
Subject: coding of dc:language

Posted by on Fri, 30 Mar 2012 17:53:30 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hallo,

I have added some examples on using Dublin Core Metadata Set to the Wiki pages. There was one response from Susanne concerning the item <dc:language>.

Generally, this item shall be used to code character set of the names (e.. g. station names a. s. o.) of the RailML file. This value is of importance in case the containing Unicode names have to be converted into a non-Unicode-string by the reading software.

Originally, I wanted it to contain the Codepage Number of the data
<dc:language>1252</dc:language> ;(1252=ANSI - Lateinisch I)
because, from my experience, one does need a Codepage Number do convert non-Unicode strings.

This did not enjoy Susanne who rather would prefer a coding like ISO 15924:
<dc:language>de-CH</dc:language>

The problem is that there is no 'conversion table' or something like that (as far as I know) to convert Codepage Numbers into ISO 15924 codes or vice versa. There is, unfortunately, no standardisation of Codepages at all. So if we do not allow the non-standardised Codepage Numbers we cannot tell the reading software how to convert the UTF-8 strings of a RailML file into non-Unicode strings. This leaves a reading software with the need to 'scan' the names for special characters and deduce a Codepage from this - a more empiric solution.

The problem with the ISO 15924 codes is not only that there is no 'conversion table'. It is also that typically a RailML file contains names of more than one language, e. g. some foreign station names also. This is normally no problem because one Codepage normally allows languages of neighboring countries. Our 'middle-European' Codepage (1250) allows German

what should we write into <dc:language> if a RailML file contains all of these three and the writing programme only know that it is CodePage 1250?

Anyway, it is not a big problem because it only applies to non-Unicode software and there should be not much non-Unicode software nowadays. It's only that we do not know...

So, from my opinion we have two possible solutions:

a) either to skip this <dc:language> at all and delete it from all examples

b) or still to allow and recommend a Codepage number (!) there because it costs nothing, may help someone, and there is no other need for this element.

It does not make sense to code it with ISO 15924 since, as I did explain, there is normally not _one_ source language for all the RailML file.

@Susanne: If nobody answers this 'post' in a near future you can tell me at any time to delete this <dc:language> from the examples without further objection from me. I leave it up to you, doing nothing more from my side..

With best regards,
Dirk.
