data roaming solutions are implemented in cellular networks. The Finnish success achieved in cellular technology was partly gained due to the Nordic cooperation between countries at a similar advanced development level. There is lots to learn from this experience now, when in addition to the mobile phone data there are needs in huge range of various data categories to be opened for Europe-wide or even global use.

The existing innovation strategy and innovation competence should be applied actively in the development of data reserves and the new SHOK network should be used as a tool in this very broad push to improve competence and business development. Currently data reserves play an important role in Tivit's programs. Cloud SW -program focuses, as the name indicates, in cloud software architecture and applications. Device

15.9.2010

and Interoperability Ecosystem -program (DIEM) aims at developing methods that enable combination of real-time data in different devices into meaningful applications. As an example, DIEM-program has published Smart-M3 solution for semantic level information interoperability, which is needed to effectively combine different data coming from different devices. Next Media -program is basically all about media content. In Co-operative Traffic -program data fusion methods are being studied and novel applications methods developed. In Future Internet -program one research topic is information networking, which may in the future cause a main change in the data routing paradigm. Tivit also has initial plans of a new program, entitled as “From Data To Intelligence”, which would be targeted to data reserves.

It is clear, however, that no single player is capable of taking responsibility in the national data reserves rush - instead we need innovative efforts on every level. One great step towards securing a lead in this area would be to appoint a Data Reserves Minister instead of Communication Minister when the next Finnish government is formed.

10 RISKS

*You'll always miss 100% of the shots you don't take.
(Wayne Gretzky)*

In Finland we have the risk that our ICT competence base is not strong enough in cloud computing and other core technologies which are needed for the successful data reserves business. However, we are not too late and when acting fast we can reach the leading position in still open business areas. Tivit has shaped its program portfolio from embedded computing towards future internet, services and cloud computing and will still continue to strengthen the efforts and focus on necessary data reserves competences. However, more national actions are needed to build the strong momentum.

Finland has no experience of being a part of an imperialistic resource rush – but now is time to take part in the global “data reserves rush”. Often the business development in Finland, where the domestic market is small, tends to take time but the data reserves business game will be global by character. This is why we should combine our strengths also in business development with dynamic business arrangements. Also the cooperation between the existing globally operating Finnish companies should be considered over the constraints of various verticals.

Although the new technologies e.g. in cloud computing will be in a key role as enabler the real core competence is in the aggressive business development. Government can support the revolution with some fresh actions, which will move the overall thinking to new paradigms like the above mentioned idea of Minister of Data Reserves, and also by encouraging Tekes to shift funding focus from R&D to business development. However, industry and companies should not wait for anything but prepare their data reserves strategy and action plans immediately.

The world is changing fast and presents us with a range of new frontiers. Finland has excellent strengths, which leave it in a strong position to capture a significant share of the future data reserves business – in the same way it was able in the field of mobile communication. Now is time act without hesitation, not only in selected companies but on a far wider basis - because data reserves will play a key role in every industry and public service.

15.9.2010

11 CONCLUSION

*“There are fish in the sea better than have ever been caught”
(Irish saying)*

Data and data reserves will be a fast growing business area of vital importance as human society enters the new phase in the knowledge society development and starts to create the next generation of global services. Data reserves have the best environment for development in the cloud computing systems - where limitless computing power is available for advanced data processing and new value adding services.

Finland has now a golden opportunity to take the leading position in data reserves development. This requires clear strategy and committed players. Smart early movers will gain the best positions. We must join the leaders who are innovating and creating the new world!