2006

The Semantic Web-Ready Postal Address Management in Global E-Commerce

Damir Trninic
Main Postal Center Belgrade, damir@trninic.com

19th Bled eConference

eValues

Bled, Slovenia, June 5 - 7, 2006

The Semantic Web-Ready Postal Address Management in Global E-Commerce

Damir Trninic

Post Serbia, Main Postal Center Belgrade, Serbia and Montenegro
damir@trninic.com

Abstract

Postal address management in e-commerce is the process of ensuring that all name and postal address data held by online businesses is up-to-date, accurate, easy-to-use and, probably most important, applicable in international business processes. There are many problems associated with name and address data in e-commerce checkout procedures. Determining whether shipping and billing addresses are valid is now a major business inhibitor. The xNAL based method of address management presented in this paper at the point of interaction with the customer (checkout procedure on website or call-center) virtually eliminates the possibility of entering wrong name and address data - all name and address data are validated against data supplied by the national postal authorities. On the other side, this semantic Web-ready solution provides the information available over the Web in a way understandable not only by humans but by computers too; this will allow a broad range of e-commerce and postal solutions to use the Web not only for displaying information as today do, but for more "intelligent" purposes, supporting sharing and reusing data across different applications and businesses, like track-and-trace or direct mail.

1. Introduction

Postal address management in e-commerce is the process of ensuring that all name and postal address data held by online businesses is up-to-date, accurate, easy-to-use and, probably most important, applicable in international business activities of one-stop global-hosting online shops. There are many problems associated with name and address data in e-commerce checkout procedures: Challenges in the treatment of name and address occur mostly during data entry in the checkout procedure, still not so user-friendly and comfortable to many Internet users; Errors and discrepancies in customer information mostly occur during the consolidation of files from different lines of business, mostly e-commerce and m-commerce; The order in which postal address elements are naturally presented varies from country to country: In some countries the house number is provided before the street name, in other countries the house number is given after the street name. For some countries the house number is essential to determine the postcode, for other countries a simple city input is sufficient; Correct entry of an address in an international environment becomes heavily dependent on the knowledge of the person performing the data entry, or the ability to interpret the appropriate address elements.

If an address database contains errors – for example the same address is entered in two different varieties the retrieval of information becomes extremely complex and unreliable.

Name and address, as a data type, has unique characteristics, which make it very difficult to manage. This data is often volatile: customers come and go, addresses change, names change. This data is often cluttered when entered. Name and address fields on front-end screens are usually free format and ripe for users to enter comments and extra data, without any edits.

Name and address is also subjective – it can be written in a number of different ways and still represents the same entities. There is no application independent standard to represent name and address data and to measure its quality in the same time. This problem is further compounded by the different cultural contexts of name and address data in a global market.

Delivery of mail and parcels to customer addresses has become a vital link in the logistic chain between suppliers and customers. This link is tenuous: an average of 15% individuals and businesses moving each year. Studies show that incorrect addresses can cost up to 8% of revenue [1], generated by double postage costs, extra printing and material cost, handling, and the related cost for organizing and administering erroneous deliveries. Determining whether shipping and billing addresses are valid is now a major business inhibitor. For e-businesses alone, between 5 to 9% of the shipments are returned due to addressing errors. In many research papers shipping issues are described as the main barrier for implementing global e-commerce. Call centers face challenges when registering correct addresses, particularly in an international environment where language differences can lead to misunderstandings and incorrect data input. The problem is the variety in international addressing systems, and the lack of knowledge on the format, structure and data involved in a correct address.

The aim of this paper is not only to introduce a new model of postal address management in e-commerce that ensures a seamless integration of all processes in information and material flows, a low-cost and platform-independent solution provided all over the global network of postal and e-commerce companies, but to promote it is a truly semantic Web-ready solution.

2. Postal Address Data in the Semantic Web Era

Traditionally, postal address data has been locked away in proprietary applications of postal companies. Historically, postal operators, locked in publicly protected monopoly and its public or state ownership and control, were seen the quality of postal address data as secondary to processing of data. Recently, both managers and engineers began to realize that postal address data is important, and it must be verified, protected and enabled for global business processes. The integration of service processes, focusing on the value chain and e-commerce, with its new forms of transactions, has further contributed to the pace of development. Once well-known as process-oriented businesses, postal companies are increasingly seen as comprehensive customer-centric businesses that offer a wide range of services bundles. However, postal address management still was kept internal to company's solutions so that each postal operator could keep data proprietary to their applications for a number of reasons in the era of liberalization and deregulation (or harmonization) [2] as well as accelerated globalization and competition. With the Web, eXtensible Markup Language (XML), and now the emerging semantic Web, the shift of power is moving from applications to data. Tim Berners-Lee, the creator of the World Wide Web and one of the main promoters of this initiative, conceives the semantic Web as “an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation” [3]. This also gives us the key to understanding the semantic Web. The path to machine-processable data is to make the data smarter. The vision of the semantic Web is a collection of data in the World Wide Web whose content is meaningful to computers. In this vision computers can directly and effectively process information stored in the Web without having to depend on human guidance [4]. In our case, it means that both e-commerce and postal applications can successfully process postal address data that provide a single customer view at anytime, stored in the Web services-enabled format with minimum risk of wrong customer’s actions. Actually, the only risk is mistyping of the global e-commerce postal code during the checkout procedure.

3. Name and Address Standards and Initiatives

Daconta [5] defines four stages of the smart data continuum; however, there will be more fine-grained stages, as well as more follow-on stages. Let’s discuss each stage and available postal name and address data solutions today:

Text and databases (pre-XML): The initial stage where most data is proprietary to an application. Thus, the “smarts” are in the application and not in the data. This stage of development can be seen in a number of legacy pre-Web applications in many postal companies.

XML documents for a single domain: The stage where data achieves application independence within a specific domain. Data is now smart enough to move between applications in a single domain. The UPU, which has 189 postal administrations as members, develops standards through its Standards Board, historically composed only of postal representatives, but opened up in the last years to industry observers as well. Two resolutions were passed by the UPU Standards Board in 2001 [6]. One calls for developing lists of address elements – the smallest meaningful parts that make up postal addresses, including individual and company names – as well as patterns, or templates, in which these elements are typically arranged. The second calls for developing standards for sending address data between parties using either traditional Electronic Data Interchange (EDI) or the emerging eXtensible Markup Language.

Taxonomies and documents with mixed vocabularies: In this stage, data can be composed from multiple domains and accurately classified in a hierarchical taxonomy. In fact, the classification can be used for discovery of data. Simple relationships between categories in the taxonomy can be used to relate and thus combine data. Thus, data is now smart enough to be easily discovered and sensibly combined with other data. This work of standardization requires both postal and industry cooperation as well as parallel efforts in different regions. In the United States, IDEAlliance has published an Address Data Interchange Specification (ADIS) in XML and relational database formats [7]. The Electronic Commerce Code Management Association (ECCMA) has agreed to serve as a Web-based repository for the address element lists and to sponsor an advisory board that votes on proposed additions [8]. In Europe, the standardization organization CEN has recently published a list of address elements applicable throughout the European Union [9]. At the UPU, the POST*Code project has published a CD-ROM with address elements and templates from all member countries.

Ontologies and rules: In this stage, new postal data can be inferred from existing data by following logical rules. In essence, data is now smart enough to be described with concrete relationships, and sophisticated formalisms where logical calculations can be made on this “semantic algebra.” This allows the combination and recombination of data at a more atomic level and very fine-grained analysis of data. Thus, in this stage, data no longer exists as a blob but as a part of a sophisticated microcosm. An example of this data sophistication is the automatic translation of a document in one domain to the equivalent (or as close as possible) document in another domain.

What will this mean for mailers? Instead of receiving address files on a line-by-line basis, with little assurance as to what data is to be found in any particular location, the data would be marked or "tagged" with codes, so that the recipient always knows the structure and format [10]. The data then is reassembled, using the templates and rules known as "rendition instructions" to produce a particular address format for a specific job. This makes it easier to combine multiple files into one, and mailings for any number of countries can be handled in a single process. There are also advantages for mail production in this approach. The task of merging and purging addresses becomes simpler when all data elements are identified in advance. When the space on a mail piece is limited, quality can be maintained by sacrificing the less important elements and reducing others so as to preserve deliverability.

4. Business Process Workflow

A number of name and address XML standards efforts are underway throughout the world. To a large extent, these standards have been designed with a particular business requirement in mind, for example, the expedient delivery of a piece of mail. This has generally meant that while the particular standard is appropriate for the purpose for which it was designed, it is frequently not suitable for a variety of other purposes. This is the key differentiator between xNAL and the other name and address initiatives throughout world.

xNAL is the only Name and Address XML Standard in industry to-date that is open, vendor neutral, application independent (i.e. independent of postal services, CRM, name and address parsing, matching and validation, etc.) and essentially, global (designed to handle names and addresses of any country in a detailed level or in an abstract level) [11]. It is compatible with the major name and address standards, including European Standardisation Body and Universal Postal Union's CEN TC 331/WG3, Universal Postal Union's own standard, UN/PROLST, ADIS, and IAEC of Electronic Commerce Code Management Association (ECCMA). xNAL is a structured XML language for representing name and address. xNAL consists of basic constructs that can form complex names and addresses. These constructs can be used for name and address data exchange or storage. xNAL consists of two parts: xNL, eXtensible Name Language, to define the name components, and xAL, eXtensible Address Language, to define the address components. xNL and xAL are part of the OASIS Customer Information Quality (CIQ) family of specifications.

This paper presents the xNAL based method of postal address management in B2C e-commerce, largely based on address data exchange from postal organizations to online businesses worldwide. The xNAL based method presented in this paper consists of a few major initiatives and data exchange flows procedures, related to the both customers' name and postal address data management and transactions. Most important, it requires an initiative of major postal companies worldwide to redesign and develop their address management solutions around the xNAL standard. The xNAL schemas are extensible and this feature enables the method of postal address management in electronic commerce, presented in this paper. This means that any number of elements from other namespaces may be added to an xNAL document where any element is allowed. It enables a simple and easy integration of postal-specific information ("GlobaleCommercePostalCompanyCode" element). As a consequence, each xNAL record can contain unique indicator of country and postal company and identify its source

(Figure 1.). Furthermore, it enables postal companies to add another "postal" code ("GlobaleCommerceUserCode").

The xNAL method of postal address management consists of two major components and several technologies of interenterprise application integration and multichannel service integration.

With the deployment of sophisticated multichannel CRM systems in postal companies, customers interact with posts in new and often innovative ways, whether they are requesting information, making purchases, inquiring about a new service, requesting a pick-up, or tracking a parcel. Posts can collect information about their customers through a variety of channels including, surface mail, personal sales calls, counter transactions, customer service call centers, faxed inquiries or transactions, email inquiries or transactions and, finally, online (Web) order forms.

The difficulty comes in trying to consolidate all "silos" of customer data together to create unified customer profiles [12]. The xNAL standard helps postal industry to overcome all major problems, including integration of legacy systems, collaboration between different operating platforms and database systems, further deployment of XML-enabled postal technologies and services. Furthermore, the eXtensible Customer Relationships Language (xCRL) standard, also a part of the OASIS Customer Information Quality (CIQ) family of specifications, allows the creation of customized customer profiles containing other data elements, such as gender, age, function, education, children etc. With the customized customer profiles, direct mail campaigns have another powerful tool. For example, understanding gender for different cultures can be a difficult task for e-commerce companies in international business environment. So, if an "e" is added to "Pascal" (Pascale), it becomes a female name, which may change the way online shops communicate and promote products and services; just imagine the value of the method for online shops of cosmetics and beauty products, categorized on two major products groups – those for women and those for men.


```xml
<?xml version="1.0"?>
<!DOCTYPE xNAL SYSTEM "xNAL.dtd">
<xNAL>
<Record>
<xNL>
<NameDetails PartyType="Person">
<PersonName>
<Title>Mr</Title>
<FirstName NameType="GivenName">NAME</FirstName>
<MiddleName Type="Initial">I</MiddleName>
<LastName NameType="SurName">LAST NAME</LastName>
</PersonName>
<Function>FUNCTION</Function>
<DependencyName PartyType="Organisation" DependencyType="C/O">
<OrganisationNameDetails>
<NameLine>COMPANY</NameLine>
<OrganisationType>BUSINESS</OrganisationType>
<OrganisationKnownAs>KNOWN AS</OrganisationKnownAs>
</OrganisationNameDetails>
</DependencyName>
</NameDetails>
</xNL>
<xAL>
<AddressDetails AddressType="Postal" ValidFromDate="1 Jan 2006" ValidToDate="31 Dec 2006"
Usage="Communication">
<Country>
<CountryName>COUNTRY</CountryName>
<AdministrativeArea Type="State">
<AdministrativeAreaName>STATE</AdministrativeAreaName>
<Locality Type="City">
<LocalityName>CITY</LocalityName>
<Thoroughfare Type="Street">
<ThoroughfareNumber>2</ThoroughfareNumber>
<ThoroughfareName>STREET</ThoroughfareName>
```

```
<Premise Type="Building">
<BuildingName>BUILDING</BuildingName>
</Premise>
</Thoroughfare>
<PostalCode>
<PostalCodeNumber>11111</PostalCodeNumber>
</PostalCode>
</Locality>
</AdministrativeArea>
</Country>
</AddressDetails>
<!-- Extension: Global e-Commerce Postal Company and User Code -->
<GlobaleCommercePostalCode Type="Active">
<GlobaleCommercePostalCompanyCode>XX-001</GlobaleCommercePostalCompanyCode>
<GlobaleCommerceUserCode>1000000001</GlobaleCommerceUserCode>
</GlobaleCommercePostalCode>
</xAL>
</Record>
</XNAL>
```


Figure 1: Example of xNAL-formatted postal address (Note: the record has been changed due the blind-review requirements)

After successful introduction of the xNAL standard in particular postal authority, its users can submit their name and address data for further validation against the xNAL standard. This postal address cleaning and verification procedure must ensure very good results, because this process is controlled by local postal company, not by any international postal body or private organizations. After successful verification of data, each user (citizen) is provided with the unique xNAL record, consists of three major components: country code, postal company code and, most important, global e-commerce user code, sometimes already integrated in national system of postal addresses and its services (track-and-trace, for example). Otherwise, the global e-commerce address code and “traditional” address code are placed in the same xNAL record and have different values for postal company and e-commerce business (postal company can easily translate the global e-commerce user address code to the “traditional” address code, already integrated in the mail processing systems and mail piece-related applications and services).

5. Web Services in the Semantic Web

Traditional knowledge management techniques have faced new challenges by today’s Internet: information overload, the inefficiency of keyword searching, the lack of authoritative (trusted) information, and the lack of natural language-processing computer systems. The semantic Web can bring structure to information chaos. The method presented in the paper enables all “players” in the e-commerce value chain to use tagged postal address data with computer-understandable markup, and every side would be able to know what data is authoritative. It enables e-commerce companies to have proof that they can indeed trust the provided name and postal address data, and then they need to be able to correlate it with the other information that they already have. Finally, they need the tools to take advantage of this new knowledge.

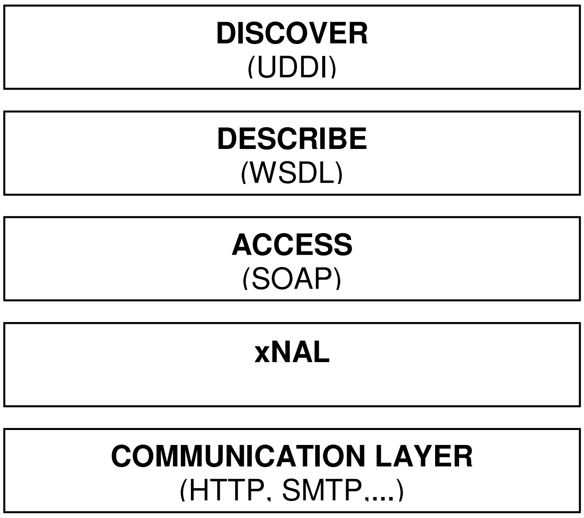

Figure 2: The components of Web services

Web services are software applications that can be discovered, described, and accessed based on XML and standard Web protocols over intranets, extranets, and the Internet. Web services are software applications available on the Web that perform specific functions. Built on XML, a standard that is supported and accepted by thousands of vendors worldwide, Web services first focus on interoperability. XML is the syntax of messages, and Hypertext Transport Protocol (HTTP), the underlying protocol, is how applications send XML messages to Web services in order to communicate. Web services technologies, such as Universal Description, Discovery, and Integration (UDDI) and ebXML registries, allow applications to dynamically discover information about Web services. The message syntax for a Web service is described in WSDL, the Web Service Definition Language. When most technologists think of Web services, they think of SOAP, the "accessed" part of our Web services definition. SOAP, developed as the Simple Object Access Protocol, is the XML-based message protocol (or API) for communicating with Web services. SOAP is the underlying "plumbing" for Web services, because it is the protocol that everyone agrees with.

If e-commerce companies know where Web services are and know what they do, they can easily get software to communicate with them. Finding Web services based on what they provide introduces two key registry technologies: UDDI (Universal Description, Discovery, and Integration) and ebXML registries. UDDI, introduced in 2000 by Ariba, Microsoft, and IBM, was created to facilitate the discovery of business processes. Introduction of UDDI technology in the solution presented in this paper would be one of the main tasks of postal companies; however it must be coordinated by utmost industry authority like the Universal Postal Union. So, postal organizations would register public information about their Web services of postal address management with UDDI, and applications could view information about these Web services with UDDI.

One of the most advanced characteristics of the presented solution is orchestration of Web services. Orchestration is the process of combining simple Web services to create complex, sequence-driven tasks. This process involves creating business logic to maintain conversations between multiple Web services. Orchestration can occur between an application and multiple Web services, or multiple Web services can be chained into a workflow, so that they can communicate with one another. We can imagine a number of possible implementation of this feature – one of them is the integration of track-and-trace Web services of all postal companies in the e-commerce logistics chain, even in case that every postal company runs its own track-and-trace system on different platform and with different business process workflow. Another feature of the method and in the same time an example of orchestration would be a possibility to validate the shipping data against the entered credit card address data and add another layer of fraud protection. To underline, applications and Web services built around the presented solution can be used on all operating systems (Windows, Linux, Solaris...) and with all major Web and mobile programming languages (JAVA, ASP.NET, PHP etc). Therefore, there are no problems to add the method (in the "plug-in" format) to existing shopping carts and B2C auction software.

Finally, let's discuss about the taxonomy and ontology of the presented method. Taxonomy could be defined as the classification of information entities in the form of a hierarchy, according to the presumed relationships of the real-world entities that they represent. Looking at the Figure 1. we can recognize a number of the real-world entities structured in a few hierarchical levels and rigorously defined position of appearance. The most common use of taxonomies is thus to browse or navigate for information, especially when we only have a general idea of what you are looking for. With well-developed taxonomy incorporated in the presented method, users can browse and find the provided postal address data very successfully. Taxonomy in the presented method is a way of structuring postal address data, information entities, and of giving them at least a simple semantics. On the Web, it can be used to help e-commerce companies find customers information. Taxonomies can also help you get a handle on your own information needs, by classifying your interests (whether they include products and services or not). Because taxonomies are focused on classifying content (semantics or meaning), they enable search engines and other applications that utilize taxonomies directly to find information entities much faster and with much greater accuracy. So, Internet user has to provide only two mandatory data during the checkout and shipping info procedure – own global e-commerce postal company code and global e-commerce user code. For many online businesses the requested data will be only information they will ever know, because they are able to outsource its supply chain management entirely to postal company of their choice. The presented method can be a component of a system for controlling the entire delivery chain of a company's online sales, acquisitions, stockholding, deliveries, invoicing and book-keeping. Furthermore, they can request the "traditional" view of customers' name and address data using secure and capable Web services of postal company. Moreover, it is possible to integrate provided Web services with internal direct marketing initiatives or value-added options in fulfillment procedures (track-and-trace, SMS notification of expected delivery, change of destination etc.).

Ontologies are complex to build as a whole in one step. So, building them step-by-step, empirically testing each new (kind of) addition and developing appropriated learning techniques for each step, businesses may automate the process; and, next time, they may build a new one in a more systematic way. Therefore, the presented solution should be seen as a starting point to the final one-ontology postal address management solution.

6. Security and Privacy Issues

Obviously, there may be security concerns about placing information in a public registry. Although it can be seen that such a technology could be very powerful in a dynamic e-business environment, placing name and address information in a public registry may be risky.

One of the biggest concerns in the deployment of Web services today is security. In a distributed Internet environment where portals may talk to other Web services, which in turn talk to other Web services, how can we know the identity of who's getting the information? How can we know what information that user is allowed to see? With online transactions, how can we have some assurance that the transaction is valid? How can we keep sensitive information transfers confidential? How can we prove, in a court of law, that someone accessed information? How can we know that a user's transmission hasn't been intercepted and changed?

With the presented method of postal address management in e-commerce, huge amounts of structured information are collected and shared globally by and among postal companies and authorities, individuals and e-commerce businesses. First, using this method of postal address management in e-commerce individuals agree that privacy-as-secrecy is not secured; they look for privacy-as-control, the best available form of privacy in this case. Privacy-as-control is at the heart of the EU's Data Protection Directive, the act with the main objective to allowing individual customers to control how their personal (name and address) information is gathered and used. Today, customer name and address data always costs something to collect and maintain, however it only has value when the customer gives permission for its use. On this premise individuals decide whether they want to use the presented method or enable particular online business to get "in-depth" view of their personal data; this is the very first step in the presented method's workflow. Web services technologies provide a very good basis for integration of privacy-as-control tools and strategies, like authentication, authorization, single sign-on, confidentiality, integrity and nonrepudiation.

7. Conclusions

Electronic commerce knows no boundaries, and neither do the solutions facilitating it. The presented method of postal address management in e-commerce is a truly global solution. Thanks to the xNAL standard, it can handle names and addresses of any country in a detailed level or in an abstract level, and provide a single customer view at anytime, over all available channels (Web, email, call center, direct mail).

Initially, the method and its applications were developed with only one major goal in mind – at the point of interaction with the customer (checkout procedure on website or call center) it should virtually eliminates the possibility of entering wrong name and address data – all data are validated against data supplied by the national postal authorities, reducing the costs and improving the quality of service. Further work on the method was focused on possible "translation" in the semantic Web-ready solution, so in this paper we tried to connect the method with the main components and features of the semantic Web. The xNAL's greatest strength is also its biggest weakness. It is so flexible that different vocabularies/structures can be written to define the same type of data. The xNAL based method of postal address management in electronic commerce will succeed only when businesses or processes agree on a standard XML vocabulary and the principles of the semantic Web can contribute in many valuable forms. Future work will include the closer examination of ontological issues and possible ways for better integration with it and many other important "faces" of the semantic Web.

References

The OASIS Customer Information Quality TC website, http://www.oasis-open.org.

King, A., (2001): PostEurop Plenary Assembly and Postal Forum, "The Postal Industry 2010", Sarajevo.

Berners-Lee, T., Hendler, J., Lassila, O., (2001): The Semantic Web, Scientic American.

Patel-Schneider, P., (2004): Fifth Mexican International Conference in Computer Science, "The Semantic Web and Knowledge Representation", Colima.

Daconta, M., Obrst, L., Smith, K., (2003): " The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Managemen," Wiley Publishing, Indianapolis.

The United Postal Union website, http://www.upu.int.

Grein, M., (2001): ECCMA Meeting Addressing Standards in the New Millenium, “Piecing Together the Address Puzzle”, Santa Clara.

The ECCMA website, http://www.eccma.org.

The CEN TC 331: Postal Services website, http://www.nen.nl/cen331.

Lubenow, J., (2001): International Address Management Issues are Focus of a Global Effort, Cross-Border Report, Vol. 3, No. 3.

Kumar, R., (2002): XML Standards for Global Customer Information Management, DM Review, May Issue.

Martin, H., (2004): Re-defining data, Postal Technology International, March Issue.

Hossapple, C., (2002): A collaborative approach to Ontology design, Communications of the ACM, February Issue.