



# Lessons Learned from Applying ITS 2.0 onto Web Platform

Jirka Kosek (UEP)

Dave Lewis (TCD)

Felix Sasaki (DFKI)



# What is ITS 2.0?



- Meta-data annotations addressing gaps in I18n and L10n interoperability
  - CMS, L10n Tools, Language Technology
- Defines 18 separate Data Categories
- Implementation-neutral specification of data categories
- Implementations in HTML5, XML and RDF

# ITS 2.0 Data Categories

## ITS1.0

- Translate
- Localization Note
- Terminology
- Directionality
- Lang info
- Element within text

## I18n

- Locale Filter
- External Resource
- Preserve Space
- Allowed Characters
- Storage Size
- ID Value

## Language Technology

- Domain
- MT confidence
- Text Analysis

## Provenance & QA

- Quality Issue
- Quality Rating
- Provenance

# ITS 2.0 examples

- Namespaced attributes are used for capturing ITS metadata in XML documents

```
<para>See you in <phrase its:term="yes">Berlin</phrase>.</para>
```

```
<para>But there are other cities to go if you like <phrase  
its:translate="no">foie gras</phrase>.</para>
```

- In HTML5 its-\* prefixed attributes are used

```
<p>See you in <span its-term="yes">Berlin</span>.</p>
```

# Problems we faced in Web Platform

- CSS selectors can not address attributes and thus can not replace XPath even for quite simple and common scenarios (e.g. title attribute)
- HTML5 extensibility is very limited and implementing its-\* prefixed solution is more expensive than using namespaces (e.g. updating validator)
- HTML5 does not have feature for embedding XML islands with additional (meta)data

# What is missing for smooth workflow

- Majority of websites is managed in CMS
- Easy way for importing/exporting content together with ITS annotations, translation memory, terminological vocabulary, ... is needed for localization interoperability
- CMIS (Content Management Interoperability Services) profile could be solution