Kim H. Veltman

Standard and Variant Names and Classifications: Cross-Cultural Challenges for Europe in a Global Context[1]

European Association of Directory and Database Publishers (EADP), Rey Juan Carlos Hotel, 19 September, 2002, Barcelona, 2002 (in press).
Cf. http://www.eadp.org/main2/main2.html

---

1. Introduction

Naming is a process as old as mankind.[2] Names play a central role in directories and databases. The challenge is straightforward: when we name things we need ways to find (back) those names efficiently. Enormous progress has been made in this context over the past fifty years. There are global trends towards convergence. There are new global solutions with respect to electronic data in general and names in particular. On the surface we are moving from strength to strength. Efforts which began in the mid 1970s as Electronic Data Exchange (EDI), have led to Electronic Data Interchange For Administration, Commerce and Transport (EDIFACT). It has also led to X.500, LDAP and more recently to a host of new acronyms, which promise to solve all our problems: SOAP, UDDI, WSDL, UCLP and XML/EDI.

The first part of this paper reviews these developments, which are described in more detail in Appendix 1, and outlines a trend whereby a) Europe tends to develop complex solutions which are then b) adapted in a simplified form in the United States and subsequently c) these simplified adaptations pose as solutions for Europe. We shall note dangers and challenges entailed in this trend before outlining a more comprehensive solution for the future.

2. Developments

Roger Clarke has traced the origins of Electronic Data Interchange (EDI) in the United States back to the time of the Berlin airlift (1948), although standardisation in this field did not begin until 1968 when a committee set out to coordinate the development of translation rules among four existing sets of industry-specific standards.[3] This led to the

X.12 and later the X.25 standards. According to Jason Levitt, Electronic Data Interchange (EDI) in the context of businesses communicating over a network was launched formally in 1975.[4] In the United States, widespread use of these methods was hampered by three obstacles: complexity, cost and lack of interoperability among competing architectures (e.g. Common Object Request Broker Architecture (CORBA), Distributed Component Object Model (DCOM), Unix Remote Procedure Call, and Java Remote Method Invocation).

Meanwhile in Europe, the United Nations, Economic Commission for Europe (UN/ECE) established a Working Party on the Simplification and Standardization of External Trade Documents in 1960. This led, in 1972, to a new group, which aimed at "a standard international trade data terminology and a uniform system for use in automatic processing and transmission of trade data. This group worked together with the forerunner of the International Telecommunication Union (ITU), and the International Standards Organisation (ISO TC 97/SC 6 "Data Communication")[5] to ensure interoperability. This resulted in **E**lectronic **D**ata **I**nterchange **F**or **A**dministration, **C**ommerce and **T**ransport (EDIFACT).[6] While this was an excellent contribution it took many decades to develop and was too complex for simple everyday use. By 1999, final rules for EDIFACT were still being discussed.[7] Nonetheless, it provided a high level technical solution to many underlying problems.

The rise of computers in the 1980s introduced a more specific, challenge: how to integrate the naming problems of telephone companies with those of computer companies. In Europe this led to new co-operation between the standards bodies, which were developing EDIFACT (namely, ISO and what is now ITU-T) and the European Computer Manufacturers Association (ECMA). The goal was to use computers to create an integrated approach to national and international directories using a standard interface. The X.500 efforts began in 1984 and were completed in 1994.

X.500 is a complete solution. It defines a global directory structure and also defines the Directory Access Protocol (DAP) for clients to use when contacting directory servers. This heavyweight protocol runs over all seven layers of the International Standards Organization (ISO). While thoroughgoing and excellent X.500 and its Directory Access Protocol still had three significant drawbacks. It was expensive, required significant computing resources to run and was complex to use.

In response to these challenges, individuals in the United States developed a Lightweight Directory Access Protocol (LDAP) originally as a front end to X.500, the OSI/ISO directory service. LDAP provided much of the functionality of the Directory Access Protocol (DAP) a much lower cost. As a next step they developed two more products:

1) a stand-alone LDAP daemon (slapd) to remove much of the burden from the server side just as LDAP removed much of the burden from clients.[8]
2) a stand-alone LDAP update replication (UNIX) daemon (slurpd) to distribute changes in the master slapd database to various slapd replicas.[9]

These LDAP developments were two edged. The good news was that there now seemed to be an easy solution to directory problems which sidestepped the high computing and administrative costs of X.500. The bad news was that a number of non-standardized implementations of LDAP soon arose. Of these, the most widely used implementation was written at the University of Michigan and developed with a National Science Foundation (NSF) grant.[10] To deal with this fragmentation an Open LDAP Consortium was founded.[11] Ironically, X.500, a serious solution developed in Europe, was taken up in a too simplified form in the United States, only to be taken up anew in Europe where it caused havoc because the compromise solution was too simplistic.

During the 1960s, at the time that the possibility of the Internet was being discussed, those in the text world faced a fundamental problem: the basic information in manuals frequently needs to be displayed in alternative forms for users at various professional levels with different needs. This led to the development of Standard Generalized Markup Language (SGML), which separated content from form in order to reduce the costs of printing different versions for various audiences.[12]

Not unlike the EDIFACT and X.500 solutions, SGML is a marvelous universal solution but much too complex for everyday use. The World Wide Web (W3) Consortium thus developed a subset of SGML better suited to the needs of everyday use, called eXtensibe Markup Language (XML). This provides a common denominator for regular content. Specialist users are writing extensions in specialized markup languages.[13] While the W3 Consortium is fully international, it has once again been individuals in the United States who have championed XML, a simplified version of SGML, and made it a center of attention and hype. As a result, XML, which began as a solution for the publishing world, has become intimately connected with the future of the business world. Closely linked with this have been four other key developments in the past years: SOAP, WSDL, UDDI, and XML/EDI.

The W3 Consortium which has developed XML, has also developed Simple Object Access Protocol, (SOAP), "a lightweight protocol for exchange of information in a decentralized, distributed environment, linked specifically at this point with the Hyper Text Transfer Protocol (HTTP) and the HTTP Extension Framework.[14]

In addition, the W3 Consortium is developing a Web Services Description Language (WDSL), an "XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information."[15]

As was the case with EDI in its early days, different vendors have developed a number of competing implementations each with its own flavour, thus bringing to the fore once again problems of compatibility and interoperability. And once again users are rediscovering the need for a centralized registry if various distributed, local databases are to be integrated. Here the latest candidate is Universal, Description, Discovery, Integration of business for the web (UDDI).[16] Meanwhile another member of the W3 Consortium, SAIC/Bellcore (now Telcordia) has submitted a proposal for a Universal Commerce Language and Protocol (UCLP). [17]

Parallel with these efforts is an attempt to combine the latest developments in eXtensible Markup Language (XML) with the specific characteristics of existing Electronic Data Interchange (EDI) to produce a new synthesis called XML/EDI. This provides a standard framework to exchange different types of data so that information "can be searched, decoded, manipulated, and displayed consistently and correctly by first implementing EDI dictionaries and extending one's vocabulary via on-line repositories to include one's own business language, rules and objects.[18]

Initially many of these very useful developments arose from the everyday realities and needs of the business world as transactions, which were traditionally local and paper based have increasingly become global and electronic. Increasingly, the needs of science and the military have also become reflected in these developments.

There are very good reasons for this. Business is becoming global. Science depends on the latest global standards. Science in its narrowest sense is concerned only with the latest developments, not with what came before. This approach is analogous to what we look for in telephone books: we want the names and numbers that are functioning today. We want live names of today rather than dead ones of yesterday.

Scientific laws and principles need to have the same name in Beijing and in New York. In science, variant names are perceived as a nuisance and a hindrance. The global laws of science extend equally to the latest technology, which is vital for the military. A tank or rocket, which works on a test field needs to work equally when it is shipped to a battlefield in some far off country. Engineers at home and abroad have to have the same up to date terms if operations are to run smoothly.

International Standard (developed in Europe by ISO, ITU etc).	United States Standard (based on a simplification of the international standard)
EDIFACT	EDI
X.500 DAP	LDAP stand-alone LDAP daemon (slapd) stand-alone LDAP update replication
SGML	XML

Figure 1. Examples of international standards in Europe, which become a starting point for simplified versions thereof in the United States.

Thus in business, science, technology and the military the quest is to agree on standard names and forget the variants. There is a temptation to focus exclusively on current, new knowledge: to build directories and databases, which are like the names in telephony:

single names with no variants. Given the telephone number mapping (ENUM) project and the new initiative to create a Universal Whois (Uwho) function, there is a temptation to treat Internet names in the same way.

3. Dangers

Science is clearly an important success story and is seen as key to future economic success and political power. There is enormous funding in this field. It is an obvious focus of attention. The scientific world has the full support of the business world. All this is clearly feasible[19] and there is nothing wrong with all this. The danger lies rather in assuming that this is all that there is: the methods and standards required for the development of science as e-science, e-business and e-warfare can serve as a model for everything else.

Our specific concern is that these fields in many ways non-cumulative: i.e. once a new principle is developed, previous principles are theoretically of no further interest. The latest is of paramount importance. Everything else is outmoded and effectively useless. This applies equally in the case of the latest scientific law, business practice and military gadget.

Where it does not apply, however, is the realm of culture, which is fundamentally cumulative in nature. Shakespeare or Leonardo are not less relevant because they are a year older. Paradoxically, because more things continue to be said about them, cultural heroes such as these become more important with each passing year. This applies to Renaissance figures and equally to individuals form older periods and other cultures such as Homer, the Buddha, Confucius and Lao Tse.

As a result of these cultural traditions, parallel with contemporary directories and databases for new knowledge, there are myriad catalogues, directories and databases in the realm of enduring knowledge in memory institutions (libraries, museums, and archives). These directories are cumulative, historical, multi-lingual, and multi-cultural.
Hence, the larger challenge which concerns us is how to link these current names of new knowledge with the names of enduring knowledge of our memory institutions; how to include variant names such that we gain access to the richness of cultural and historical knowledge that is available.

This concern has a very practical dimension. Europe has traditionally created models, which are considered too complex by Americans for the very simple reason that the realities, which we need to model are more complex than those of the United States which supports a theory of melting pot whereby the many cultures become one. By contrast, Europe needs to take into account many languages, even more dialects, many cultures and thousands of years of recorded history.

Of these factors, languages are the most obvious ingredient to complexity. In the United States anyone who does not speak English is considered as a lesser among equals at best. Languages are considered an interesting curiosity, which have nothing to do with the

everyday real world of English. Even very serious Americans, including former heads of ICANN, still frequently assume that English is the only real language of the Internet. In Europe, the notion that languages are unimportant is simply not a serious view. Multilingualism is clearly a challenge and has been addressed in many programmes and need not be considered further at this juncture.

More specifically our concern lies with four other topics relating to names and naming where there are serious differences between the United States and Europe: 1) in the fascination with fuzzy versus exact methods: 2) in the treatment of variant names; 3) in a more systematic treatment of domain names whereby Internet names can be treated more systematically to include a) a geographical component that can be linked with Global Positioning Systems (GPS) and Geographical Information Systems (GIS); b) authority lists based on subject headings which can be linked to existing classification systems. Needed also is 4) a new integration of local, regional, national, international and global directories and databases such that knowledge at all levels can be more systematically accessed.

4. Fuzzy versus Exact

The earliest search engines searched for a term which was an exact match. Assuming one had the right spelling, this process was accurate but slow. To cope with this drawback statistical probability methods were introduced. This speeded up the process so enormously that the positive impact of so many new hits seemed to outweigh the negative impact of a number of false hits. As a result most search engines today use approximate names. Even Google has an automatic spelling adjuster. It guesses what we might have meant. In vague searches this continues to be a useful strategy.

Sometimes exactness is important. Unless I am the owner of a harem, I want to know precisely how many wives and children I have. Being told I have approximately 3 children is not useful for my accountant. Similarly, in cases where I know the precise name of the person involved, I do not want someone who is almost the same as John Smith. I do not even want any John Smith. I want a specific John Smith who lives in my home town. The quest for a universal Whois, or Uwho as it is so charmingly called, once it looks beyond a few local directories and databases to consider the enormous richness of names with all their variants will have to be very precise if it is to find just the name that I want and not just any John Doe.

5. Variant Names

Moreover, it will need to cope with variant names. To take a very personal example, my name is different in Canada than when I am in the Netherlands due to translation problems. Historically, this was a much greater problem. Complex or unfamiliar names

How?	Where?	Why?	What? Who? When?
System	Geographical Info	Category	Subject Headings, Classifications
	CcTLD	TLD	2LD, 3LD etc.

Figure 2. Basic characteristics of a new Domain Name scheme.

often had different names, sometimes with a half dozen or ten variants, bearing no simple relation to one another. For instance, the Italian Renaisance architect who rebuilt the Castello Sforzesco in Milan is remembered as Filarete but his official name was Antonio Averlino. The author of the first French, printed treatise on perspective, Jean Pélerin, was called Pelegrinus, Viator, Viateur or even Jean Gast. No amount of probability could predict such a range of variant names and still hope to find a precise individual. If we eliminate variant names then shall find only the standard names which we know already. Variant names are alternative entry points into knowledge from the past.[20] Hence, unless we have complex authority files with all these variants, which have occurred over the ages we cannot have a proper Uwho. For general searches we can continue to use all the tricks of probability. But when we are sure we know what we want to find then we need exact tools to find precisely what we need.

6. Domain Names and Classification Systems

At the moment Domain Names are confusing at best. A typical name such as www.louvre.edu.fr has four ingredients: 1) the type of server (i.e. www as opposed to an alternative such as ftp); 2) an institution (known in American as a Second Level Domain name or 2LD; 3) a type of organization (known in American as a Top Level Domain name or TLD) and 4) a country (in this case France as fr). In the present system these domain names at various levels were conceived before there were country codes and are positioned from right to left in the manner of Arabic.

One shortcoming with this approach is that if one wishes to add further terms (2LDs, 3LDS) then one has to alter the whole address. There is no basic part, which remains constant to which one can add extras. In the long term this is very inefficient. A second shortcoming is that geographical elements (where?) and content elements (what? or who?) are mixed up such that there is no way to search simply for Who? What? or Where? We need a system where these questions are treated systematically (figure 2)

Such a logical approach (figures 3) would reverse the order of the domain names and introduce a more systematic order. A first section answers the question How information is transported? Hence, one begins as usual with 1) the system (www, ftp, telnet etc.). A second section answers the question Where? and has a geographical component. This includes the 2) the continent (e.g. .eu); 3) the country (.fr); 4) the institution (Louvre). This section can be linked with Geographical Information Systems (GIS) and Global Positioning Systems (GPS) which can provide tourists with maps showing the location of the Louvre on their mobile telephones and devices. A third section answers the question Why the site is there? It does so through 5) a Top Level Domain (TLD), which identifies

1. System		www, ftp, etc.
2. Continent		eu
3. Country		fr
4. Institution		Louvre
5. Kind of Organization	(TLD or 1LD)	edu
6. Description linked with Classification Systems	(2LD)	paintings
7. Further Description linked with Calssifications	(3LD)	Renaissance

Figure 3. A concrete example of basic ingredients in a new domain name scheme. the kind of organization, whether it is government, business, education, military, etc. In the new system these first three sections of the address are fixed.

By contrast a fourth section, which answers the questions What, Who and When? will constantly have add-ons. This section will use authorized subject headings and will follow from left to right such that one will have 6) Second Level Domains (2LDS) and 7) Third Level Domains (3LDS) successively. If one organizes domain names in this way then there is no problem in adding elements to the description, which serve further to limit and define the topic at hand. Such constraints as they are called in computer jargon help us to navigate from the generality of the universal to the specifics of the particular. To continue with the example cited in figure 3 it would be easy to add some further descriptors such that one had: 5) Paintings/Renaissance/Italian/ Florentine/Leonardo/ MonaLisa.

How will this approach affect companies concerned with their trademarks and brand names? Multinationals, which do not wish to be identified with a single address, will remain the way they are today: e.g. www.ibm com will continue to exist as before.

Because private individuals linked with a specific address will begin with Internet addresses such as www.eu.fr.etc. there will be less confusion between multinational companies and individuals who might wish to comment on them. On the other hand, organizations such as the European Brands Association are discovering the need to use the recognized headings of Yellow Pages, and professional classification systems to identify their materials. This will become all the more essential as they move towards making their product catalogues available on line. Such descriptions would simply continue from left to right and thus represent simple add-ons rather than requiring that the entire address be changed with each addition. The new system thus allows cumulative change rather than the constant disruption implicit in domain names as they exist today.

While large companies clearly want a first entry point via a brand name such as www.ibm.com when it comes to meeting customers, companies need as much as anyone to be readily findable. It may therefore prove useful to have individual branches have websites complying with the general model such as www.eu.it.ibmnaples.com in order that these can readily be found geographically as well as conceptually. Further details would then follow in the left to right manner using names and subject headings from authority files.

There is a very simple reason why the telcos have such strict rules in making their unique lists and why librarians insist on authority files when classing objects in their collections. As humans we have a natural tendency to associate a given item or concept with a number of different terms. Unless these are carefully defined then we inevitably class the same things under different headings at different times. Unless we have authority files to tell us which is the accepted term and which are the variants only confusion will result.

With the help of technology and authority files such as subject headings, which are then arranged as thesauri and classification systems what was traditionally a source of confusion can become a source of new insight. As in the case of variant names, as long as one has a clear record of equivalents, one can map systematically across various classification systems.[21]

Classification systems are much more than simple lists of contents. They offer insights into the various methods of cubbyholing information in different cultures. They are records of the practical categories of thought in a given culture. If these can be systematically accessed then we have new methods of gaining insight into ways of organizing the world, new insights into world-views. This quest is linked in turn with another concern: a new integration of local, regional, national, international and global knowledge.

7. Local, Regional, National, International and Global

The first three of these categories have existed for over two millennia. The significance of international and global has evolved rapidly in the course of the past two centuries. In the nineteenth century the problem appeared fairly straightforward. Local knowledge was housed in local libraries and archives, regional and national knowledge was housed in regional and national institutions respectively. There were a few organizations, inspired mainly by Panizzi's vision for the British Museum, that a single institution could be the repository of all knowledge. This vision was as old, of course, as the Library of Alexandria and it took more than a century to discover that no building, no matter how large, not even a *Très Grande Bibliothèque* could house everything systematically.

Parallel with this quest for a universal collection was a quest for a universal system of organizing knowledge which led to the visions of Otlet and LaFontane for a new Encylopaedia Universalis Mundaneum. This quest gradually inspired the growth of many international organizations such as the International Telecommunications Union (ITU) and the International Standards Organization (ISO) in the technical domain, but also their equivalents in the cultural field: the International Federation of Library Associations (IFLA), the International Committee on Museums (ICOM), the International Federation of Documentation (FID) and the Union Internationale des Associations (UIA).

For the founders of these organizations the goal seemed clear. Local and regional agreement was not worth much. National agreement was useful. International agreement was the goal. In one sense they were absolutely right. Without these international

frameworks there would be no hope at arriving at international authority files which pay due attention to the complexities of multilingualism.

Once they began reaching agreement at a global level the internationalists thought that they saw an easy solution for the future. In their view, all they needed to do was to establish a standardized method. This could then be sent around, used by everyone and the problems would be over. In the library world, for instance, this seeming solution took the form of Machine Readable Cataloguing (MARC). What happened was something unexpected. Almost invariably countries introduced their variations, their own flavours to the so-called MARC standard such that there were soon USMARC, UK MARC, AUSMARC, etc. Contrary to the dictum on the American dollar bill (*E pluribus unum*) from the one came many.

Some thinkers thought the problem lay in the complexity. If only one could reduce this complexity, they reasoned, then one might find a small number of essential categories on which there could be universal agreement. This was the assumption behind the 15 categories of the so-called Dublin (Ohio) Core project. This idea was originally presented in 1995. In theory it was very simple. Despite considerable investments and high level politics there has been great resistance to this top down model. Paradoxically we are content to argue that the former Soviet Union collapsed because of its top down bureaucracies and yet we frequently assume that the same methods can be applied in another context.

It is important to recognize that even if the Dublin Core were successful it would be a Pyrrhic victory. To understand this it is important to return to the distinctions between science and culture outlined at the outset of this essay. Science depends on universal laws and standards, which are the same everywhere and thus naturally tends towards the international and the global. Science such as physics is the same in China as it is in the United States.

Culture, by contrast is not the same everywhere. The 19th century tried to make culture a national thing and to a certain extent this is still so today. There are cultural traditions, which distinguish an Italian from a Frenchman or a German. Yet in its more complex forms, culture is almost always regional and local. What makes Italy unique is not that it is a place where persons speak Italian and eat pasta but rather because there are many Italian dialects and many dozens of different pasta. As Stillman Drake, the leading authority on Galileo used to say, Italy is only country where every little town acts as if it were the centre of the world and where in a certain sense that is also perfectly true.

In very simple terms, the quest for a single universal, global system may solve the /problems and challenges of science, which strives for a single law concerning every phenomemon. It may solve the needs of the military where a single will needs to dominate in the interests of order. It might seem to fit in with the needs of business which imagines that a single solution would be much more efficient than a series of alternative ways of doing things. But it cannot solve the challenges of culture. How much poorer

would the world be if there were only one kind of beer, only one kind of (Kraft) pasta, only one kind of (McDonald's) food, only one kind of art, music etc.?

Indeed, in retrospect we can see that the richest expression of human culture came precisely from contexts where persons shared familiar traditions, usually a holy book such as the Bible, the Koran, the Mahabharata, the Buddhist, Confucian or the Taoist texts) and used these as a point of departure for diversity rather than homogeneity. To take a simple example for the Italian Renaissance. It was the great familiarity with the *New Testament* and specifically Luke 1:25-27,[22] that inspired artists to paint the *Annunciation* as a particularly important scene from the *Life of Christ*. Some such as Leonardo da Vinci painted it as taking place in a garden. Other such as Crivelli painted the event in a courtyard. Northern painters such as Van der Weyden preferred to paint the same scene indoors. It is thanks to the enormous diversity of their interpretation of the same passage that we rightly view the Renaissance as one of the richest cultural periods in Western history.

The fifteen fields of the Dublin Core cannot do justice to this complexity and diversity. A closer study of the past century could lead us to a much more dramatic conclusion. In the realm of culture no external system imposed from the outside can do justice to the local and regional complexities wherein the richness of culture lies.

According to thinkers such as Benjamin Barber we are therefore faced with a curious dichotomy: a Herculean choice between disembodied and bloodless globalism that entails a McWorld effect and a bloody regionalism in the form of jihads and the like. It is not necessarily a question of either or. There is a better set of solutions.

While we cannot simply impose a template externally in a top-down manner, we can reasonably use a global template for an umbrella function of bridging to various national, regional and local databases. Instead of asking everyone to rewrite their databases and directories, we need only ask them to provide their rules so that we can bridge their legacy systems with ours. Hence, instead of imposing a single standard, we introduce a centralized authority register, which becomes a standardizing instrument insomuch that it allows connections and bridges to diversity.

The nineteenth century mentality, which is also the mentality dominant in science, the military and business, diversity in the form of many languages and cultures were inevitably seen as inconveniences at best. In the new approach, the quality of the system is judged by the extent to which it offers access to the complexities of regional and local language and culture. This implies that there also will be differences in what one records at various levels. At the global level one is interested only in a standard terminology and dictionary to define that terminology (with a list of variants for bridging purposes). At the local level, by contrast, one has a local terminology, local dictionary, a corpus, oral recorded material, and perhaps even oral material, which has not yet been systematically recorded.

Once we begin to think in these terms there is as usual good news and bad news. The good news is that we escape from what Barber perceived as a dichotomy between the regional and the global. We can establish a global system and use that system to ensure access to local and regional diversity, rather than pretending that global standards are only possible by destroying that diversity. The imperialism of the 19th century was a mistake, but if we can use that experience in order not to repeat that mistake, then there is some hope.

What some would call the bad news is that the problem of a universal Whois, which persons in Verisign think can be solved with a part of $200 million or with at least $200 million, is a problem which may cost much more if it is to be solved properly on a global scale in such a way that the full richness of culture is properly reflected at the regional and local levels. Scholars such as Manfred Thaller in his Prometheus project,[23] are showing that these ambitions have practical solutions.

In the 1960s, the late Marshall McLuhan warned us that we were entering the age of the global village. If we follow Barber's vision of McWorld then this means we are entering an age where every village on the globe is the same as every other village. By the same token, there is no reason why we have to make use the melting pot mentality which has inspired the motto: from the many one. We could have a globe where every village can communicate with others around the globe and use others' diversity as a source for courage to develop one's own uniqueness. If the path to this goal is less simple, the fruits of this goal are less simplistic. They promise access to the full range of human expression which is different in every place and changes over time rather than promising us only what happens to be the flavour of the month. European Directors of directories and databases have a choice. They can follow the simple formula and let an external body tell us what we need. Or they can tackle the deeper problems and find a solution that will answer the complexities of Europe and offer a new model for appreciating the richness of cultures everywhere around the globe.

8. Conclusions

We noted a) that there is a well established tradition for Europe to develop complex solutions to deal with its rich cultural and historical traditions and b) an equally well established tradition whereby complex solutions developed in Europe are simplified and often made even more successful in the United States. We have intimated that there is unfortunately also a third tendency whereby c) these simplified, popularized American imitations of what was developed in Europe is exported back to Europe as if it were the only viable solution. Herein lie grave dangers insomuch that we cannot expect "quick and dirty" fixes to become genuine solutions for immensely complex problems.

At the level of international science we have recognized that the problems are so immense that they cannot be solved by the scientists of a single lab or a single country. It is high time for us to recognize that, if anything, the same holds even more true in the case of culture. In the realms of science, the military and business the answers to global names in directories and databases are relatively straightforward. In the cultural world, in

memory institutions (libraries, museums, and archives) we need a much more complex approach that fosters and maintains cultural diversity which is as crucial as biodiversity for our future survival. We need to include variants in multiple languages and dialects, at multiple levels from the local and regional to the national and global.

If all this requires much more effort and investment than a simple directory of current names, it also brings profound advantages. On a purely economic front, it means that cultural and historical knowledge can be made available for the tourist industry. At an educational level it means that children and students have a means of comparing their own traditions with those of others in order both to understand what they share and what is unique therein, thus leading to greater tolerance. At a societal level it means that access to the richness of cultural diversity can be continued and fostered. Tolerance and understanding in this domain is not a luxury. It is essential to the fabric of civil society and civilized society. Unlikely as it may seem, multi-lingual, variant name directories may be one of the keys to the future.

The pioneers of computing had a saying: Garbage in, garbage out they warned. They recognized at an early stage that imput and output are related. It follows that if we want to get the richness of culture out of our systems we must begin by assuring that that richness is put in in the first place. Unless we enter all the complexities of names, we will not find them back.

Shakespeare once asked: What's in a name? America has demonstrated that the answer is much more than that for which Romeo bargained. Names can be precious also in the monetary sense. But we must take care not to fall into the realm of Oscar Wilde's cynic who knew the price of everything and the value of nothing. Superficially names can be bought. At a more profound level there is something priceless about names, just as there is something nameless about culture which cannot be bought or sold, and constitutes the essence of what is human, which some link with freedom and others with the human condition.

Maastricht, 5 July 2002.

Acknowledgements

I am most grateful to my colleague Drs. John Beckers for kindly reading the text and suggesting corrections. I also thank warmly my colleague Chris Zielinski for reading the paper and offering suggestions.

Appendix 1. Key Developments in naming and directories over the past fifty years.

Electronic Data Interchange (EDI)

Electronic data exchange is actually older than the Internet. Roger Clarke, in his *Electronic Data Interchange (EDI): An Introduction* (1998) traces the idea back to the:

> 1948 Berlin Airlift, where the task of co-ordinating airfreighted consignments of food and consumables (which arrived with differing manifests, languages and numbers of copies) was addressed by devising a standard manifest. Electronic transmission commenced during the 1960s, initially in the rail and road transport industries. The standardisation of documents was a necessary concomitant to that change. In 1968 the United States Transportation Data Coordinating Committee (TDCC) was formed, to coordinate the development of translation rules among four existing sets of industry-specific standards. A further significant move towards standardisation came with the X12 standards of the American National Standards Institute (ANSI), which gradually extended and replaced those created by the TDCC.
>
> At about the same time, the U.K. Department of Customs and Excise, with the assistance of SITPRO (the British Simplification of Trade Procedures Board), was developing its own standards for documents used in international trade, called Tradacoms. These were later extended by the United Nations Economic Commission for Europe (UNECE) into what became known as the GTDI (General-purpose Trade Data Interchange standards), and were gradually accepted by some 2,000 British exporting organisations.
>
> Problems created by the trans-Atlantic use of two different (and largely incompatible) sets of standardised documents have been addressed by the formation of a United Nations Joint European and North American working party (UN-JEDI), which began the development of the Electronic Data Interchange for Administration, Commerce and Transport (EDIFACT) document translation standards. A full range of business documents is in the process of being developed.[24]

According to Jason Levitt, in his *Brief History of Web Services*, the quest for Electronic Data Interchange (EDI) was launched formally in 1975 as a first attempt to create a standard way for businesses to communicate over a network:

> In the 25 years since EDI came on the scene, there have been numerous attempts at a universal conduit for connecting business logic over a network: Common Object Request Broker Architecture, Distributed Component Object Model, Unix Remote Procedure Call, and Java Remote Method Invocation. Each of those technologies failed to gain significant market share or enough momentum to succeed. All of them exist today--each still has its uses--but each failed to gain a broad reach.

> EDI was difficult to implement because of its complexity and cost. Corba, deployed mostly by Unix systems vendors, and DCOM, a Microsoft technology, competed with each other for many years. Both were relatively difficult for programmers, and neither gained broad industry support.[25]

While EDI was clearly not a perfect solution, a serious community of users continued to develop.[26] At least four developments helped this process: the rise of EDIFACT, X.500, LDAP and more recently the advent of XML/EDI.[27]

Electronic Data Interchange For Administration, Commerce and Transport (EDIFACT)[28]

Meanwhile the United Nations, Economic Commission for Europe (UN/ECE) established a Working Party on the Simplification and Standardization of External Trade Documents in 1960. In 1972 this became the:

> Working Party on Facilitation of International Trade Procedures, with terms of reference extended to cover, inter alia, the development of a standard international trade data terminology and a uniform system for use in automatic processing and transmission of trade data….International co-ordination in the development of uniform communication protocols for various types of communication networks (both private and public) is ensured through such specialized international bodies as the International Telecommunication Union (ITU), International Telegraph and Telephone Consultative Committee (CCITT), ISO TC 97/SC 6 "Data Communication", etc.[29]

	Service String Advice	UNA Conditional
_____	Interchange Header	UNB Mandatory
| ___	Functional Group Header	UNG Conditional
| | _	Message Header	UNH Mandatory
| | |	User Data Segments	As required
| | |_	Message Trailer	UNT Mandatory
| |___	Functional Group Trailer	UNE Conditional
|_____	Interchange Trailer	UNZ Mandatory

Figure 3. Basic EDIFACT Components.[30]

While EDIFACT was an excellent contribution it took many decades to develop and was too complex for simple everyday use (Figure 3). By 1999, final rules for EDIFACT were being discussed.[31] Nonetheless, it made an important contribution in providing a high level technical solution to many of the underlying problems.

X.500

The rise of computers in the 1980s introduced a new challenge: how to integrate the naming problems of telephone companies with those of computer companies. In Europe this led to new co-operation between the leading standards bodies:

> On the one hand there was the CCITT (now the ITU-T), whose major concern was to provide a white pages service that would return either the telephone numbers or X.400[32] O/R addresses of people, and on the other there was the International Standards Organisation (ISO) and the European Computer Manufacturers Association (ECMA), who were concerned mainly with providing the name server service for Open Systems Interconnection (OSI) applications. Inevitably, the two tracks would merge, and they did in 1986, with the formation of the Joint ISO/CCITT working group on Directories (w/w 1.1). The interests of ECMA were subsumed by ISO, since the former sent delegates to meetings of the latter.[33]

The goal posed by this group was effectively a new synthesis of many existing efforts in the directories field:

> Ideally we would all like access to more than the conventional telephone directory. As it stands, a paper directory, or directory enquiries service, is very limited in the service that it offers. For a start, the paper one is always at least six months out of date. And the directory enquiry service is very expensive to use in some countries. Not only that, but the national and international ones are accessed differently. Computer access to a conventional database gives you many more features such as: rapid scanning of thousands of entries, retrieval of entries with similarly spelt names, and retrieving the name of a person with a given telephone number or address. If only we could computerise the entire set of global telephone directories, and interconnect them, and give people access to them all via a standard interface, then we would have a real directory service. X.500 of course is designed to provide this service, and many more besides.[34]
>
> A major part of X.500 is that it defines a global directory structure. It is essentially a directory web in much the same way that http & html are used to define & implement the global hypertext web. Anyone with an X.500 or LDAP client may peruse the global directory just as they can use a web browser to peruse the global Web. Additionally, with the help of web<->X.500 gateways, you can use your favorite web browser to peruse both![35]

The X.500 efforts began in 1984 and took a decade to complete. When they were finished in 1994 the European Commission had large-scale plans for its deployment. At about the same time Al Gore introduced his idea of an information highway, which inspired the Bangemann Report n Europe and a decision to create a G7 Exhibition and Conference in 1995. It is rumoured that the monies planned for X.500 were diverted to pay for this event.

X.500 is a complete solution and also defines the "Directory Access Protocol (DAP) for clients to use when contacting directory servers." This heavyweight protocol runs over a full OSI stack and "requires a significant amount of computing resources to run."[36]

Lightweight Directory Access Protocol (LDAP)

The Lightweight Directory Access Protocol (LDAP) was originally developed as a front end to X.500, the OSI directory service:

> LDAP runs directly over TCP and provides most of the functionality of DAP at a much lower cost. …This use of LDAP makes it easy to access the X.500 directory, but still requires a full X.500 service to make data available to the many LDAP clients being developed. As with full X.500 DAP clients, a full X.500 server is no small piece of software to run. The stand-alone LDAP daemon, or slapd, is meant to remove much of the burden from the server side just as LDAP itself removed much of the burden from clients. [37]

Linked with the stand alone LDAP daemon is a stand-alone LDAP update replication (UNIX) daemon (slurpd) "that helps slapd provide replicated service. It is responsible for distributing changes made to the master slapd database out to the various slapd replicas."[38]

The LDAP development was two edged. The good news was that there now seemed to be an easy solution to directory problems which sidestepped the high computing and administrative costs of the X.500 solution. The bad news was that a number of non-standardized implementations of LDAP soon arose. "The most widely used implementtation of LDAP was written at the University of Michigan, and development is currently being supported by the National Science Foundation under Grant No. NCR-9416667."[39] To deal with this fragmentation an Open LDAP Consortium was founded. [40]

One of the major ironies of these developments is that serious solutions such as X.500 developed in Europe, are taken up in a too simplified form in the United States, and then taken up anew in Europe where they do not work and cause havoc to the community at large.

SGML- XML

During the 1960s, at the time that the possibility of the Internet was being discussed, those in the text world had become aware of a fundamental problem: the basic information in manuals frequently needed to be displayed in alternative forms for users at various professional levels and with different needs. This led to the development of Standard Generalized Markup Language (SGML), which separated content from form in order to reduce the costs of printing different versions for various audiences.[41]

Not unlike the EDIFACT and X.500 solutions, SGML is a marvelous universal solution but much too complex for everyday use. The World Wide Web (W3) Consortium thus

developed a subset of SGML better suited to the needs of everyday use, called eXtensibe Markup Language (XML). This provides a common denominator for regular content. Specialist users are writing extensions such as Chemical Markup Language (CML) and Mathematical Markup Language (MML).[42] As Mike Ricciuti has noted:

> XML grew out of the publishing world as a way to describe the contents of documents exchanged over the Web. One of the language's key qualities is that it is readable by people, not just machines, making systems easy to debug.…
> The language provides terms used to define a Web document's tags--that is, the elements of the document that describe its various pieces--and the relationships between them. Developers at either end of a data exchange then agree to use a common set of tags. In this way, XML is uniquely flexible and versatile: It can be used to describe tennis balls or tires, employment contracts or engine parts.
> That flexibility, however, can also present a problem. Because each XML transfer includes a large amount of information describing the data contained in it, file sizes can quickly mushroom.[43]

Although XML began as a solution for the publishing world, it has become intimately connected with the future of the business world. Here there have been four key developments in the past years: SOAP, WSDL, UDDI, and XML/EDI.

Simple Object Access Protocol (SOAP)

TheWorld Wide Web (W3) Consortium which has developed XML, has developed Simple Object Access Protocol, (SOAP), "a lightweight protocol for exchange of information in a decentralized, distributed environment….SOAP can potentially be used in combination with a variety of other protocols." In its present form SOAP has bindings for the Hyper Text Transfer Protocol (HTTP) and the HTTP Extension Framework.[44]

Web Services Description Language (WSDL)

Another initiative of the W3 is Web Services Description Language (WDSL) which is an: "XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information. The operations and messages are described abstractly, and then bound to a concrete network protocol and message format to define an endpoint."[45] Like SOAP, WDSL can be extended to various message formats and network protocols, but in its present state the only bindings are in conjunction with SOAP 1.1, HTTP GET/POST, and MIME.[46]

Universal Description, Discovery, Integration of business for the web (UDDI)

As so often with successful initiatives, potentially successful solutions frequently have a number of different implementations. In the case of SOAP and WSDL major vendors have produced their own flavours, which reintroduce problems of compatibility and interoprerability. Hence, once again there has been a recognition that some kind of centralized registry is needed if the various local, distributed databases are to be

integrated. Here one of the candidates is Universal, Description, Discovery, Integration of business for the web (UDDI):

> Web services is a hot, relatively new technology that is poised to connect enterprises to their business partners and customers, as well as build bridges between disparate IT systems--all without traditional software and extensive development effort. Web services will fail to become mainstream for complex business usage if a Web services registry is not used. The current leading technology for this role is the UDDI registry[47].

XML/EDI

Meanwhile, another approach has been to align the developments of extensible Markup Language (XML) with those of Electronic Data Interchange (EDI):

> XML/EDI provides a standard framework to exchange different types of data -- for example, an invoice, healthcare claim, project status -- so that the information be it in a transaction, exchanged via an Application Program Interface (API), web automation, database portal, catalog, a workflow document or message can be searched, decoded, manipulated, and displayed consistently and correctly by first implementing EDI dictionaries and extending our vocabulary via on-line repositories to include our business language, rules and objects. Thus by combining XML and EDI we create a new powerful paradigm different from XML or EDI.[48]

Notes

[1] European Association of Directory and Database Publishers (*EADP)*, Rey Juan Carlos Hotel, 19 September, 2002.

[2] Genesis 2:20.

See: http://www.biblegateway.com/cgi-bin/bible?search=Names&SearchType=AND&version=KJV&restrict=&StartRestrict=&EndRestrict=&rpp=25&language=english&searchpage=0&x=11&y=7

[3] Roger Clarke, *Electronic Data Interchange (EDI): An Introduction*, Xamax Consultancy Pty Ltd, Canberra.

See: http://www.anu.edu.au/people/Roger.Clarke/EC/EDIIntro.html**.**

> Unix RPC, which refers to several flavors of technologies that are available on Unix systems, was never widely deployed outside the Unix vendors. Sun's Java RMI technology is a recent addition; I don't believe it will get complete industry support, because of Microsoft's break with Java.

Jason Levitt, “From EDI To XML And UDDI: A Brief History Of Web Services “Oct. 1, 2001.

See: http://www.informationweek.com/story/IWK20010928S0006.

Cf. B.L.M. Goldstein, S. J. Kemmerer, and C. H. Parks, "A Brief History of Early Product Data Exchange Standards," NISTIR 6221, Natl. Inst. Stand. Technol. (U.S.), 17 pages (Sept. 1998). See: http://www.eeel.nist.gov/811/design.html

[5] Reference should also be made to the Basic Reference Model for Open Systems Interconnection (OSI), ISO 7498.

See: http://www.unece.org/trade/untdid/welcome.htm

[6] See: http://www.edifactory.de/edifact/edimain1.html

[7] UN/EDIFACT.

See: http://www.unece.org/trade/untdid/sessdocs/ewg_0598.htm

[8] See: http://www.umich.edu/~dirsvcs/ldap/doc/guides/slapd/1.html#RTFToC1

[9] It frees *slapd* from having to worry that some replicas might be down or unreachable when a change comes through; *slurpd* handles retrying failed requests automatically. *Slapd* and *slurpd* communicate through a simple text file that is used to log changes.

See: http://www.umich.edu/~dirsvcs/ldap/doc/guides/slapd/1.html#RTFToC1

[10] See: http://www.umich.edu/~dirsvcs/ldap/

[11] See: http://www.openldap.org/project/

[12] With respect to elementary word processing, this is apparent in packages such as Microsoft Word, whereby the process of entering content (by typing) is separated from questions of format such as adding bold, italic or underlined scripts, centred, left or right aligned scripts etc.

In the eyes of the world, the introduction of Hyper Text Markup Language (HTML), in 1989, marked a revolutionary step forward, because it made the World Wide Web an almost instant reality. From a technical standpoint, it marked a step backwards insomuch that this interim solution re-conflated the functions of content and form. As a result any change in the form of an HTML document, required re-writing the entire contents of the document.

[13] These include:

Chemical Markup Language	(CML)
Handheld Device Markup Language	(HDML)
Hardware Description Language	(HDL)
Mathematical Markup Language	(MML)
Web Interface Description Language	(WIDL)
Precision Graphics Markup Language	(PGML)
Extensible Hyper Language	(EHL).

[14] See: http://www.w3.org/TR/SOAP/

[15]Related concrete endpoints are combined into abstract endpoints (services). WSDL is extensible to allow description of endpoints and their messages regardless of what message formats or network protocols are used to communicate.

See: http://www.w3.org/TR/wsdl; cf. http://xml.coverpages.org/wsdl.html

[16] See: http://www.uddi.org/

[17] See: http://www.w3.org/Submission/1999/02/

[18] See: http://www.geocities.com/WallStreet/Floor/5815/

[19] There are cynics who object that much of this is hype: that the visions of convergence of broadcast and telephony paradigms with those of internet are much exaggerated; that

there is a great discrepancy between the talk of gigabit and terabit networks to link exabyte databases and the everyday shortcomings of networks which struggle with single gigabit throughput. At the frontiers of progress there is always hype, there are always versions which overestimate the possibilities of new developments. Even so there is little reason to doubt the main lines of these developments.

[20] "Past Imprecision for Future Standards: Computers and New Roads to Knowledge", *Computers and the History of Art*, London, vol. 4.1, (1993), pp. 17-32. Cf. "Four Ways that Digital Communications are Transforming Scholarship: Sources, Names, Claims and Scope," *INET 2003*, June 2003, (in press).

[21] See the author's: **"**Four Ways that Digital Communications are Transforming Scholarship," prepared for submission to INET 2003 (in press).

Sources, Names, Claims and Scope.

[22] See: http://www.biblegateway.com/cgi-bin/bible?passage=LUKE+1:25-27&language=english&version=KJV&showfn=on&showxref=on

[23] See: http://www.prometheus-bildarchiv.de/

[24] Roger Clarke, *Electronic Data Interchange (EDI): An Introduction*, Xamax Consultancy Pty Ltd, Canberra.

See: http://www.anu.edu.au/people/Roger.Clarke/EC/EDIIntro.html**.**

Unix RPC, which refers to several flavors of technologies that are available on Unix systems, was never widely deployed outside the Unix vendors. Sun's Java RMI technology is a recent addition; I don't believe it will get complete industry support, because of Microsoft's break with Java.

Jason Levitt, "From EDI To XML And UDDI: A Brief History Of Web Services "Oct. 1, 2001.

See: http://www.informationweek.com/story/IWK20010928S0006.

Cf. B.L.M. Goldstein, S. J. Kemmerer, and C. H. Parks, "A Brief History of Early Product Data Exchange Standards," NISTIR 6221, Natl. Inst. Stand. Technol. (U.S.), 17 pages (Sept. 1998).

See: http://www.eeel.nist.gov/811/design.html

[26] EDI Source.

See: http://www.1edisource.com/

[27] See: http://www.xmledi-group.org/xmledigroup/ediindex.htm

[28] See: http://www.edifactory.de/edifact/edimain1.html

[29] Reference should also be made to the Basic Reference Model for Open Systems Interconnection (OSI), ISO 7498.

See: http://www.unece.org/trade/untdid/welcome.htm

[30] See: http://www.edifactory.de/edifact/edimain1.html

[31] UN/EDIFACT.

See: http://www.unece.org/trade/untdid/sessdocs/ewg_0598.htm

[32] x.400 is an ISO and ITU standard for addressing and transporting e-mail messages. It conforms to layer 7 of the OSI model and supports several types of transport mechanisms, including Ethernet, X.25, TCP/IP, and dial-up lines.

See: http://www.webopedia.com/TERM/X/X_400.html.

[33] See: http://www.isi.salford.ac.uk/staff/dwc/Version.Web/Chapter.1/Chapter1.htm#1.1%20EVERYONE%20NEEDS%20DIRECTORIES

[34] See: http://www.isi.salford.ac.uk/staff/dwc/Version.Web/Chapter.1/Chapter1.htm#1.1%20EVERYONE%20NEEDS%20DIRECTORIES

[35] Jeff Hodges, An LDAP Roadmap & FAQ. A tutorial aid to navigating various LDAP and X.500 Directory Services resources on the Internet
See: http://www.kingsmountain.com/ldapRoadmap.shtml

[36] See: http://www.umich.edu/~dirsvcs/ldap/doc/guides/slapd/1.html#RTFToC1

[37] See: http://www.umich.edu/~dirsvcs/ldap/doc/guides/slapd/1.html#RTFToC1

[38] It frees *slapd* from having to worry that some replicas might be down or unreachable when a change comes through; *slurpd* handles retrying failed requests automatically. *Slapd* and *slurpd* communicate through a simple text file that is used to log changes.
http://www.umich.edu/~dirsvcs/ldap/doc/guides/slapd/1.html#RTFToC1

[39] See: http://www.umich.edu/~dirsvcs/ldap/

[40] See: http://www.openldap.org/project/

[41] With respect to elementary word processing, this is apparent in packages such as Microsoft Word, whereby the process of entering content (by typing) is separated from questions of format such as adding bold, italic or underlined scripts, centred, left or right aligned scripts etc.
In the eyes of the world, the introduction of Hyper Text Markup Language (HTML), in 1989, marked a revolutionary step forward, because it made the World Wide Web an almost instant reality. From a technical standpoint, it marked a step backwards insomuch that this interim solution re-conflated the functions of content and form. As a result any change in the form of an HTML document, required re-writing the entire contents of the document.

[42] These include:

Chemical Markup Language	(CML)
Handheld Device Markup Language	(HDML)
Hardware Description Language	(HDL)
Mathematical Markup Language	(MML)
Web Interface Description Language	(WIDL)
Precision Graphics Markup Language	(PGML)
Extensible Hyper Language	(EHL).

[43] XML: Mike Ricciuti, “Star: XML could become new lingua franca,” CNET News.com, November 8, 2001.
See: http://news.com.com/2009-1017-275448.html?legacy=cnet

[44] See: http://www.w3.org/TR/SOAP/

[45]Related concrete endpoints are combined into abstract endpoints (services). WSDL is extensible to allow description of endpoints and their messages regardless of what message formats or network protocols are used to communicate.
See: http://www.w3.org/TR/wsdl; cf. http://xml.coverpages.org/wsdl.html

[46] Yasser Shohoud, Introduction to WSDL, Learn XML Web Services Development
See: http://www.learnxmlws.com/tutors/wsdl/wsdl.aspx.

[47] See: http://www.uddi.org/

[48] See: http://www.geocities.com/WallStreet/Floor/5815/