ITS categories transfer in CMS <-> LSP scenario

[This blurb is intended as a germ of a best practice document that could be produced by the WG at a later stage]

# Definitions [these will need to be merged with defs from the requirements grandfathered document]

## CAT

Computer Aided Translation. Tooling making use of Translation Memory.

## LSP

stands for a Language Service Provider, a company to which Localization Services are being outsourced by corporations and SMBs. In some cases LSPs can be internal corporate departments, such as Oracle WPTG (World Wide Product Translation Group).

## Localization Buyer

Corporations, small or medium businesses that need to make their content or products multilingual and work in other than their home markets.

# Status Quo:

Localization Buyers store HTML fragments in CDATA sections of XML documents. This practice is common but far from being commendable, as the CDATA sections are out of scope of any rules set in the carrier XML document.

LSPs are used to coping with this bad practice and they normally have cascaded parsing mechanisms that can handle the CDATA sections in an intelligible way. However the CDATA can be just anything and so the welformedness issues are dumped onto the LSP and LSPs stab on parsing the CDATA as HTML or any other syntax is just a stab into the dark and backfires every now and then.

# Best Practice

The options in case you want to transfer useful metadata onto your localization service provider are the following.

1. Send valid ITS 2.0 decorated HTML 5 with an external XML rules file.
   1. This is a valid and conformant way, and all localizable content is in scope of the externally provided rules. Thre is however the risk of separation of the rules file that would make the its- prefixed mark up within HTML useless in most cases.
2. Use XLIFF with ITS 2.0 mapping [XLIFF 1.2 is available, XLIFF 2.0 mapping to be finalized within 2013]
   1. This is a clean and conformant solution, but may no tbe feasible if you do not have localizable content extraction know how or if your target CAT tool does not support XLIFF
3. Use XHTML 5 serialization that allows use of XML based ITS scoping mechanisms in the same file as the content payload.

## Technical caveats

TBD

[See current discussion on use of Tidy in PHP etc.]