The Global Cross-Reference Data Base –
a Tool for Information Logistics in e-business

Mag.art. Helge G. Simonsen
Siena AS, Norway
helge.simonsen@siena.no

Paper presented at the 5th eGov-conference
University of Antwerpen June 16. – 17. 2005

Abstract

1. Background

A number of information classification systems are related to different aspects of international trade, e.g. customs, e-business, statistics, special interest etc. Most of them live in splendid isolation from each other. E.g. the international customs authorities have one system (HS) which does not “speak” with other commerce “languages” like CPV or UNSPSC, thus making it necessary to fill in information manually multiple times – although it could have been done by computers only once and at very low costs.

Lack of information standardisation is a major problem in e-business, especially in the context of e-catalogues – the backbone of e-business.

This is a recognized problem and is sought addressed through different means.
• Competition between the nomenclatures.
• Creation of a super-mega-metadata system with semantic interoperability covering all interests

2. Suggested action

There is a third way, i.e. the establishment of a Global Cross-Reference Database for all major nomenclatures. As they have been created for different special interests, there is no reason why these interests should be given up when the nomenclatures can be correlated and thus “speak” with each other.

The governments should have an inherent interest in establishing cross-references between different nomenclatures as a part of their e-government programmes in order to promote and facilitate efficient e-business.

The data base should be managed by a pertinent international body, e.g. the OECD.

3. Benefits

A database giving correlations between the major international nomenclatures will give enormous advantages to both private and public (government) sectors in e-business.

The correlation data base can
• eliminate the need for multi-entering the same information,
• reduce errors in e-business
• give better e-catalogues
• automate and speed up transaction processing
• reduce transaction costs
• give better statistics for public authorities on e.g. electronic trade
• give enterprises a tool for analysis on their spending
• provide transparency in the procurement process

This will especially help the SMEs who lack expertise in dealing with such advanced aspects of e-business.

An extra bonus will be exchange of trade information between organisations and countries that fight illegal capital flight, organised crime, money laundering, fraud and terrorist financing.
1. The Impact of Globalisation and the Internet on Commerce

The combination of the Internet and globalisation has had a profound impact on commerce and trade in the industrialised world. Electronic commerce is now a well-known concept for both the big industrial companies, SMEs and the individual consumers. A new supply value-chain is developing, changing the way we buy goods and services whether one wants a new book or a 5-year contract on producing bumpers for a new car-model.

It is impossible to say if globalisation is a practical consequence of the Internet, or if globalisation has helped to increase the commercial usage of the Internet. Probably we are dealing with forces promoting each other. Together the forces are (inter alia):

- Reducing or altogether eliminating manual processes in commerce and trade, thus increasing productivity
- Integrating or eliminating hitherto independent manual and electronic processes
- Changing the business models of trading parties, promoting networking

We are taking part in perhaps the fastest and most radical process of change in the history of mankind. I order to understand what is going on, it is necessary to pay a short visit to the Spanish sociologist Manuel Castels, who has coined the concepts of “informationalism” and The Network Society. Castels (vol. 1, 1996, p.17) claims that “knowledge and information are critical elements in all modes of development, since the process of production is always based on some level of knowledge and in the processing of information”. Hence “informationalism” is oriented towards technological development, that is “toward the accumulation of knowledge and towards higher levels of complexity in information processing”. This is the back curtain for this paper.

2. Information, Logistics and Infrastructure

Information is a commodity necessary to fulfil any commercial process in a rational way for both vendor and customer. Timely, reliable and relevant information is a *sine qua non* in all commercial activities. What is the customer going to buy from the vendor, and what are the properties of the respective goods or services? Are the producers and their representatives reliable and credit-worthy, are the goods meeting required specification, is the quality up to standard, will the goods be delivered on time, has information arrived in time to make the correct commercial decision etc.

This becomes especially important in both e-commerce and e-business where transactions to some extent are computerised and not scrutinised by humans unless some irregularities are discovered. If the information is not good enough, e-business systems are just speeding up the process of making bad decisions, losing millions much faster than a manual process will be able to do.

In a small, local market where everybody know each other, good information will be a matter of course and readily available. But in a global market one cannot count on informal information.

Firstly: The amount of information is overwhelming and we cannot easily know what is relevant. SAS Institute assumes that the amount of information on the Internet doubles every 18 months. Professor Umberto Eco has coined the concept of “[Information Paralysis](http://example.com)”. He says that when he is searching on the Internet for references for a paper on e.g. Thomas Aquinas, he gets tens of thousands of “hits”. He becomes paralysed and unable to relate to this mountain of information. Instead he goes back to his 10 – 12 standard book references on his subject.

Secondly: The quality of the information must be controlled. The “old” academic demands on reliability and validity are more than ever relevant. Formalized information from reliable sources must be obtained before important contracts are concluded. The information must...
also be made available to the interested parties online and quality controlled by independent parties.

Logistics is a concept originally developed by military organizations to describe the movement of troops and their supplies. Long ago it has gained a much wider meaning, covering also the civil society, including the production of goods and their movements till their end-usage. Some will also include e.g. the eventual demolition of buildings as a part of the logistics process. Logistics has gained even wider usage in the information age. We speak about information logistics, i.e. the distribution of information for a specific purpose to a specific audience at a specific time.

The concept of infrastructure has traditionally meant the basic facilities and installations necessary for inter alia transportation of goods and people and simple communication between people/organizations (e.g. roads and telephone lines). This used to be a public responsibility, but in the Western world it has now mostly become privatized or is in the process of being privatized.

However, the meaning of infrastructure is now also covering the facilities for transporting information. E.g. SIMAP (2005) defines “its objective is the development of the information systems infrastructure needed to support the delivery of an effective public procurement policy in Europe, by providing contracting entities and suppliers with the information they need to manage the procurement process effectively.” SIMAP is the acronym for Système d’Information pour les Marchés Publics. It is a EU project with the mandate “to encourage best practice in the use of modern information technology for public procurement. Initially the project aims to improve the quality of information about the EU procurement opportunities and ensure that information is made known to all potentially interested suppliers. In the longer term it will address the whole procurement process, including bids, award of contracts, delivery, invoicing and payment.”

What is the infrastructure available for meeting these ambitious goals? It consists mainly of several classification systems of varying quality, relevance and scope. The main “roads” are the CPV (Common Procurement Vocabulary” developed in order to increase competition among suppliers to the public sector of EU, and NACE which classifies the economic activities of the EU/EEA countries. In addition there is a number of international, national and branch-oriented nomenclatures. Nobody knows the exact number of classification systems, but it is probably safe to estimate that there are close to a thousand in just the EU/EAN countries.

Most of these nomenclatures have inter alia some of these properties in common:
- not suited for e-commerce or e-business as they were made for other purposes
- undergoing and/or will have to undergo extensive revisions the next few years
- proprietary, i.e. not able to communicate with each other

The situation is no better in e.g. the United States or Japan. All industrialized areas suffer from the same lack of a modern, common information infrastructure for e-business.

This situation could be compared to having an unknown and unmapped number of dirt roads, starting somewhere and ending nowhere, with no connection between them and most of the roads in need of extensive repair.

The consequences of this lack of standardisation are serious.

Let me give an example: Any actor on the global trade scene have to know the implications of international trade in relation to different national and international regulations. What are the customs tariffs for this item? Are there any export or import restrictions or embargos? How can I comply with national and international reporting demands on e.g. trade statistics and competition information (like usage of CPV for public procurement)? If I do not have à jour information on these and other questions, it could cost me dearly in terms of unknown tariffs.
or taxes. Consequently, this lack of standardisation presents the trade partners with unnecessary financial risks unless they perform extensive manual examination processes. In order to find correct information, they must manually search through cumbersome paper catalogues of different nomenclatures. I know that many of them are in electronic format, but to me – as long as they are not integrated with other electronic systems – they are still “paper” – albeit electronic paper.

The Internet has in some ways improved the search for important trade information. If e.g. the different national customs authorities have made good homepages, it is easier to find relevant information directly as compared to (perhaps) outdated written documentation. However, the Internet normally provides a wealth of unstructured information that makes the customer more often than not, uncertain about what is the correct information. Experience has also shown that public authorities are not as good as desired to update their homepages. This does not mean that the searched-for information is not obtainable. Often it is, but in a combination with different information systems and sources that do not “speak” with each other. It is often fragmented and incoherent and impossible to use unless one has specialised knowledge.

The economical consequences of this situation is that costs for all trade parties are unnecessarily high in terms of cost of manual labour needed for finding and processing relevant information. Nobody has tried to calculate the magnitude of these unnecessary costs on the trading partners, but they are no doubt in the range of millions of Euros per annum.

We can paint a parallel picture for the public authorities. E.g. the Harmonised System (HS) was developed as a tool for implementing the different tariffs agreements on products and it has served well in this respect. However, it was not developed in order to regulate the explosive development of electronic trade on the Internet – whether B2C or B2B – which also includes services. The International Trade Council (ITC)’s insistence on only using HS in relation to two other outdated systems – CPV and SITC - does not improve the situation. Consequently, HS is irrelevant for much of the explosively expanding electronic trade. This goes especially for trade in services – whether they are common engineering services, gambling, money laundering or children pornography.

This means that the ability to impose correct tariffs and taxes on electronic trade is wanting and is probably resulting in loss of revenues. It also means that organised crime may be able to utilise this lack of system to their advantage on e.g. ecological crime, smuggling, money laundering and terror financing.

As none of the official nomenclatures are able to catch the development of electronic trade, we can safely assume that neither individual countries, nor regions like EU or the World at large – have reliable statistics on global electronic trade.

To sum up – the lack of mandatory and relevant international standards on e-business is probably the most important impediment to increased e-business, both Nationally and Internationally, as well as creating conditions beneficial to international financial crime. Every 18 months the global information volume is doubled, i.e. the lack of standardisation represents an accelerating challenge!

From a business point of perspective the lack of standards in the e-business sector are also costing trading partners (private and public) millions of Euros per annum in unnecessary costs, and simultaneously challenging public authorities claiming correct tariffs and taxes.

In Castels’ terminology, this is slowing down the increase of productivity of informationalism and also slowing down the development of the Network Society, thus creating an unhealthy imbalance between the different parts of the Global system. This is especially hurting the SMEs because of their limited competence on e-business.

3. Building The Information Infrastructure
The need for an Information Infrastructure has been recognised for many years (although this term is not normally used). The term recently in vogue has been and is interoperability. Different initiatives have been able to maintain interest in this important question, albeit no significant and practical progress has been made. As of today, the term semantic interoperability seems to be the most popular. The EU and CEN are very much interested in this, and e.g. CEN has organised laudable workshops on interoperability and electronic catalogues.

Several of the major information technology companies have already taken action and started building an information infrastructure which is simple, free and readily available for everybody. In many ways it started with XML – the extensible mark-up language that simplifies information exchange and processing, especially in relation to e-commerce. XML has led to the concept of Web-services – a system of standardised information services or applications that can be accessed through open standard Web protocols. The most important in this context are the SOAP protocol, the UDDI and (unofficially) the UNSPSC nomenclature.

In addition to the Web Services there are a big number of building blocks which can be integrated into the information infrastructure. The multitude of taxonomies, nomenclatures and standards that have been developed for different purposes, can all play important roles in this context.

Among the most important are EU’s CPV, the GS1 bar codes and location numbers, GP1’s GPC (Global Product Code), UNSPSC and ECCMA’s eOTD (eccma Open Technical Dictionary).

The different nomenclatures are constructed in different ways, based on different principles and methods in order to serve their different tasks. Some are strictly hierarchical, other are less systematic. Some take their point of departure on e.g. function of the goods, while others may use material content, etc. In general, the different nomenclatures/standards have no common structure nor do they agree on semantics or synonyms or ontology. In short: They are proprietary. Thus the quest for semantic interoperability has become very hot. The same word may have different meanings, depending on which nomenclature and which context it is used in connection with. Also seemingly different words may have the same meaning, depending on the context. An additional problem is related to translations that may vary in quality.

However, semantic interoperability covers only parts of the problem. Structural interoperability is just as important.

Structural interoperability relates to how the different nomenclatures are constructed. Are they constructed according to a systematic, hierarchical plan, or are they built “by inspiration” catering for on the on-the-spot needs of a kind of system, or are they hybrids of both approaches? Or may they perhaps be object-oriented?

Structural interoperability can relatively easily be achieved if the different nomenclatures are constructed on the same or similar principles. Some basic principles are formulated by Hoffman and Chamie (1999) “Classifications group and organize information meaningfully and systematically into a standard format that is useful for determining the similarity of ideas, events objects or persons. The preparation of a classification means the creation of an exhaustive and structured set of mutually exclusive and well-described categories, often presented as a hierarchy that is reflected be the numeric or alphabetical codes assigned to them”. Too many taxonomies do not follow these simple basic principles.

The major interoperability problem is thus de facto a question of metadata (data on data). What system shall have the power of defining the rules of play within the area of electronic commerce? One strategy (it seems) has been to make one standard the only standard. Some of the proponents of special taxonomies claim that “their” taxonomy is the universal one, superior to all other and to be regarded as the meta data system. They do not take into account that a taxonomy that is good for one purpose may not be useful for another purpose. In some instances this attitude has resulted in silly, almost “religious wars” between
nomenclatures, creating confusion among both administrators and politicians with no expert knowledge on this question. Such confusion has without doubt created setbacks in the development of electronic trade.

It should be no surprise to conclude that the different attempts to make one nomenclature king of all other nomenclatures have failed. This is for both practical, professional and political reasons.

The practical reasons are sound and down-to-earth. Why should e.g. the customs HS system be replaced with e.g. EU’s CPV, when the systems have been developed for different purposes and functions? There are no sound professional grounds for that.

The lack of political reason is even more obvious. Having its own classification system is a symbol of importance for an organisation, of ability to define the area of autonomy – of defining the rules to play with. It is a question of both symbolic and real power. And who wants to give up power? EU certainly does not. It seems to insist on making the current CPV a mandatory nomenclature for e-business. However, the commission has a problem. It may decide that CPV is suited for e-business, but from a professional point of view “everybody” agrees that it is not. Shall political considerations overrule the professional expertise? If so, it is a bad omen for the development of e-business within the EU.

The second way has been to try and develop a new super meta data system overriding and replacing all the “old” systems. I will not say that this is impossible from a technological point of view. However, the political aspects remain the same. Why should one give up its power base without rewards? A new and unknown system gives no reward in itself. It will take long to persuade all the stakeholders of special-interest systems that they should give up their well-known systems in exchange of a new and untried system.

It will probably take even longer to develop a sound data model and architecture for a new meta-system replacing all other nomenclatures. To me this seems to be the modern parallel to alchemy. The practical world is so unstructured and so multi-faceted that it cannot be forced into one set of principles. And it is changing all the time!

These attempts of making one taxonomy the superior one can be compared with trying to replace the old dirt roads with either a brand new super highway or to make one of the old roads mandatory for everybody. This will of course not work – neither with transportation of goods on roads nor with transportation of information.

Hence I conclude that the way to overcome these obstacles is neither to make one current nomenclature “king of the roads”, nor to try and develop and negotiate a new meta data standard. In the first place it would be both political and practical impossible to accommodate all the different special interests. Secondly, it is not necessary. There is a third way.

4. The Global Cross-Reference Data Base

An information architecture based on a Global Cross-Reference Data Base will address the majority of these challenges and represent a core component of a business intelligence platform for information exchange based on all major nomenclatures, National or International, open or proprietary. A Global Cross-Reference Data Base will enable universal correlation and qualitative analysis of the respective systems and serve as a “bridge” between the different proprietary systems.

It would appear obvious that governments should have an inherent interest in the establishment of such a cross-reference between different nomenclatures, as part of their e-government programmes, in order to promote and facilitate efficient e-business in addition to reducing their own costs. However, this will not necessarily be the situation as governments are not always ruled by wisdom.
The figure above shows the principle of the Global Cross Reference Database (GCD). The database should give seamless integration between the most important international nomenclatures/standards like CPV, UNSPSC, GS1/EAN/UCC/GPC, UDDI, STEP/ISO 10303, CPV, NACE, NAICS/SIC, HS, SITC etc. and also correlations with important national industry nomenclatures. This will help both transparency in international trade as well as facilitate the integration with other business processes.

The benefits of such a database should be obvious. An illustrative example is correlation between HS for customs, UDDI for identifying the companies involved, CPV for identifying purchasing needs of the public sector of EU, and UNSPSC for goods and services in e-business.

The point of departure is a simple supply chain of trade. The vendor is e.g. a municipality in a EU country using the TED data base and the CPV code for identifying its purchasing needs. With correlation between CPV and UNSPSC, the vendor automatically can get access to information on relevant suppliers both within and outside the EU. One can thus get a long-list much more detailed and relevant than CPV alone is able to give. From this long-list a short-list can be made based on financial information from the respective national registers, combined with information from e.g. UDDI and/or EAN Location Codes.

When goods/services have been ordered electronically using UNSPSC and CPV, both buyer, vendor and the customs authorities will benefit from database information on which tariff should be used for specific items as well as import and export regulations and possible embargo situations.

This will make the total e-commerce process simpler, more reliable and less costly. The need to repeatedly and manually enter the same information in different systems will be eliminated. This will both reduce costs of manual processes and at the same time reduce the number of errors in the trade process. It will also contribute to the shift in power between vendors and buyers, enabling the buyers to be more active and searching the markets for suitable vendors, instead of sitting and waiting for the vendors to knock on the door.
The database would also increase transparency in the commerce process as well as being a tool for seamless communication and integration with national taxonomies. It could also contribute to eliminating the unnecessary and costly invoice, which in my opinion belongs to the “churchyard of history” because it is paper. Electronic paper is still paper. The concept of electronic invoices does not change that.

5. Who should develop and manage the GCRD?

Developing the GCRD is no big deal from a technological point of view. The technology is available and can be put into production on short notice. In Norway we have already started correlating UNSPSC with some national standards. When I was head of UNDP’s office for development and maintenance of UNSPSC’s forerunner (UNCCS) ten years ago, we made correlation tables between UNCCS, CPV, HS and SITC.

The big question is who should be trusted with the important task of developing and running this database?

An international body with no strings attached to special interests should no doubt be chosen. There are several candidates, e.g. WTO/ITC, OECD and SIMAP.

I think we can rule out WTO/ITC because they seem to be afraid of new ideas that may be controversial. We will need somebody who cares about this concept and will work enthusiastically to promote it.

SIMAP has no doubt the competence to manage and promote such a system, but may be crippled in its activities because of EU’s special political interests.

The UN system has since long decided to concentrate on its core business, which does not include nomenclatures.

This, I think, leaves OECD as the single actor on the stage of unpartisan promoter of efficient e-business through building a global information infrastructure. The alternative is to create a new international body, which I think is a long shot.

6. Industrial-Political Implications

The importance of nomenclatures and their standardisation in global e-commerce contexts have until lately been considered an exotic subject, drawing little or no interest from neither governments nor industry. However, with the recent manifest interest from international marketplaces and other e-commerce actors, there may be hope that this subject gets the attention it deserves. There are many hopeful signs.

A report from an EU Expert Group on B2B Internet trading platforms (EU 2003) finds that SMEs are lagging behind the larger corporations in utilising Internet trading platforms. That goes for both

• “traditional” e-commerce applications like home-pages and Internet shops
• e-business (integration of IT-systems, completing business transactions, providing information on goods and services (quality electronic catalogues) etc.)

The Expert Group expresses concern for this development since it will change the business structure of Europe, potentially making the SMEs of diminishing importance. The Expert Group concludes that the emphasis now should be on teaching SMEs how to develop into e-business companies. This will make them

• More efficient and thus more prone to survive and grow
• More able to adapt to the complicated business models of multinational companies

It will also stimulate

• Co-operation between SMEs both nationally and within the EU/EEA countries
• Utilisation of European electronic marketplaces as well as other markets.
The key to this development is electronic catalogues which again are dependent upon a good nomenclature suited for global electronic commerce.

Reference is also made to the work done by CEN. Its numerous activities – especially on e-catalogues – help building an information infrastructure and standards which will make it possible for SMEs to develop good electronic catalogues suitable for the different national and international electronic marketplaces.

This concept should also interest all parties involved in development aid programmes, helping the developing countries to access the global electronic market places which provide a simple and inexpensive means for companies in these countries, to be clearly identified, together with their goods and services, in the traditional maze and confusion of information on the Internet.

7. A Summary of Benefits

A data base providing correlations between the major internationally adopted nomenclatures with links to National taxonomies will offer measurable advantages to both the private and public sectors:

- Improve productivity, i.e. eliminate the need for entering the same information several times into different systems
- Increase data and information quality by reducing miscopies and faults by automating the handling of information input and transactions
- Improve critical business activities, i.e. speed up the processing of transactions
- Reduce processing costs for business partners and public authorities

It will also facilitate the collection of better statistics for public authorities on commercial activities, including services that are not covered today.

Both private and public enterprises will get a practical tool for their own spend analysis and integration of processes, thus eliminating the need for paper-based or electronic invoices.

Especially SMEs will get a tool for making good electronic catalogues, thus giving them access to both National and International markets.

This data base will also give improved transparency in the procurement process by giving solid information on how and why a product or service was purchased by whom – provided it is integrated in the financial systems of the participants in the supply chain.

Finally, such a database will have the added benefits of representing an efficient tool for fighting international financial crime by giving transparency on all registered commercial and financial activities.

References


www.uddi.org
www.wsinterop.org
www.oasis.org