

The trouble with standards

Why you're not likely to be using one for your personal names and addresses for a while yet.

Many attempts have been made to create standards related to personal name and address data, and moves in this direction are gathering pace. The chances, however, that you or your company is using any of these standards is slim, and if you are, it is likely to be a standard covering only a small area of personal name and address management, such as XML data exchange or the rendering of addresses on envelopes for a small number of countries.

Why is this? Personal names and addresses are inherently so complex and flexible that the creation of any standard is difficult, and the creation of a workable standard is nigh on impossible. I do not dispute that having universally accepted and practically applicable standards is useful in any area. For personal names and addresses standards can lead to easier data exchange, improved deliverability of mail items and improved data manipulation for such processes as de-duplication. However, with every advantage that a standard brings comes a disadvantage, which often creates a major barrier to wide uptake of that standard.

These disadvantages will be here for a long while to come, and I believe that whilst they are there the uptake of standards will always be small. Some changes in the process of standards creation, however, may well improve uptake. This article takes a look at the process of standards creation and suggests some areas for improvement.

Standards are not standard

Even where standards have been created, uptake of those standards is usually very low. For me, a standard may only call itself that if it is freely available, has general universal acceptance and practical applicability. A standard such as the ISO-3166 country code standard has almost universal acceptance and is freely available, but due to a number of significant flaws some users have to make in-house alterations to its application, damaging the aspect of practical applicability. This may also be said to apply to the Unicode standard for representing languages, which is accepted and used by most large corporations but which has still to become the norm for smaller businesses and individual users.

Why are personal name and address standards not working?

In seeking to explain the lack of take-up of personal name and address standards I believe we should seek flaws in the standards creation process as well as in the practical applicability of the standards themselves.

Too many standards. There are already a large numbers of standards for parts of personal name and address management, covering certain issues, such as data exchange, address output format or the area of an envelope which may be printed; or certain geographical areas, such as countries or trading blocks. Some of these standards are “competing” inasmuch as they cover the same issue but the standard is different. Often, groups of users are unaware of the existence of a standard or of the knowledge required to create the standard and they spend too much time and effort re-inventing the wheel, covering ground already known to other authorities.

A smaller number of standards with integration between the existing standards would be more helpful to the average user. Standards also need to be more accessible to potential users, who often have to search long and hard to find information about these standards. Once a standard is in place it needs to be “marketed” and promoted to an extent that uptake will spread beyond the companies involved in creating the standard in the first place.

Finance. Creating standards does not produce profit in the short term, and pressure for the creation of standards usually comes from users themselves. Standards committees have usually, therefore, to be self-financing. Members of committees may be expected to pay hefty membership fees in order to become committee members, and to finance international travel and hotel expenses to attend the committee meetings. This alone can often limit membership of standards committees to representatives of large and rich corporations. Furthermore, to recover costs, most organisations charge companies to view their standards. No standard will ever become universally adopted unless it is provided free-of-charge.

Committee membership. As already mentioned, given the financial considerations, standards committee members usually represent large and rich corporations. Surprisingly, users and people highly knowledgeable about the subject under discussion, at least in the personal name and address field, are rarely asked to join these committees, and most committees have not yet found ways of consulting non-committee members about the issues under discussion without applying unacceptable financial burdens to those involved. This sometimes gives the committees the air of being cliques, and the resulting standards are often useful for those corporations involved but not for the large mass of users who have not been able to have their input. Opening committee membership, and accepting more knowledge from non-committee members would help in preventing committees from covering areas already researched by other individuals and committees, and would make the resulting standards more applicable and usable.

Standards evolution. It is quite right that standards evolve and also quite right that standards are published for peer review. However, standards are sometimes released while still being developed so that they cannot be universally applied. When early adopters try to apply the standards and find them unusable, they may permanently reject the standard. Early versions are also too often overtaken by dynamic changes in global name and address systems. In personal name and address standards, for example,

versions are sometimes released which cover some countries and not others. This is hardly useful for companies wrestling with the problems of managing an international database. Standards committee need to be sure that their standard has realised a certain level of coverage and applicability before releasing it, and the limitations of all standards should be more honestly and clearly mentioned in the accompanying documentation - this happens at present all too rarely. This is an issue of expectation management, covered below.

Managing expectations. Confronted daily with the complexities of handling personal names and addresses, many data managers will grab at any proposed solution they find. This being the case, some standards need to be a little clearer in the aspects that the standard concerned covers. Standards which appear at first glance to be an all singing all dancing standard for personal names and addresses prove, upon deeper inspection, to cover only a small part of the issue, such as data transfer between postal authorities, or field naming in a specific programming language, or the area of an envelope which may be printed with the address block. These are valuable contributions in their own rights to our subject area, but often need to be presented in a clearer way.

Cultural clashes It is inevitable that the cultural background of the committee members involved in the creation of any standard has a deep effect on the resulting standard, with a standard, for example, covering initially only part of the globe and afterwards being tweaked to cover other countries. One exchange between members from different countries in one standard is instructive. Country *a* requested the removal of certain address components from the standard because they don't have them in their addresses (a highly worrying input on a committee attempting to create an international standard!), whilst country *b* requested a new field which it required and which the committee refused to implement. The resulting structure is a compromise which country *b* in any case will be unable to accept and implement, and which can therefore not be regarded as a standard. In another standard, personal name parts have been tagged as *first name* and *last name*. As people from different cultures write their names in different orders, this Anglo-Saxon approach will be inappropriate and confusing for people from other cultures.

Practical applicability. I would like to be able to say that if the issues raised above were resolved a personal name and address standard could be produced which everybody would be happy with and would be able to apply. However, after almost 15 years of research in this field I remain to be convinced that any standard can be practically applied beyond small parts of the whole issue, such as portion of the envelope printed and field tags for data transfer. Whilst committees struggle to standardise address rendition (the order of address elements on an envelope), for example, there are very many countries in the world where the postal authorities themselves still do not require data in a fixed pattern or who have not codified this information. If they have not done this, what chance that the users under their jurisdiction will apply any standard? Remember, every single person on this planet is a name and address user - what chance that they are going to accept and apply a standard imposed upon them? Standards are as much about the people who will use them as about the hardware and software systems that they will be

used upon. Even if every postal authority applied a standard tomorrow, cultural inertia is such that I doubt that application of the standards would occur for some decades. Though there have been clear guidelines for many years about how my address in The Netherlands is to be rendered, no single American or British company is able or willing to make an effort to render it correctly. Equally, whilst the Universal Postal Union (UPU) is investigating ways of standardising address rendering, its members are happily introducing postal code systems and rendering rules in quite un-standardised ways.

Our cultural diversity dictates that our personal name and address systems are highly diverse, and are unlikely to converge into a single system. Most standards recognize this and attempt to encompass this cultural diversity into their standards. This, in turn, leads to the inability to practically apply the resulting standard. For example, a standard might recognize that a house number may be split into the building number, door indicator, floor indicator, staircase indicator and wing indicator for data transfer. Anybody who has attempted the parsing of addresses will know how incredibly difficult it is to do this even manually, let alone programmatically, with any degree of accuracy. It is problematic enough for a single country - for global data the difficulty is multiplied many fold. Equally, a data entry form with these fields will leave data-entry staff or users totally perplexed and guarantee the worst quality data. In effect, the standard becomes impossible to apply to anybody except for data originators, such as postal authorities. Users are forced to compromise the standard in order to achieve a practical aim. Faced with a choice of wanting to follow the standard to enable effective data exchange and improved control over their data quality, but by doing so increasing the cost, time and knowledge required to manage their data, most companies choose not to follow the standard. This issue is not likely to be resolved in the short term.

Graham Rhind is an acknowledged expert in the field of data management. He runs his own consultancy company, GRC Database Information, based in The Netherlands, where he researches postal code and addressing systems, collates international data, runs a busy postal link website and writes data management software. Graham speaks regularly on the subject and is the author three books on the topic of international data management.

<http://www.grcdi.nl>
graham@grcdi.nl

© Graham Rhind 2002