

Applying RDF Site Syndication (RSS) and Zthes to International Standards Organization (ISO) Standards.

By Ronald P. Reck
RRecktek LLC
Chantilly, VA. 20151, USA

ABSTRACT

Consumers concerned with adhering to International Standards Organization (ISO) Standards are all faced with the same two challenges. There is no uniform way for users to identify changes to the standards, nor does ISO supply the values in a “standard” electronically accessible format.

This paper proposes using a “standards” approach to support the use of standards. Clearly, it would be reasonable for there to be an automated subscription system for notifying interested parties of updates. The syndication of information through RDF Site Summary (RSS) would allow users to subscribe to an information feed and stay abreast of current changes for any of the standards in which they are interested.

Many standards are only published in the Portable Document Format (PDF) or Microsoft Access database format. While the PDF format can ensure that information is not easily modified, it is not very electronically useful. Users expecting to electronically access any of the values from an ISO standard would need to design a process for obtaining the values from either of these two proprietary formats. The additional effort associated with extracting information from these proprietary formats could easily be avoided if the standards were supplied in an XML format such as the family of specifications called “Zthes”. Zthes is intended for use in presenting thesaurus and hierarchical information in XML and could be useful as a format for publishing ISO standards.

Keywords: controlled vocabularies, data standards, International Standards Organization, metadata, RDF Site syndication, XML

Introduction

Given the modern day reliance on computers, companies, organizations and individuals of the twenty-first century produce significant amounts of electronic data on a daily basis. This information can be a powerful asset, but the challenge that information architects and system

implementers face is to make use of information with the least amount of effort and resource expenditure. The value of electronic information is increased when the cost of using it is controlled because data quality is consistent and predictable.

One strategy for ensuring good data quality is to define both a data format and a range of acceptable values. While it takes energy to specify a format and enumerate values, it can be well worth the cost. Those willing to make this type of expenditure are fortunate because several data formats and possible values are already defined and articulated in the form of standards. The International Standards Organization (ISO) offers standards for many domains in the form of country codes, units of measure, and currency codes just to name a few.

Given that a community chooses to use an ISO standard to articulate the format and acceptable values some challenges still persist. The first major challenge users face is that there is no formal schedule nor calendar for the release of standards nor is there a notification mechanism for letting communities who depend upon the standards know when ISO standards have been updated. Hence, one challenge users tackle is staying current with the standard itself. Standards are not updated extremely often, yet people who depend upon them need to know when changes have been made. Whether they are producers or consumers of information it is important that they know when new versions of standards are released. Currently, users need to repeatedly check the ISO website when simple mechanisms exist for news syndication using technologies such as RDF Site Syndication (RSS) or ATOM.

Simple XML based syndication mechanisms have been used successfully for years, but unfortunately organizations like ISO have not yet adopted them. In time it is likely they will adopt these mechanisms but in the meantime it is up to users to manually check for information about standard’s updates.

A second limitation in the adoption of standards is that ISO does not consistently present code lists using a standard format. Several formats exist for the

representation of code lists like ISO 2788 or the family of specifications known as Zthes. Using a consistent format for the presentation of ISO code lists would be a useful strategy for limiting the costs associated with using standards and thereby would be likely to increase their adoption. Unfortunately, ISO has not adopted this viewpoint but explanations about the merits may indeed provide a compelling enough argument to motivate this change.

Notification: RDF Site Syndication & ATOM

Since it is not always clear when a new version of the standard will be produced it makes sense to use modern day mechanisms. Syndication notification like RDF site syndication (RSS) is one example.

RSS uses an XML-based format that leverages the Resource Description Framework (RDF) for supporting the syndication of lists of hyperlinks, along with other information and metadata. RSS allows a person's computer to retrieve and understand information, so that a consumer can easily track, on a regular basis, updated information. It is a format that's intended for use by computers on behalf of people, rather than being directly presented to them (like HTML). Browsers allow a user to subscribe to an RSS feed and then quickly and easily monitor that feed for updates.

An example of what a RSS notification for an update of an ISO 3166 country code list would look like Figure 1.

```
<?xml version="1.0" encoding="windows-1252" ?>
- <rss version="2.0">
- <channel>
  <title>ISO 3166 Country Codes</title>
  <description>The international standard for Country Codes ISO 3166 is one of the most widely used standards maintained by ISO TC 46. It provides a standard numeric and 2-letter and 3-letter alphabetic codes for 237 countries or areas of special sovereignty. First released in 1974, ISO 3166 has grown to encompass three parts, including two new sections on codes for subdivisions (states, regions, major cities, etc.,) and a listing of retired codes.</description>
  <link>http://www.iso.org/iso/en/prods-services/iso3166ma/index.html</link>
  <copyright>ISO/IEC 2007 All Rights Reserved.</copyright>
  <language>en-us</language>
  <managingEditor>countrycodes@iso.org</managingEditor>
  <pubDate>Tue, 01 Jan 1974 12:00:00 GMT</pubDate>
  <webMaster>countrycodes@iso.org</webMaster>
  <generator>Vocabutek (1.1)</generator>
- <item>
  <title>ISO 3166 has been updated to revision 5</title>
  <description>There have been 4 records updated, 1 deletion and 1 addition</description>
  <link>http://www.iso.org/iso/en/prods-services/iso3166ma/03updates-on-iso-3166/index.html</link>
  <pubDate>Mon, 22 Jan 2007 12:00:00 GMT</pubDate>
</item>
</channel>
</rss>
```

Figure 1: An example of an RSS Notification for an ISO 3166 country code list update.

Similarly, an example of what an ATOM notification for an update of an ISO 3166 country code list would look like is shown in Figure 2.

```
<?xml version="1.0" encoding="utf-8" ?>
- <feed xmlns="http://www.w3.org/2005/Atom">
  <title>The international standard for Country Codes ISO 3166</title>
  <link href="http://www.iso.org/iso/en/prods-services/iso3166ma/index.html" />
  <updated>2003-12-13T18:30:02Z</updated>
- <author>
  <name>International Standards Organization</name>
</author>
  <id>urn:uuid:60a76c80-d399-11d9-b93C-0003939e0af6</id>
- <entry>
  <title>ISO 3166 has been updated to revision 5</title>
  <link href="http://www.iso.org/iso/en/prods-services/iso3166ma/03updates-on-iso-3166/index.html" />
  <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
  <updated>2007-01-22T18:30:02Z</updated>
  <summary>There have been 4 records updated, 1 deletion and 1 addition.</summary>
</entry>
</feed>
```

Figure 2: An example of an ATOM Notification for an ISO 3166 country code list update.

Creators of ATOM and RSS can validate the format of their messages using the W3C's validation website at <http://validator.w3.org/feed/check.cgi>.

Code List Representation: Zthes

According to the Zthes website at <http://zthes.z3950.org> the Zthes family of specifications are intended to "facilitate interoperability for applications that deal with thesauri - semantic hierarchies of terms as described in ISO 2788 and ANS/NISO Z39.19." This means that as a representation modality, Zthes provides more than sufficient semantics for representation of code lists such as ISO 3166 (country codes), ISO 639 (language codes) or ISO 4217 (currency codes).

The Zthes abstract model may be represented using the extensible markup language (XML). An XML document in this format consists of an <Zthes> element which contains a sequence of <term> elements, each representing a single term. Figure 3 shows an example of how a term an ISO 3166 country code term would be represented using Zthes.

```

<?xml version="1.0" encoding="utf-8" ?>
- <Zthes>
- <term>
  <termID>931375</termID>
  <termName>208 (1974-12-01)</termName>
  <termVocabulary>ISO 3166-1 Numeric</termVocabulary>
  <termStatus>Active</termStatus>
  <termApproval>Standard</termApproval>
  <termLanguage>English</termLanguage>
  <termType>PT</termType>
  <termModifiedBy>ISO</termModifiedBy>
- <relation>
  <relationType>NumericFor</relationType>
  <termID>928520</termID>
  <termName>DENMARK (1974-12-01)</termName>
  <termVocabulary>ISO 3166-1</termVocabulary>
  <termStatus>Active</termStatus>
  <termApproval>Candidate</termApproval>
  <termLanguage>English</termLanguage>
  <termType>PT</termType>
  <termModifiedBy>ISO</termModifiedBy>
</relation>
</term>
</Zthes>

```

Figure 3: ISO 3166 country code term represented using Zthes.

Since Zthes can be represented in XML it lends itself well to the representation of code lists. Past implementations for the representation and dissemination of code lists in the Zthes format have been successful for companies using the Vocabutek Vocabulary Management System (<http://www.vocabutek.com>). Dozens of code lists have been successfully represented in the Zthes format, and XSLT transformations have easily transformed Zthes into other more semantically expressive formats such as RDF SKOS.

The Zthes specification is currently at revision 1.0 and can be considered mature as it has been developed over the last 6 years.

Code List Representation: Genericode

Since the original inception of this paper an XML format has emerged that is designed primarily for the representation of controlled vocabularies. It offers such promise that it will likely replace Zthes as the best possible open standards representation format for controlled vocabularies. This XML format is called “genericode”. The project’s website at <http://www.genericode.com> describes the efforts thusly: “Genericode started as a project to define a standard format for defining code lists (also known as enumerations or controlled vocabularies). Code lists occur throughout all parts of IT (information technology), but there isn’t a single standard way to manage them. This means that each new standards effort creates its own new solution, and this has caused a *balkanization* of the management of code lists.”

The orientation as a format intended expressly for representing code lists means that genericode is likely the premier XML format for controlled vocabularies for the foreseeable future. Users who are looking for a robust XML format for controlled vocabularies would do well to carefully consider it before adopting any other format.

While space constraints prohibit the presentation of a complete sample of genericode in this paper, an abbreviated sample is presented in Figure 4. This sample contains much of the same information that the Zthes sample does. An important component and strength of genericode is omitted but should be mentioned. Genericode permits the definition of columns and data types similar to a database. Hence, this provides the ability to clearly represent the diversity of a wide range of values with more precision than is possible with Zthes.

Irrespective of format for code lists in either of the XML formats Zthes or genericode, ISO would decrease the cost for utilizing standards if they were to present all their standards in a standards’ based format that did not require the use of proprietary formats like PDF or MS Access. Many users do not have the means to extract values from PDF, and other users do not necessarily work with Microsoft technologies.

Conclusions

The value of using standards is balanced with the cost of adopting them. If standards’ communities lessen the resource expenditure necessary for working with the standards it is likely to increase their adoption. As is often the case, proponents for a technology have been focusing on the technology for so long they forget what the barriers are for entering their community. If more consumers articulate the need for a standard representation format for ISO standards, as well as a practical notification mechanism, then it is likely that the ISO community will listen.

Acknowledgements

This paper was prepared through the kind support of Olga M. Lorincz-Reck, and Dr. Ruth Reck. Anthony Coates was also kind enough to assist in the representation of ISO 3166 values using genericode. Any errors or omissions are surely the fault of the author and not the contributors.

```

<?xml version="1.0" encoding="UTF-8"?>
<!-- Generated from '3166.xml'. --><!-- This is only an example, and should not be used in production. -->
<!-- Produced by Anthony B. Coates, abcoates@mileywatts.com -->
<gc:CodeList xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:gc="urn:oasis:names:draft:genericcode:schema:xsd:CodeList-1.0-pre-2"
  xsi:schemaLocation="urn:oasis:names:draft:genericcode:schema:xsd:CodeList-1.0-pre-2 CodeList.xsd">
  <Identification>
    <ShortName>CountryIdentificationCode</ShortName>
    <LongName xml:lang="en">Country</LongName>
    <LongName Identifier="listID">ISO3166-1</LongName>
    <Version>0.3</Version>
    <CanonicalUri>urn:oasis:names:specification:ubl:odelist:gc:CountryIdentificationCode</CanonicalUri>
    <CanonicalUri>urn:oasis:names:specification:ubl:odelist:gc:CountryIdentificationCode</CanonicalUri>
    <CanonicalVersionUri>urn:oasis:names:specification:ubl:odelist:gc:CountryIdentificationCode-2.0</CanonicalVersionUri>
    <LocationUri>http://docs.oasis-open.org/ubl/os-ubl-2.0/cl/gc/default/CountryIdentificationCode-2.0.gc</LocationUri>
    <Agency>
      <LongName xml:lang="en">United Nations Economic Commission for Europe</LongName>
      <Identifier>6</Identifier>
    </Agency>
  </Identification>
  <SimpleCodeList>
    <Row>
      <Annotation>
        <AppInfo>
          <Extract xmlns="http://www.example.com/Zthes">
            <term xmlns="">
              <termID>931375</termID>
              <termName>208 (1974-12-01)</termName>
              <termVocabulary>ISO 3166-1 Numeric</termVocabulary>
              <termStatus>Active</termStatus>
              <termApproval>Standard</termApproval>
              <termLanguage>English</termLanguage>
              <termType>PT</termType>
              <termModifiedBy>ISO</termModifiedBy>
              <termNote>

```

Figure 4: Genericcode sample of an ISO 3166 code list modified from Zthes.

BIBLIOGRAPHY

- Chaudhri, A., Rashid, A., & Zicari, R. (2003). *XML Data Management: Native XML and XML-Enabled Database Systems Addison Wesley*
- Coates, Anthony B. Why are simple code lists so complex?, XML 2004. Washington D.C., November 2004.
- Daconta, M., Obrst, L., & Smith, K. (2003). *The Semantic Web Wiley Publishing Inc.*
- Hjelm, J. (2001). *Creating the Semantic Web with RDF John Wiley & Sons Inc.*
- ISO 2788 (1986). Available at:
<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=7776&ICS1=1&ICS2=140&ICS3=20>
- ISO 3166-1 (2006). Available at:
<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39719>
- Oasis Code List Representation TC. (2007). *Code List Representation (Genericcode) Version 1.0*. Retrieved from: <http://docs.oasis-open.org/codelist/cd-genericcode-1.0/doc/oasis-code-list-representation-genericcode.pdf>
- Powers, S. (2003). *Practical RDF O'Reilly & Associates Inc.*
- RDF (2004) Resource Description Framework. Retrieved from <http://www.w3c.org/RDF/>
- RFC 1766 (1995). Tags for the Identification of Languages Retrieved from:
<http://www.faqs.org/rfcs/rfc2046.html>
- RFC 4287 (2005). The Atom Syndication Format Retrieved from: <http://www.ietf.org/rfc/rfc4287>
- Sall, Kenneth B. (2002). *XML Family of Specifications: A Practical Guide. Addison-Wesley*
- SKOS (2006) Simple Knowledge Organization System (SKOS) Core Vocabulary Specification Retrieved from: <http://www.w3.org/TR/swbp-skos-core-spec/>
- XML (2006) Extensible Markup Language. Retrieved from <http://www.w3c.org/TR/REC-xml/>
- XSLT (1999) XSL Transformation. Retrieved from <http://www.w3c.org/TR/xslt/>
- Zthes Working Group. (2006) Zthes version 1.0. Retrieved from:
<http://zthes.z3950.org/model/index.html>