

The Universal Postal Union (UPU) international address standard, International Postal Address Components and Templates, designated as UPU S42, has now been published, having been approved for publication and further testing by the UPU Standards Board in November 2002. The S42 standard was developed in the UPU POST*Code group under the leadership of Guy Goudet and in its technical committee led by Ruth Jones of USPS.

The standard is based upon a comprehensive list of name and address elements that originated in the work of the European standardization organization CEN, which has an agreement with the UPU to work together on postal standards. These elements define the smallest meaningful parts of names and addresses. The set of elements is extended as necessary to cover additional situations, but so far has been sufficient to represent names and addresses in a number of non-European countries, including the US, taking account of some terminological differences.

A second major concept within UPU S42 is the address template, which describes unique combinations and orderings of elements, or in more general terms, address types, within a country. Templates in UPU S42 are described both in natural language and using an XML format known as the Postal Address Template Description Language (PATDL). PATDL supports multiple sub-templates with branching based on field values, business rules, decision tables, or other defined algorithms. Templates refer to elements by their names or by using codes assigned by the UPU, and can also utilize externally defined elements or code sets. Templates for some countries, such as the United Kingdom, are substantially more complex than for others, such as the United States. By using the templates, the names and addresses can be stored in a permanently parsed format and reconstituted when necessary according to the requirements of a specific situation.

The third major part of the standard has to do with rendition, or the production of addresses on an output medium such as an address label or a computer display screen. Included in the standard is a registry of rendition instructions, which can be formatting rules for final presentation, including abbreviation and prioritization of data elements when there are constraints on available space, and upstream procedures designed to govern the rendition process as a whole, to decide among alternatives, or to implement user preferences. A simple example of rendition instructions is the formatting of a postal code, while a more complex example is the movement of apartment information as recommended by the USPS to the line above the street address as an alternative to abbreviating or shortening the street name or omitting the apartment information.

It turns out that within a template an element such as the UPU “thoroughfare qualifier” may have multiple occurrences in different positions, such as pre-directionals and post-directionals in US addresses, and other elements such as the UPU “postcode” need to be divided into parts in order to be properly rendered, such as the US ZIP+4 code with its hyphen after the first five digits. These situations have in common that they raise issues of cardinality not dealt with in the list of elements itself. The POST*Code group agreed to define element sub-types in order to handle the issue of cardinality in both forms by making it possible to represent any multiplicity or subdivision of elements in the templates. These element sub-types are explicitly defined within the standard as the need for them is recognized.

Through surveys and discussions at the UPU, it has been learned that at least twenty countries either have or are developing a delivery point database. By this is meant a full definition of the specific addresses to which deliveries are made, without resort to summaries, range files, or other methods that cause loss of information about whether a certain set of address elements represents a complete and correct address. Without such a database, the technology that the UPU standard facilitates can only distinguish between addresses that might be valid and those that are definitely invalid. But with a delivery point database, the same technology can distinguish between the addresses that are valid and those that are invalid. Current approaches to address maintenance typically store composite address lines and cannot always make those key distinctions correctly because they require an additional step of parsing the address elements and mapping them to the database fields, which can fail if there are extraneous, misplaced or ambiguous elements.

The standard needs additional testing and development of more templates before it can be utilized on a worldwide basis. Currently fourteen countries have agreed to participate and half of those countries have provided mappings of elements, natural language templates, basic rendition rules, and sample addresses representing the known address types that involve different orderings of elements. There are two approaches to deriving the formal PATDL XML template from the inputs provided, which can be deployed separately or together. One is to translate directly from the rules implicit in the natural language template, and this works if that template is sufficiently precise and complete. The other is to generalize upward from the sample addresses to find a PATDL template that can generate all the renditions correctly, and this works if the sample is robust enough. Actually some combination of deductive and inductive approaches is needed in order to ensure that the template is capable of accomplishing the objective of properly formatting all valid addresses for the country. Thereafter any template may be further elaborated and customized with options and user preferences. Both the natural language and PATDL templates will be published by the UPU if appropriate approvals have been obtained.

The standard can be obtained from the UPU only by ordering the complete set of UPU standards, which costs 400 Swiss francs for a one time distribution either on hard copy or CD-ROM, or 750 Swiss francs for an initial distribution and a one year subscription to receive updates that are normally issued quarterly. The URL to start with is www.upu.int/publications/en/upu_technical_standards.html. This reporter would recommend the subscription since updates to S42 are expected to occur as work continues. Understandably, some may prefer that the standards be made available separately and that they be directly downloadable, but at this time neither of these options is offered by the UPU. The S42 standard may be of greater interest to industry users than many of the other UPU standards which are focused more exclusively on internal postal needs.

What are the different uses that can be made of the UPU S42 standard? Some users may be primarily interested in the element list, others in elements and templates, and still others in elements, templates and rendition. The element list in itself serves as an excellent foundation for consolidating separate country-based addressing applications around a common data definition. Some users of related addressing standards such as OASIS xNAL and HR-XML may be able to benefit from the templates, assuming that a mapping

between the element lists is constructed. Those who mail across borders can utilize the registry of rendition instructions to learn about formats for multiple countries from a single source. Beyond that there are extensions that will add to the scope of what can be implemented. This reporter has participated directly in the UPU POST*Code process and is also the chair of an IDEAlliance work group developing an implementation of the UPU S42 standard within the broader context of business mail. This project, known as the Address Data Interchange Specification (ADIS), facilitates many additional user options and extensions while supporting all the elements of UPU S42 and the PATDL template description language. While UPU S42 only uses XML to describe templates, ADIS also allows address data to be described in XML. The upcoming 2003 version of ADIS is expected to offer many useful extensions to the published UPU S42 standard.

Available within ADIS are the alternatives of an XML or a relational database architecture, use for preparation of a specific mailing or for long term address management, support for bulk mailings as well as collections of individual mail pieces, and the option for fully element based storage or to carry out a transition from line-by-line methodology. Features supported include documentation needed for qualification for postal rates, multiple production groups to support multiple countries with different address formats within a single campaign, addition of mail production variables to the address block, and formatting of ink jet messages used for marketing purposes. Application areas that are enabled include late address changes after the mailing file has been presorted, support for combined mailings of different levels of granularity, and support for intelligent mail with sender identification and the capability to request ancillary services through mail piece barcodes.

When addresses are originally collected, they are often obtained in a line-by-line format. It is unlikely that a Web site would be designed with separate fields holding several dozen name and address elements. However, immediately upon acquisition, and while direct feedback is still possible, the UPU S42 technology will help accomplish permanent parsing and validation of addresses. It facilitates address updating and matching to databases, and when a mailing is planned, allows for simpler and quicker de-duplication. During mail production or rendering on a computer display, the UPU S42 element based technology expressed through a rendition engine, such as ADIS specifies, enables the final presentation of the address to achieve the best possible quality. Through this new methodology, address quality can approach more closely to the elusive goal of 100% correctness that provides greater efficiency for both postal services and their customers.

For information on POST*Code or UPU S42 contact Guy Goudet at (+41 31 350 3156) or Ruth Jones at (+1 901 681 4585). For information on IDEAlliance, contact David Steinhardt at (+1 703 837 1066). For questions or comments on any subject in this report, contact Joe Lubenow at (+1 773 478 2249) or lubenow@msn.com.

12 June 2003