Standards to be obeyed by the masses: Representing temporal and monetary data on the Web

1 Introduction

Human beings often make references to time and money. Temporal and monetary data are very relevant in everyday life, in communication between humans, and, thus, also in information systems (IS) and the data exchange between them.

People living in different cultures developed distinct attitudes for representing time and money. It is easier to obtain a common understanding between members of the same community than between members with different cultural background. Standards for representation of temporal and monetary data are helpful to ease communication and to avoid ambiguities and misinterpretations. Whereas many standards are relevant only for a small group of professionals, the audience for standards for representing temporal and monetary data is extremely large.

The Web has become the main medium for global communication. Temporal data are relevant as well in the contents provided for the global readership as in managing Web pages. Web standards help ensure that everyone has access to and a common understanding of the information provided [Web Standards Project (2001)].

In section 2 we give an overview on standards that are available for representing temporal and monetary data. In section 3 we show the relevance of temporal data presented on the Web and discuss the importance of their unique representation for managing Web sites. Section 4 shows that also monetary data are represented very differently on the Web. With respect to the broad audience affected by temporal and monetary data, bipolar representations are suggested that accommodate as well to traditional, widely used forms of local representations as to the global standard notations. Section 5 presents conclusions.

2 Standards for Representing Temporal and Monetary Data

2.1 Standards for Temporal Data

The measurement of time and its representation is one of the most fundamental problems in the history of standardization. A formalized system to measure days is called a calendar. For intercultural communication usually the Gregorian calendar is applied.

Since ancient times, dates are usually referenced as a triple of year, month, and day. Differences in representing temporal data exist because full or (differently) abbreviated texts (e.g., for naming the months) or numerical data may be used. In the latter case, the year may be represented by four, two or even only one digit, the first nine months of the year and the first nine days of the month by one or two digits. Furthermore, the sequence of the components may differ: The first component may represent a day (often in Europe), a month (often in the US) or a year (according to the ISO 8601 standard). Special characters can be used for concatenating the components of the date element. The variety of resulting representations is discussed in Knolmayer/Buchberger (1997). Different systems like the Jewish, Chinese or Muslim calendar [Timechange Creativity Services (2001)] make the identification of a certain day even more difficult.

In 1988, ISO issued its standard 8601[1] that is similar to the ANSI X3.30 standard published in 1985. It defines formats for numerical representation of dates, times, and date/time combinations. Dates can be given in year-month-day, year-week-day or year-day formats. All formats present the larger units first, i.e., from left to right the sequence is year, month, week, day, hour, minute, and second. Any particular date/time format is a subset of these values and the standard lists various permissible subsets: Truncated or abbreviated representations are allowed.

ISO 8601 has been suggested for internal representation of dates in IS but also for human use. Thus it is allowed to write a date either without or with hyphens to separate the elements. Hyphens are superfluous for technical use; however, humans are better able to recognize structured than unstructured displays.

[1] http://www.mcs.vuw.ac.nz/technical/software/SGML/doc/iso8601/ISO8601.html

ISO 8601 has been adopted as European Standard EN 28601; in all EU countries the conflicting national standards have been changed accordingly. For instance, in 1996-05 the German standard DIN 5008 that specifies typographical rules for German texts has been updated: the German numeric date notations DD.MM.YYYY and DD.MM.YY have been replaced by the ISO date notations YYYY-MM-DD and YY-MM-DD [Kuhn (2001)]. In 2000-12-21, the ISO 8601 standard has been updated without generating much public interest.[2]

The Federal Information Processing Standard FIPS 4-1 references the ANSI X3.30-1985 standard[3] and, therefore, indirectly ISO 8601. The most important standard for EDI, established by the United Nations as UN/EDIFACT (Electronic Data Interchange for Administration, Commerce and Transport), provides about 250 segments to specify date, time or period [Axiom (1998)]. UN/EDIFACT makes reference to ISO 8601[4] as several other proposals do.

2.2 Standards for Monetary Data

As for temporal data also for currency symbols one or more local and an international representation may exist. Furthermore, in different cultures distinct characters are used to structure monetary amounts and to express the decimal point. The currency information is sometimes missing, occasionally placed to the left, and sometimes to the right of the amount.

The international financial community uses three capital letter abbreviations defined by ISO 4217 (Codes for the Representation of Currency and Funds) [XE.com (2001)]. These codes have been partly accepted also for local communication. In UN/EDIFACT, more than 20 sections make reference to currency data [Axiom (1998)].

3 Temporal Data on the Web

The World Wide Web is based on several consortium specifications and standards. The Hypertext Transfer Protocol (HTTP/1.1) [Fielding et al. (1999)] is used for communication between servers and clients. Structuring and also the representation of content on the Web are based on the Hypertext Markup Language (HTML). In 1999,

[2] http://lists.ebxml.org/archives/ebxml-core/200104/pdf00005.pdf
[3] http://www.itl.nist.gov/fipspubs/fip4-1.htm
[4] http://www.unece.org/cefact/docum/sessdocs/r1157a2.htm

the W3C Recommendation for HTML 4.0[5] became the basis for the international standard ISO/IEC 15445 HTML.[6] Both are based on ISO/IEC 8879 that defines a Standard Generalized Markup Language (SGML).

Recently, the Extensible Markup Language (XML)[7] gains much attention. Unlike HTML which is based on a SGML document type definition (DTD) and is, therefore, restricted to a set of predefined markup elements, XML allows to define elements for specific purposes and how they relate to each other.

In HTML and also to a certain extent in XML the support of temporal data is rather poor. Thus, for the majority of temporal data on the Web there is no formal requirement how they have to be presented. This leads to grave consequences in using temporal data.

3.1 Relevance of Temporal Data on the Web

Temporal data is abundant on the Web as is illustrated in Figure 1; even if not all hits represent years this will be true for a very high percentage. This section shows that using a standard for representing temporal data on the Web is very important because otherwise the communication between members of different cultures becomes error-prone.

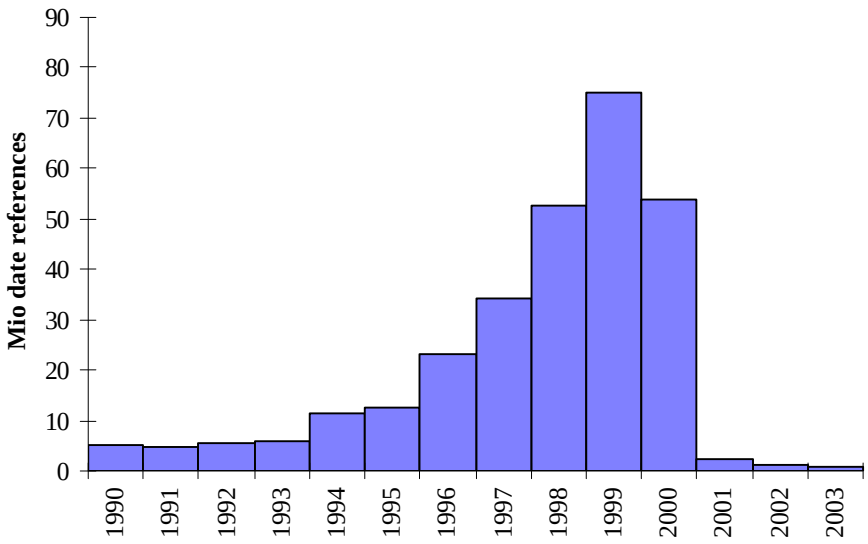

Figure 1: Number of date references on the WWW (as of 2000-05-03)
(*** Graphic will be updated for final version ***)

[5] http://www.w3.org/TR/1998/REC-html40-19980424/
[6] http://xml.coverpages.org/related.html#isoHTML
[7] http://www.w3.org/XML/

Millions of people are developing and maintaining Web-pages and using quite different representations for temporal data in their documents. If we select a certain day, for instance, the American Independence Day in 2001, we see that this date is represented quite differently (Table 1). An expression like 07/04/01 (which was found about 186,000 times by Google [as of 2001-08-27]) may be clear in the communication between, e.g., an American author and an American reader. However, the Internet is a global, worldwide accessible medium and the reader is often unaware of the cultural environment of the author. The string mentioned above may be interpreted in different regions, e.g., as 2001-07-04, 2001-04-07 or 2007-04-01.

Search string	**Number of hits obtained by Altavista (advanced mode)**
07-04-01	53,063
07-04-2001	43,668
04-07-01	41,834
04-07-2001	34,680
01-07-04	21,510
2001-07-04	13,196
July 4, 01	37
July 4, 2001	8,528
Jul 4, 01	1,293
Jul 4, 2001	163
4th of July 01	6
4th of July 2001	65
4. Juli 01	22
4. Juli 2001	701

Table 1: Potential representations of the American Independence Day in 2001 [as of 2001-08-06]

In addition to the presentation of dates on static Web pages many scripts are used that generate temporal data dynamically. This has become evident at the date change 1999/2000 which resulted in Y2K-glitches (Table 2) that became broadly visible on the Web [cf. y2kmistakes.com (2000); Knolmayer/Myrach (2000)].

The problems with a proper presentation of dates on the Web have led to several initiatives trying to bring the Internet world to a common standard for representing temporal data. In 1997-09-15, a proposal has been submitted to W3C from employees of Reuters for using the ISO 8601 standard for date representation on the Internet [Wolf/Wicksteed (1997)]. In general, W3C references ISO 8601 in its recommendations.

Owner of Web page	Erroneous date representation
Yahoo	3-Jan-0
George Mason University	1/3/0 Today@MASON
U.S. Naval Observatory	2 Jan 0100 (693960 days 6 hours 48 minutes 44s) to the Year 2000
Worldtimezone	01-01-100 ... 01-01-100
France2	Mar 01 Fev 100
Lycos	date: 2 jan 100
Nepal Information Center	Current Date and Time in Nepal: 03-01-100 .. 8:53 a.m.
AMD Zone Bulletin Board	Last Post 01-02-100
Startrek	1/1/1900
New York Times	January 1, 1900
Year 2000 Information Center	Updated January 10, 1900
Microsoft MCSE Online Training Kit	Release Date: January 1900
Microsoft Portugal	Ultima Modificacao: 1 Janeiro, 1970
Amazon.co.uk	This item will be released 10 October, 2011.
Netscape Netcenter America Latina	Sabado, 1 de enero de 19100
TerraServer.com	Sunday, January 2, 192000
Gamelan	Monday, January 3, 192000
Nintendo Pokemon	Sunday, January 2, 3900
RDWeb, Datum.nl	het is nu zondag 2 januari 3900, 21:18 uur

Table 5: Erroneous date representations on Web pages at the beginning of the year 2000

The representation of time according to ISO 8601 looks unfamiliar for many persons. If they are aware that the standard exists, they may refuse to use it because they prefer to present temporal data in the way used in their local community. Due to the distributed organization of the Web, nobody can force millions of authors and webmasters to use a standard notation. Most of the Web pages do not even follow the rules of the HTML standard strictly.[8] Thus, a first step in the right direction would be if Web development tools offered an option to translate a date formulated in a non-standard way into the ISO 8601-format. Due to the existing ambiguities, the author or webmaster may be asked to confirm the interpretation proposed. As with other

[8] http://www.ilrt.bristol.ac.uk/people/cmdjb/projects/uksites/survey/

changes that have effects on the masses (e.g., familiarizing people to use the metric system or the centigrade scale) one could add the ISO 8601 representation in parentheses to the locally accepted display.

Web documents do not only contain user-directed content but also meta-data that is used to describe the resource or its content. Meta-data is normally hidden from the user. One has to use specific functions to view the meta-data of a document. User-agents like Web browsers and search machines pick up in particular the attributes of the HTTP header of a document sent from a responding Web server.

Common browsers make subsets of the temporal data available in the document information. By Netscape Navigator 4.77 [en] such information can be retrieved by the command sequence View -> Page Info. The following temporal attributes are shown:

- “Last Modified” (in Local Time and GMT)
- “Expires” (without specifying the time zone)

Using MS Internet Explorer 5.50 the commands File -> Properties provide the following temporal attributes:

- “Created” (date only)
- “Modified” (date only)

For example, for the Web page http://www.iwi.unibe.ch/ the following information was shown by the Netscape Navigator [as of 2001-08-15]:

```
"Last Modified:    Donnerstag, 7. Dezember 2000 10:26:30 Local time
 Last Modified:    Donnerstag, 7. Dezember 2000 09:26:30 GMT
…
 Expires:          No date given"
```

For the same page on the same day, MS Internet Explorer presented

```
"Created:   15.08.2001
 Modified:  15.08.2001"
```

One recognizes the different ways of representing date and time by the two browsers. None of them shows the date according to ISO 8601. Even worse, the information about the date of the last modification differs because MS Internet Explorer does not show the date of the last modification on the Web server but of the last modification

in the cache of the client. This is described in the Microsoft Product Support Services for version 5 of the Explorer in the following way:

"Modified and Created Time Information Is Not Listed Correctly on the File Properties

When you create a file on another computer on your network, and then view the file properties, the created time may be replaced with the last accessed time. Also, the modified time may be listed as the same time as the last accessed time, instead of the last modified time in Temporary Internet Files. … This problem can occur if Internet Explorer obtains this time information from the cache (Temporary Internet Files)."[9]

Some search engines also show the last-modified date. For instance, Altavista can be customized to present the last-modified date. For the page mentioned above this date is given as 20-Feb-2001. It is different from the information shown by the browsers and the format also does not follow ISO 8601.

For http://www.acm.org/ Netscape states

```
Last Modified: Unknown
```

However, Altavista shows

```
Last modified on: 2-Jul-2001
```

and the text mentions the date of today (Friday, August 24, 2001; as of 2001-08-24).

The present way of handling temporal data by Web development tools and user agents is rather confusing. This is especially undesirable if references to online resources are made, e.g., in scientific work. Many citation rules for referencing online resources demand that the date of the last modification is quoted [Lamp (2001)]. ISO 690-2 requires that for electronic documents the date of publication, of update/revision, and of citation has to be given.[10] Many users and search machines want to judge the actuality of a retrieved web document. Because it is not obvious how to get and interpret the respective meta-information it is quite common to introduce at least the date of the last modification in the user-directed content. Because this is often done manually there is considerable risk that this date is not updated correctly and therefore misleading. There are many examples where the last modification date specified in

[9] http://support.microsoft.com/support/kb/articles/Q243/7/04.ASP?LN=EN-US&SD=gn&FR=0&qry=last%20modified%20&rnk=1&src=DHCS_MSPSS_gn_SRCH&SPR=IE
[10] http://www.nlc-bnc.ca/iso/tc46sc9/standard/690-2e.htm#5.1

the HTTP header differs from the information in the content of a document [Knolmayer/Myrach (2001)].

3.2 Support of Temporal Data by Web Standards

This section shows that a common representation of temporal data on the Web is very important because the underlying technical standards do not provide sufficient datatype mechanisms. With respect to the issue of a standard representation of temporal data there is no principal difference whether Web documents are static or have been generated dynamically. This is particularly true if the documents are processed on the server-side and sent to the user-agent as a plain HTML or XML document.

3.2.1 Hypertext Markup Language (HTML)

HTML as well as XML are primarily designed to structure text-oriented contents. The content of a document is decomposed in hierarchically ordered elements that begin and end with tags. Basically, the contents of HTML and XML documents are untyped: Every element is treated as character data. In contrast to general programming languages or database systems no specific data types like DATE are available in HTML. For instance, the content of the following paragraph-element is treated as a string:

```
<p>The ACM symposium will start on 2002-03-10</p>.
```

The existence of a date element in this sequence of characters cannot be exactly determined. To identify a (sub-) string as a date, one has to search for typical patterns of date representations. This is cumbersome if many representation formats exist [Knolmayer/Buchberger (1997)] and it is obvious that a standard representation of dates is very desirable also for managing Web documents.

The DTD of HTML provides the attribute DATETIME. Although the value of this attribute is specified to be character data, the standard informally requires the ISO 8601 format. This is specified in the DTD of HTML 4.0 by the entity "%datetime"[11]:

```
<!ENTITY % Datetime "CDATA" -- date and time information. ISO date
format -->
```

The attribute DATETIME is only used by the elements INS and DEL that may be applied to mark content that has been inserted or deleted. A browser may use these

[11] http://www.w3.org/TR/1998/REC-html40-19980424/sgml/dtd.html#Datetime

elements for emphasizing the parts of the document that are marked accordingly. An example illustrates the intended use[12]:

```
<INS datetime="1994-11-05T08:15:30-05:00"> Furthermore, the latest
figures from the marketing department suggest that such practice is
on the rise.</INS>
```

HTML supports elements for specifying meta-data that may be used in the HEAD-section of a document. One example is the META element that remains by definition without content; meta-data can be defined by attributes of the META element. One may alternatively employ the attributes HTTP-EQUIV and NAME for specific types of meta-data. The HTTP-EQUIV attribute is used to define a header field name for HTTP and its values must correspond to valid HTTP header attributes; the value of the NAME attribute may be chosen deliberately and remains without impact on the transfer protocol.

Because the META element in connection with attributes of the type HTTP-EQUIV manipulates the content of a HTTP header, the value of the associated content attribute must conform to the HTTP standard. HTTP defines several time-related attributes to control caching. The attribute values LAST-MODIFIED and EXPIRES represent date values. HTTP applications allow for historical reasons three different formats for the representation of date/time stamps (Table 3)[13]. Neither of these corresponds to the ISO 8601 standard.

Standard	**Example of Notation**
RFC 822, updated by RFC 1123	Sun, 06 Nov 1994 08:49:37 GMT
RFC 850, obsoleted by RFC 1036	Sunday, 06-Nov-94 08:49:37 GMT
ANSI C's asctime() format	Sun Nov 6 08:49:37 1994

Table 3: Representation of temporal data in HTTP/1.1

A date value in the attribute "content" must conform to either of these notations. Thus, a META declaration for the HTTP attribute EXPIRES may look like[14]:

```
<META http-equiv="Expires" content="Tue, 20 Aug 1996 14:25:27 GMT">
```

This HTML element will result in the HTTP header:

```
Expires: Tue, 20 Aug 1996 14:25:27 GMT
```

[12] http://www.w3.org/TR/1998/REC-html40-19980424/struct/text.html#h-9.4
[13] http://www.rfc-editor.org/rfc/rfc2616.txt
[14] http://www.w3.org/TR/1998/REC-html40-19980424/struct/global.html#adef-http-equiv

3.2.2 Extended Markup Language (XML)

Unlike HTML, XML may be adapted to support temporal elements or attributes. One may, e.g., define a DATE-Element for the markup of explicit temporal data:

```
<p>The ACM symposium will start on <date>2002-03-10</date></p>.
```

In XML it is possible to introduce time-related attributes for certain elements. Thus, one may implement an implicit time-stamping of states or events [Myrach (2001)]:

```
<p valid-to="2002-06-20">We will meet every week at Monday, 10:00</p>
```

Because the Document Type Definition (DTD) of an XML document in general follows the same rules as the DTD of HTML, the document's content consists likewise of untyped character data. Thus, user agents may deduce from the markup that the content of the specified DATE element is meant to be a date but there is no automatic validation whether the content corresponds to a certain date representation or represents a valid date.

In the new W3C Recommendation XML-Schema[15] typed elements are introduced. XML-Schema provides several built-in datatypes related to temporal information, among them date and datetime[16]. Both demand a lexical representation conforming to ISO 8601.

Standardization efforts for a common understanding how meta-data should be declared in XML have led to the Resource Description Framework (RDF)[17]. Within this framework one may employ different standards like the Dublin Core Metadata Standard[18] that also covers temporal information.[19]

4. Currency data on the Web

To some extent the ambiguity discussed above for temporal data also holds for currency data. Table 4 shows that also in this domain a broad spectrum of different, locally used notations exists. When considering the number of hits obtained for the names of the currencies, it must be recognized that not all strings found refer to currencies. To reduce this effect, the advanced search mechanism of Altavista was used and all searches mentioned in Table 4 were supplemented by

---

[15] http://www.w3.org/XML/Schema/
[16] http://www.w3.org/TR/xmlschema-2/
[17] http://www.w3.org/RDF/
[18] http://dublincore.org/
[19] http://dublincore.org/index.shtml.rdf

NEAR (1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 0)

Partial search string	ISO Country Code	Number of hits obtained by Altavista (advanced mode)
USD	US	473,791
$	US	syntax error
dollar*	US	2,117,310
dollar* NEAR US	US	249,699
DEM	DE	303,174
DM	DE	939,411
Deutschmark	DE	1,826
CHF	CH	113,328
SFr.	CH	36,464
SFr	CH	22,637
Sfrs	CH	57
ATS	AT	164,887
schilling*	AT	71,096
schilling* NEAR Austria*	AT	9,735
GBP	GB	105,735
pound*	GB	2,344,628
British NEAR pound*	GB	54,757
£	GB	syntax error

Table 4: Frequency of different currency representations [as of 2001-08-06]

Usually a Web page shows the currency information in a static way. If the currency in which price information is shown is not the reference currency of the user, it is generally assumed that he will convert the price into his own currency. A personalization concept could allow the user to specify the currency in which he would like to see the price-information in the future. In this case, the server of the information provider would have to use the current exchange rates to compute the price to be offered to a foreign customer. The database would contain static data about the names or abbreviations of the currencies and dynamic data about the varying exchange rates. The exchange rates must be documented for several time intervals. In this case monetary and temporal data are logically connected in a database. The information provider must decide whether he wants to represent the currency in the locally accepted way or in the form proposed by the ISO standard. Again, the locally usual representation may be accompanied by the standard notation in parentheses, resulting in a bipolar representation of currency notations.

Altavista does not offer to distinguish several characters used for decimal points or structuring many-digit numbers. In scanning for monetary data this ambiguity also provides problems.

5 Conclusions

If only a small number of professional people decides whether they obey a standard or not, one could hope that it is comparatively easy to address this group and to convince them that the benefits of following the standards are larger than the costs associated with using them. However, in many areas of professional and private life people do not care too much about existing standards and on using them.

The quality of data on the web is often perceived as worse than in traditional media. Significant deficiencies have been found with respect to access security, representational consistency, objectivity, accuracy, and concise representation [Klein (2001)]. Thus, presenting temporal data in a more standardized way would contribute to the representational consistency, accuracy, and concise representation of data and would improve the appreciation of the web as a reliable source of information.

Standards are available but many authors and webmasters do not care. A first step toward a broader use of standards for representing temporal and currency information on the Web could be to make people more familiar with the existing standards by showing these data in the locally preferred as well as in the internationally standardized representation.

References

Axiom: Draft Native ASN.1 Representation of a UN/EDIFACT 92.1 PAYORD Message. http://www.axiomservices.freeserve.co.uk/ewfbaa/ep921ia/ep921ia.htm [1998-12-12; as of 2001-08-25]

Fielding, R., et al.: Hypertext Transfer Protocol -- HTTP/1.1 (1999). http://www.ietf.org/rfc/rfc2616.txt [2000-12-27; as of 2001-08-25]

ISO 8601: Data elements and interchange formats – Information interchange – Representation of dates and times. http://www.mcs.vuw.ac.nz/technical/software/SGML/doc/iso8601/ISO8601.html [1995-10-12; as of 2001-08-06] http://lists.ebxml.org/archives/ebxml-core/200104/pdf00005.pdf [2000-12-19; as of 2001-08-16]

Klein, B.D.: User Perceptions of Data Quality: Internet and Traditional Text Sources, in: Journal of Computer Information Systems 41 (2001) 4, pp. 9 - 15.

Knolmayer, G., Buchberger, T.: Maintaining temporal integrity of World Wide Web pages. In: S. Jajodia, W. List, G. McGregor, L. Strous (Eds.): Integrity Control in Information Systems, Volume 1: Increasing the confidence in information systems, London et al., Chapman&Hall, 1997, pp. 43 - 63.

Knolmayer, G., Myrach, T.: Y2K: Much ado about nothing? In: Informatik-Spektrum 23 (2000) 2, pp. 131 - 137.

Knolmayer, G.F., Myrach, T.: Concepts of Bitemporal Database Theory and the Evolution of Web Documents. In: R.H. Sprague (Ed.): Proceedings of the 34th Hawaii International Conference on Systems Sciences (HICSS 2001). IEEE Computer Society, Los Alamitos 2001 (CD-ROM)

Kuhn, M.: A Summary of the International Standard Date and Time Notation. http://www.cl.cam.ac.uk/~mgk25/iso-time.html [2001-08-01; as of 2001-08-25]

Lamp, J.: Citation Styles for Electronic Media. http://lamp.infosys.deakin.edu.au/citation.html [2001-08-13; as of 2001-08-25]

Myrach, T.: Temporale Datenbanken in betrieblichen Informationssystemen: Prinzipien, Konzepte, Umsetzung. Habilitation Thesis at the University of Bern 2000.

The Web Standards Project: What are Web standards and why should I use them? (no date). http://www.webstandards.org/edu_faq_n4.html [as of 2001-08-25]

Timechange Creativity Services: The Current Calendar system (no date). http://www.timechange.com/3m/ [as of 2001-08-25]

Wolf, M., Wicksteed, C. (1997): Date and Time Formats. http://www.w3.org/TR/NOTE-datetime [2000-03-13; as of 2001-08-06]

XE.com: ISO 4217 Type Currency Symbol List - About the Currency Symbols. http://www.xe.com/iso4217.htm [2001-07-10; as of 2001-08-06]

y2kmistakes.com: The Big List. http://y2kmistakes.com/biglist.htm [2000-04-15; as of 2001-08-06]