


```

    <rdf:li>Places</rdf:li>
    <rdf:li>Germany</rdf:li>
    <rdf:li>Baden-W&#xFC;rttemberg</rdf:li>
    <rdf:li>Umlaut&#xDF; &#xFC; &#xF6; &#xE4; &#xDC; &#xD6; &#xC4;</rdf:li>
  </rdf:Bag>
</dc:subject>
<lr:hierarchicalSubject>
  <rdf:Bag>
    <rdf:li>Places</rdf:li>
    <rdf:li>Places|Germany</rdf:li>
    <rdf:li>Places|Germany|Baden-W&#xFC;rttemberg</rdf:li>
    <rdf:li>Places|Germany|Umlaut&#xDF; &#xFC; &#xF6; &#xE4; &#xDC; &#xD6; &#xC4;</rdf:li>
  </rdf:Bag>
</lr:hierarchicalSubject>
</rdf:Description>
</rdf:RDF>
</x:xmpmeta>
<?xpacket end="w"?>
659 rmills@rmillsmm:~/gnu/github/clanmills/exiv2 $

```

When the data is output to the terminal (without the filter: **xmllint --format -**), Windows and the Mac do not display bytes 128-255 with the same glyphs. However, that is the different behaviour by the Mac's Terminal and Window's cmd.exe console.

To my knowledge, metadata strings (in IPTC and Exif) is stored as count + binary sequence of bytes. The standards do not say that it has to be stored as UTF-8, or anything else. On the exiv2 command-line, there is a switch `-n (--encoding)` to define the Charset for User comments.

One matter that could be investigated is to modify the Windows samples to operate in UNICODE. Be default, we build the Exiv2 library without enabling `EXV_UNICODE_PATH` and therefore only offer ascii interfaces into the file-system. When `EXV_UNICODE_PATH` is enabled, the only sample which takes advantage of this is `exifprint`. A useful project could be to:

- 1) Always build msvc libraries with `EXV_UNICODE_PATH` enabled.
- 2) Convert all sample applications to use `wmain()` for msvc builds.
- 3) Respect `-n --encoding` option to tell the samples how to convert wstring to binary.
- 4) Study what should be done for Cygwin and `msgs/2.0`

This could be a good project for Google Summer of Code. If we recruit a Chinese or Indian student, they would have the language to effectively test this.

Files

metadata-test.jpg	70.6 KB	06 Sep 2017	Robin Mills
-------------------	---------	-------------	-------------