

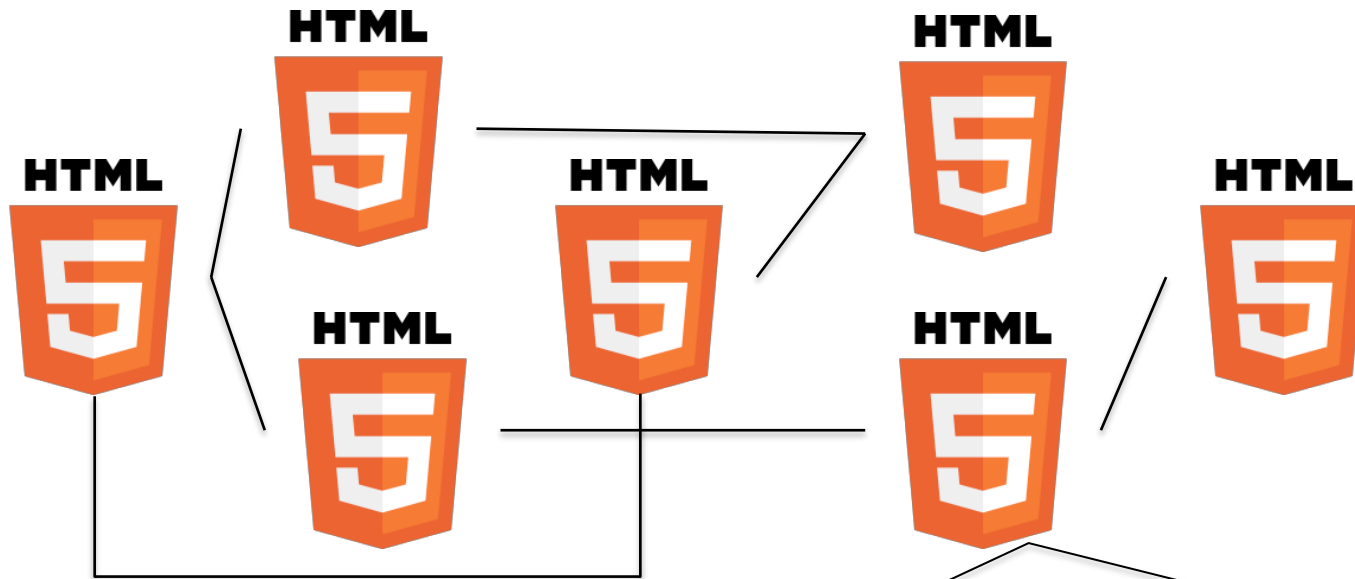
New developments in Semantic Web and Internationalization at the W3C

Felix Sasaki

DFKI / W3C Fellow

WHAT IS SEMANTIC WEB? INTRODUCTION IN 3 MIN 45 SEC ...

Building blocks of the Web



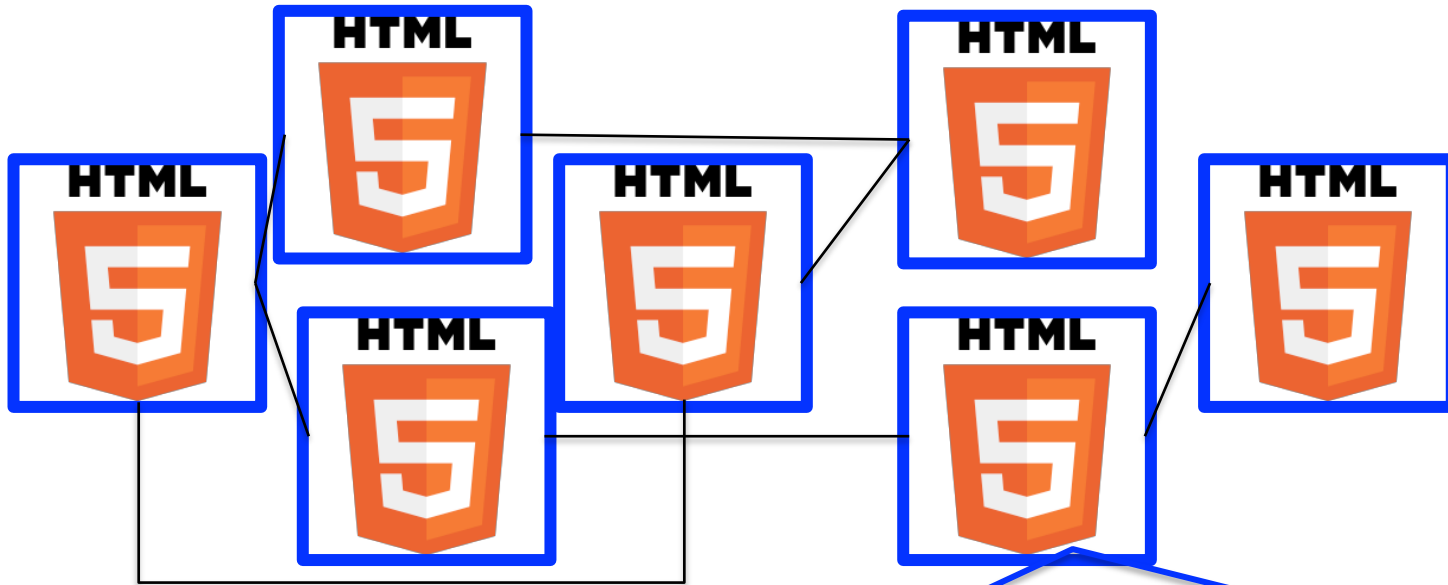
`<p>All content on this site is licensed under`

`<a`

`href="http://creativecommons.org/licenses/by/3.0/">`

`a Creative Commons License. </p>`

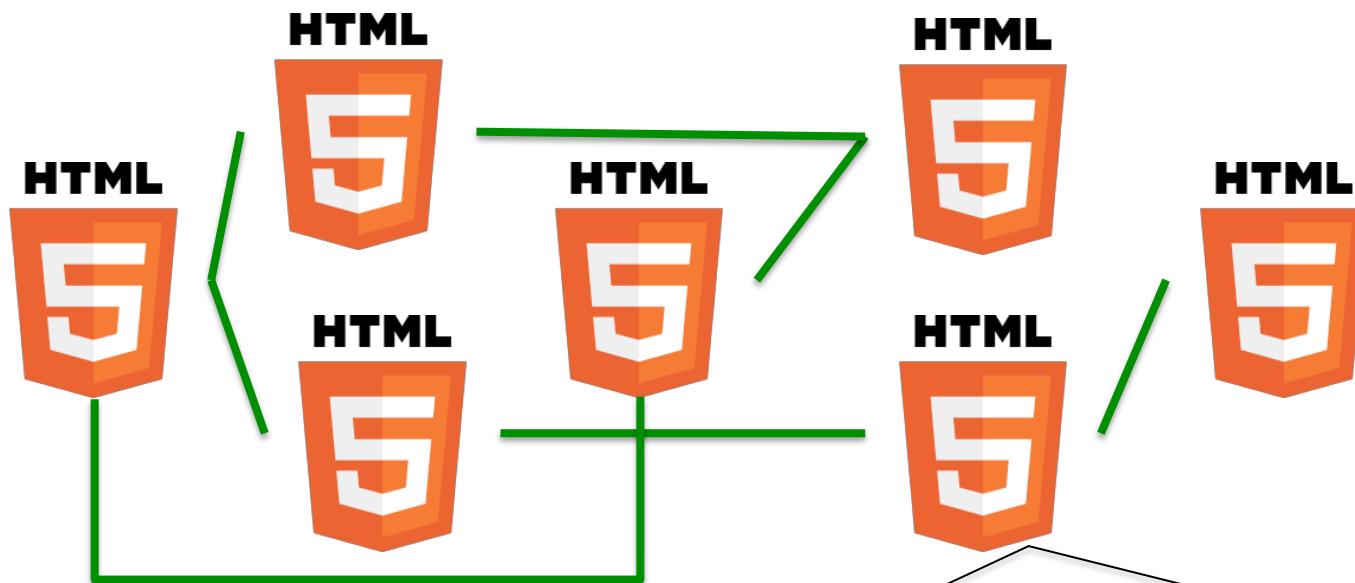
Content



<p>All content on this site is licensed under

a Creative Commons License. </p>

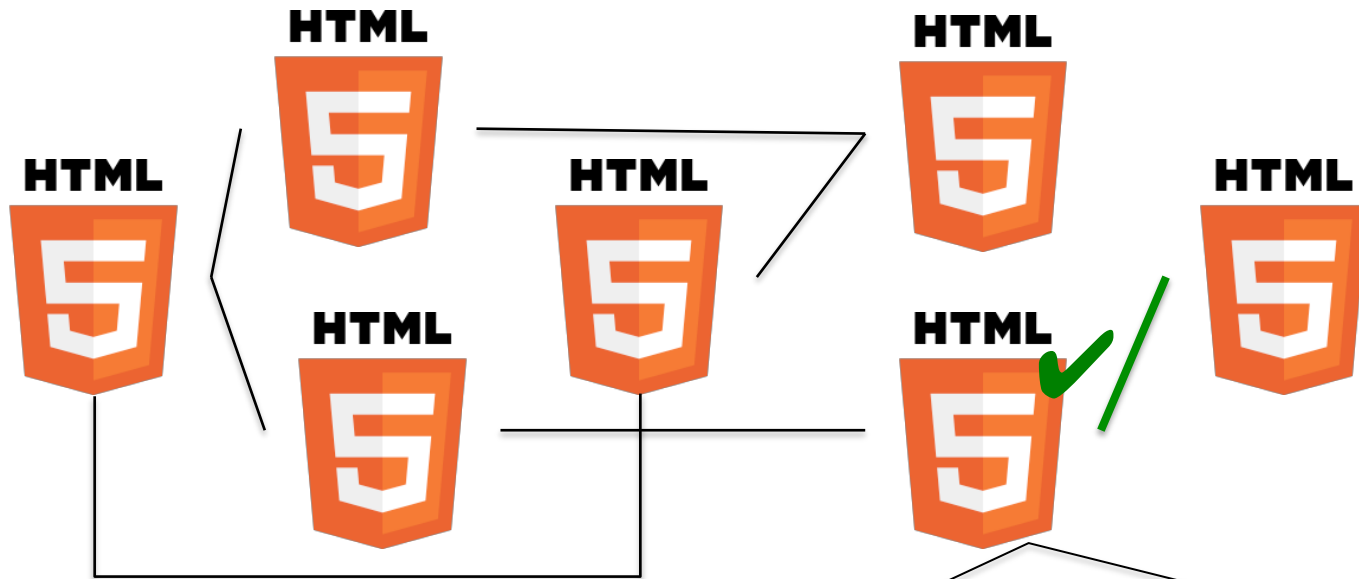
Links (or “identifiers”)



<p>All content on this site is licensed under

a Creative Commons License. </p>

Easy: “Find all content that links to <http://creativecommons.org/licenses/by/3.0/>”



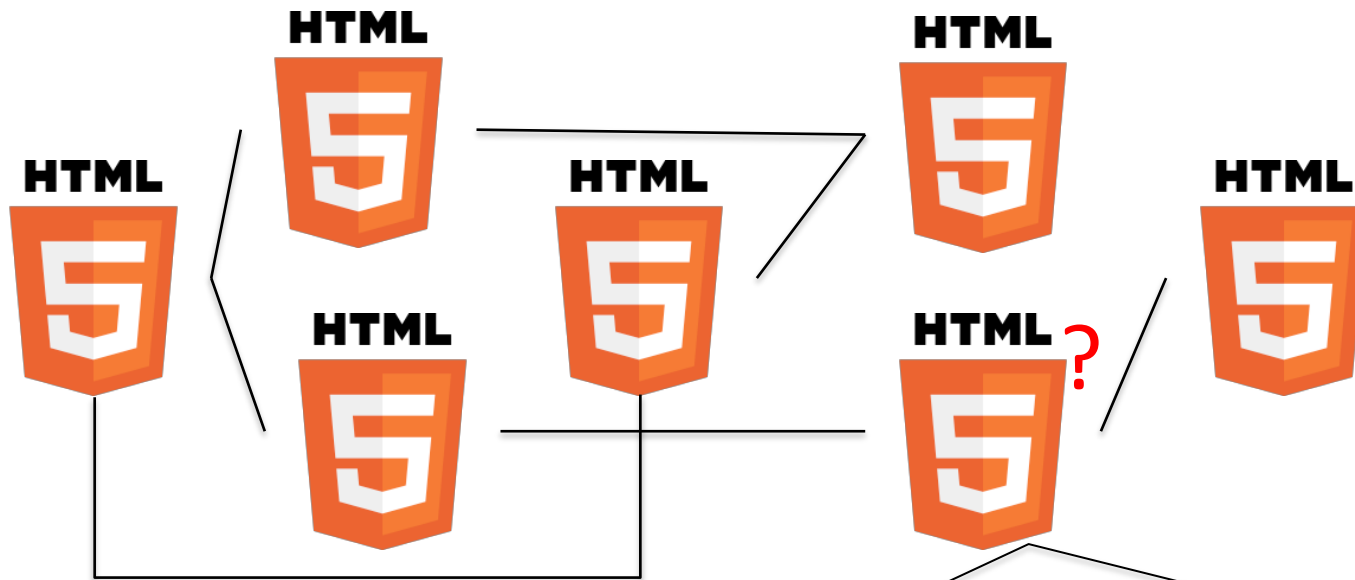
```
<p>All content on this site is licensed under
```

```
<a
```

```
href="http://creativecommons.org/licenses/by/3.0/">
```

```
  a Creative Commons License</a>. </p>
```

Still **difficult**: “Find all content that links to a creative commons license”



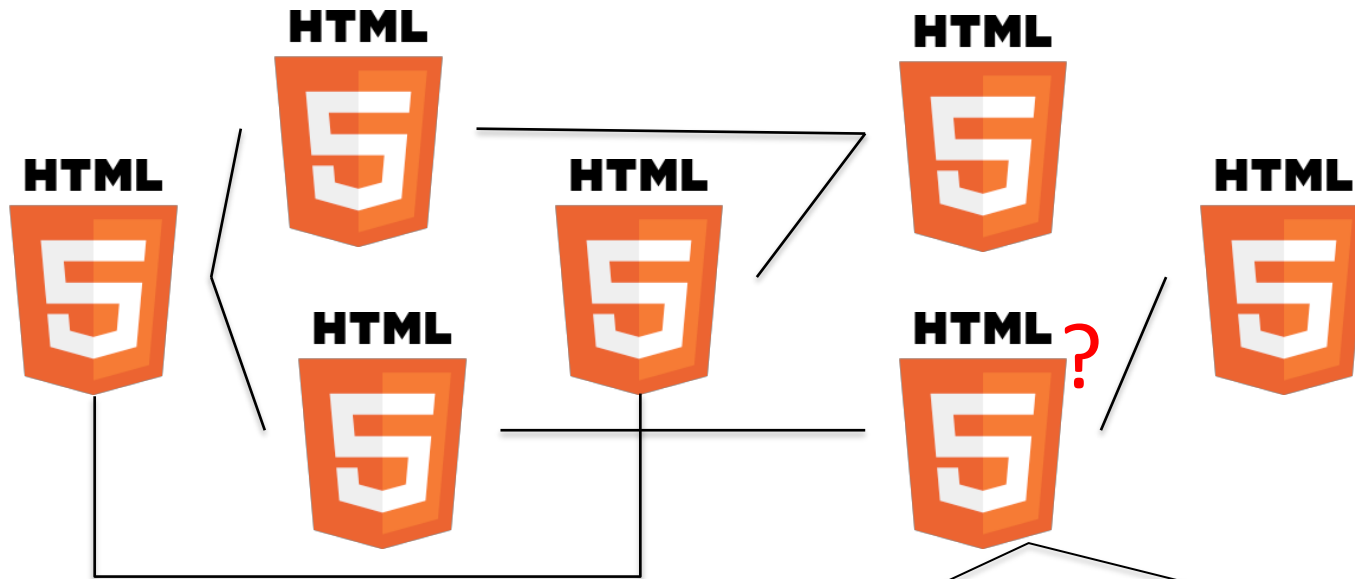
<p>All content on this site is licensed under

<a

href="<http://creativecommons.org/licenses/by/3.0/>">

a Creative Commons License. </p>

Semantic Web to the rescue = Providing machine readable information on the Web

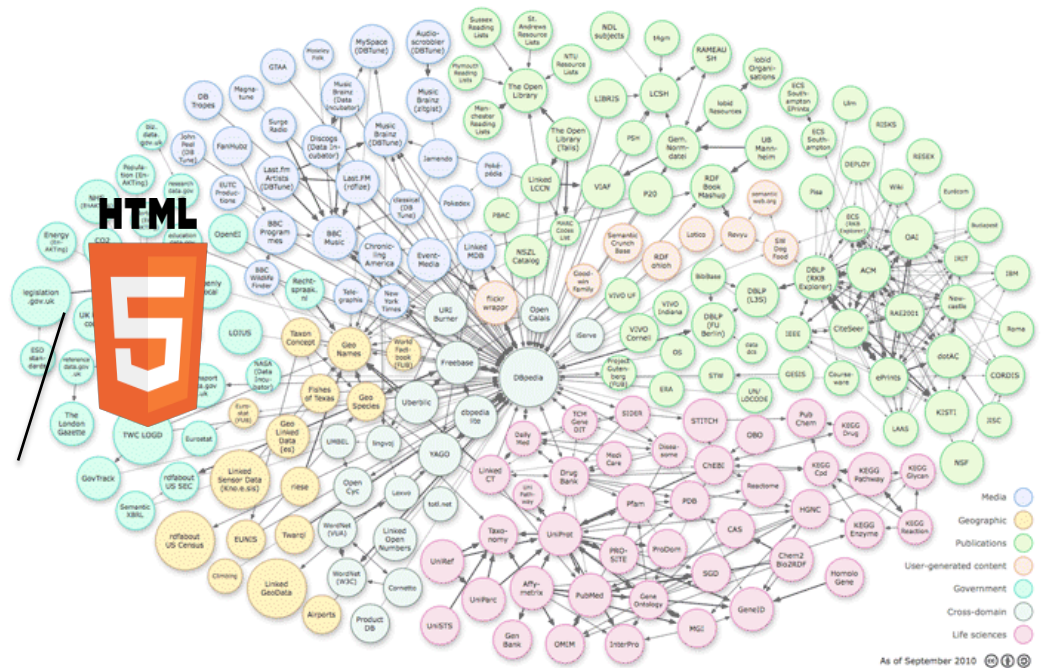
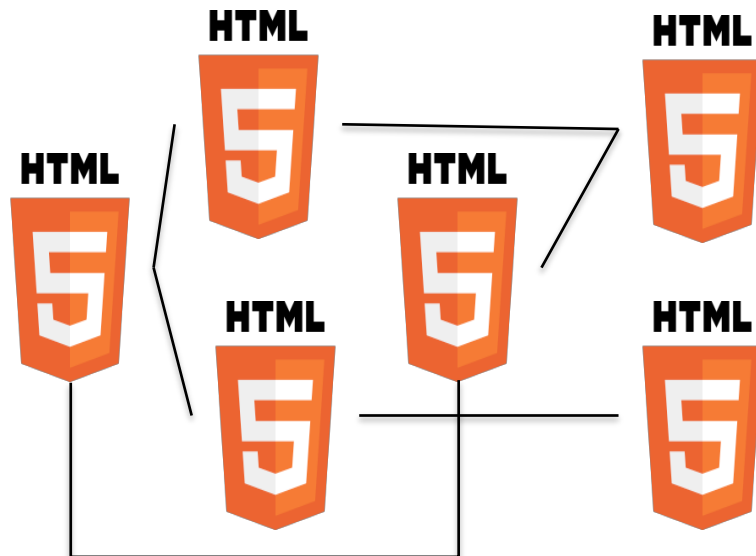


```
<p>All content on this site is licensed under  
<a property="http://creativecommons.org/ns#license"  
href="http://creativecommons.org/licenses/by/3.0/">  
a Creative Commons License</a>. </p>
```


Semantic Web = Providing machine readable information on the Web

Web of documents

Web of data





TASKS & TECHNOLOGIES (CURRENT + UPCOMING)

Tasks ...

Write Semantic Web data

Create Semantic Web vocabularies

Representation and data creation

Query

... and technologies: past and presence

Write Semantic Web data: RDF (Resource Description Framework)

Create Semantic Web vocabularies: RDFS, SKOS, OWL (for complex ontologies)

Representation and data creation: Turtle, RDFa, R2RML, ...

Query: SPARQL

Write Semantic Web Data: RDF “statements”



Reference to license CC0 with an RDF statement (visualization):

<http://www.w3.org/Talks/2012/1124-SW-MLW-Moscow/>

<http://creativecommons.org/ns#license>



<http://creativecommons.org/licenses/by/3.0/>

Turtle syntax

Reference to license CC0 with an RDF statement (visualization) + Turtle syntax:

<http://www.w3.org/Talks/2012/1124-SW-MLW-Moscow/>

<http://creativecommons.org/ns#license>

<http://creativecommons.org/licenses/by/3.0/>

```
@prefix cc: <http://creativecommons.org/ns#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
<http://www.w3.org/Talks/2012/1124-SW-MLW-Moscow/>
  cc:license
    <http://creativecommons.org/licenses/by/3.0/>.
```

RDF statements ...

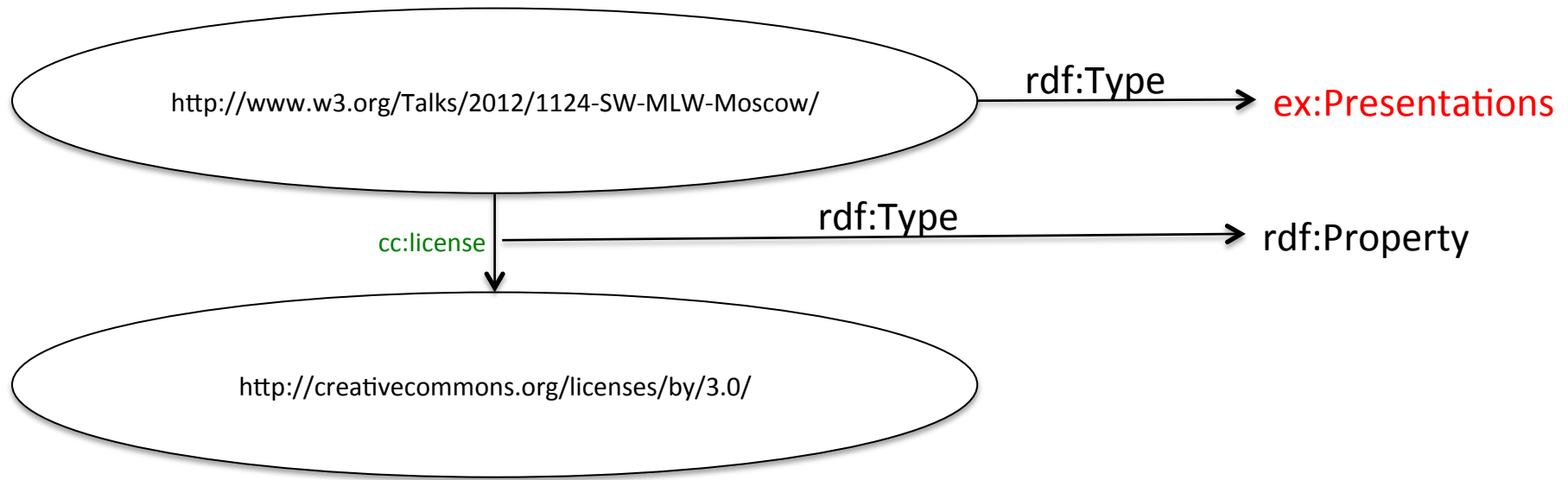
<http://www.w3.org/Talks/2012/1124-SW-MLW-Moscow/>

<http://creativecommons.org/ns#license>



<http://creativecommons.org/licenses/by/3.0/>

... can be based on vocabularies



RDF Schema

- For defining **Classes** (example “Presentations”) and **properties** (like “cc:license”)

OWL (Web Ontology Language)

- For defining further constraints for vocabularies

SKOS

- For describing e.g. thesauri, taxonomies, classification schemes

Current activities of Semantic Web vocabulary work in W3C

- Work on actual vocabularies, e.g. in the area of eGovernment
 - “An Organization Ontology”
 - <http://www.w3.org/TR/vocab-org/>
 - “RDF Data Cube Vocabulary”
 - <http://www.w3.org/TR/vocab-data-cube/>
 - “Data Catalog Vocabulary (DCAT)”
 - <http://www.w3.org/TR/vocab-dcat/>
- Discussing how to actually use open data
 - “Using Open Data” workshop June 2012
<http://www.w3.org/2012/06/pmod/>
- *Join the discussion* in the eGov Interest Group
http://www.w3.org/egov/wiki/Main_Page !

Ways to create and store Semantic Web data



```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cc="http://creativecommons.org/ns#">
  <rdf:Description
    rdf:about="http://www.w3.org/Talks/2012/1124-SW-MLW-Moscow/">
```

```
  @prefix cc: <http://creativecommons.org/ns#>.
  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
  <http://www.w3.org/Talks/2012/1124-SW-MLW-Moscow/
    cc:license
    <http://creativecommons.org/licenses/by/3.0/>.
```

RDF/XML

Turtle

```
<p>All content on this site is licensed under
<a property="http://creativecommons.org/ns#license"
  href="http://creativecommons.org/licenses/by/3.0/">
  a Creative Commons License</a>. </p>
```

RDFa (embedding in HTML)

- RDFa Lite 1.1 <http://www.w3.org/TR/rdfa-lite/>: one-to-one mapping to microdata
- Microdata to RDF <http://www.w3.org/TR/microdata-rdf/>: extracting RDF from microdata
- HTML Data Guide <http://www.w3.org/TR/html-data-guide/>: guidance about RDFa vs. microdata vs. microformats
- R2ML <http://www.w3.org/TR/r2rml/>: Mapping relational data bases to RDF

New work: Linked Data Platform

- See <http://www.w3.org/2012/ldp/>
- HTTP patterns for read & write of linked data
 - Retrieve / create / delete / update / aggregate / ...
linked data resources
 - Using HTTP GET / PUT / DELETE / ...
- Lot's of user scenarios
 - Maintaining social contact information, software development tool integration, library linked data, data catalogs, ...

Query - SPARQL

- Query language for RDF
- Patterns in link (=graph) structures
- E.g. “find all presentations with CC0 license”

Query will return

<http://www.w3.org/Talks/2012/1124-SW-MLW-Moscow/>

```
PREFIX cc: <http://creativecommons.org/ns#>
SELECT ?presentation WHERE {
?presentation cc:license <http://creativecommons.org/licenses/by/3.0/>.
}
```

SPARQL Query with dbpedia data set: "People who where born in Moscow before 1900":

<http://tinyurl.com/dbpedia-example>

SPARQL Explorer for <http://dbpedia.org/sparql>

SPARQL:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dbo: <http://dbpedia.org/ontology/>

SELECT ?name ?birth ?death ?person WHERE {
  ?person dbo:birthPlace :Moscow .
  ?person dbo:birthDate ?birth .
  ?person foaf:name ?name .
  ?person dbo:deathDate ?death .
  FILTER (?birth < "1900-01-01"^^xsd:date) .
}
ORDER BY ?name
```

Results:

SPARQL results:

name	birth	death	person
"Aleksandr Danilovich Menshikov"@en	"1673-11-06"^^xsd:date	"1729-11-12"^^xsd:date	:Aleksandr_Danilovich_Menshikov ↗
"Aleksandr Golovin"@en	"1863-03-01"^^xsd:date	"1930-04-17"^^xsd:date	:Aleksandr_Golovin_(artist) ↗
"Aleksandr Nekrasov"@en	"1883-12-09"^^xsd:date	"1957-05-21"^^xsd:date	:Aleksandr_Nekrasov ↗
"Aleksandr Yakovlevich Golovin"@en	"1863-03-01"^^xsd:date	"1930-04-17"^^xsd:date	:Aleksandr_Golovin_(artist) ↗
"Alexander Alexandrovich Alekhine"@en	"1892-10-31"^^xsd:date	"1946-03-24"^^xsd:date	:Alexander_Alekhine ↗
"Alexander Dmitriyevich Kastalsky"@en	"1856-11-28"^^xsd:date	"1926-12-17"^^xsd:date	:Alexander_Dmitriyevich_Kastalsky ↗

MULTILINGUAL WEB

And the Semantic Web?

“4.78% human readable descriptions have one language tag”

“Only 0.7% datasets contain several language tags”

“Most commonly language used:

44.72% (en), 5.22% (de), 5.11% (fr), 3.96% (it),...”

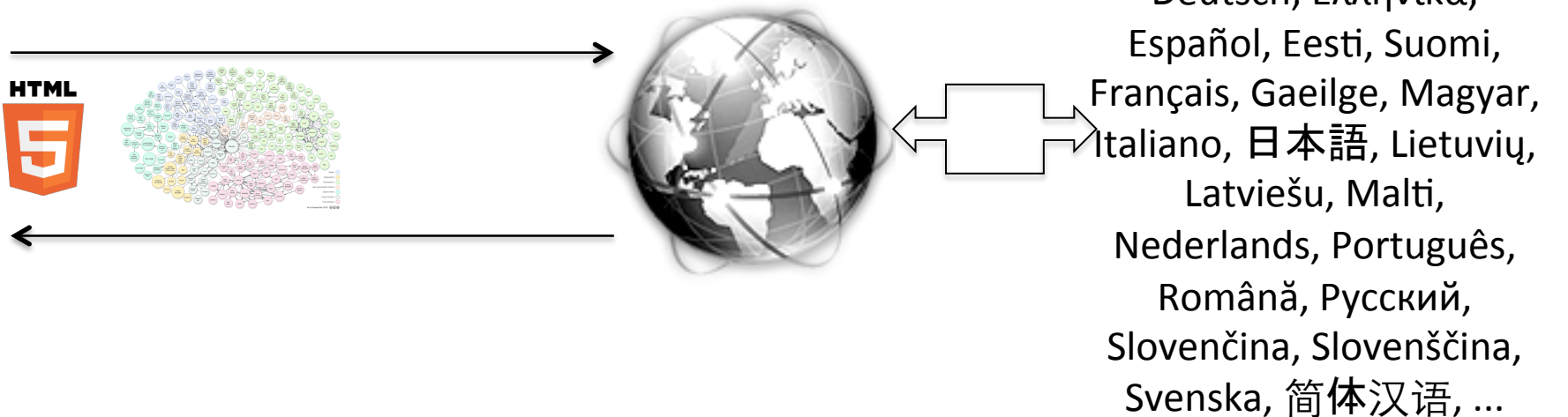
Source: Ell et al, 2011 Labels in the Web of Data, ISWC 2011

Getting multilingual: a process ...

Create
content & data

Organize
localization

Translate &
adapt content

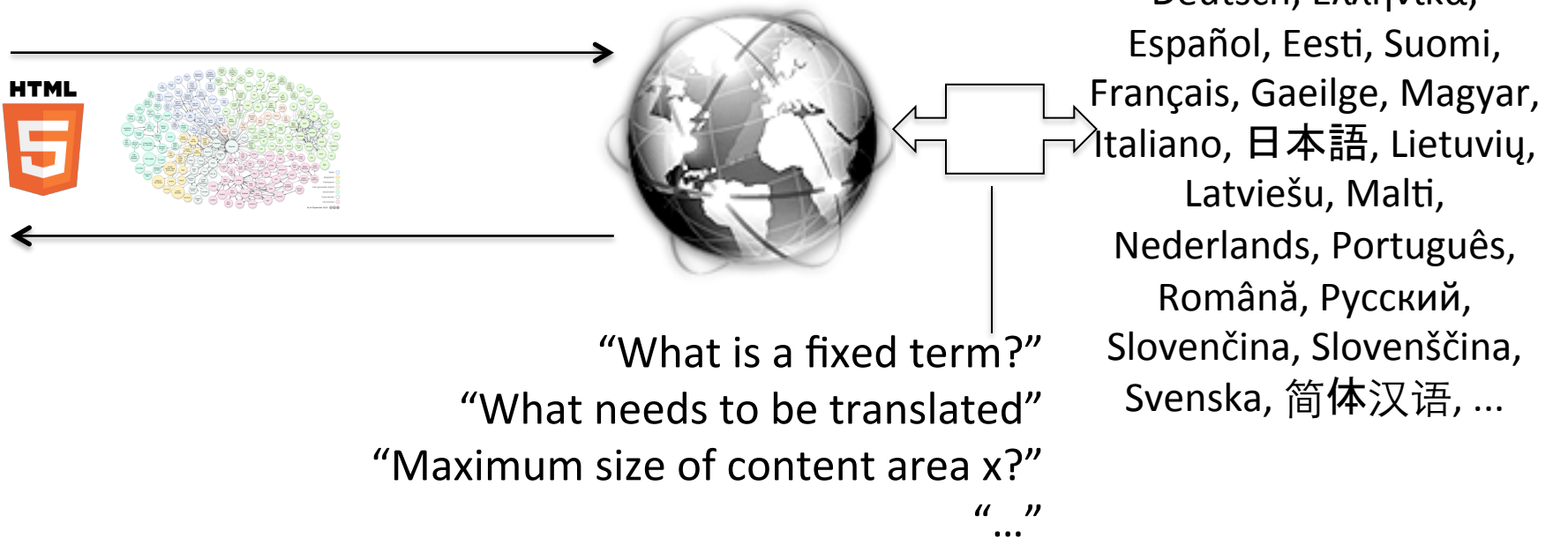


... with lots of questions coming up

Create
content & data

Organize
localization

Translate &
adapt content

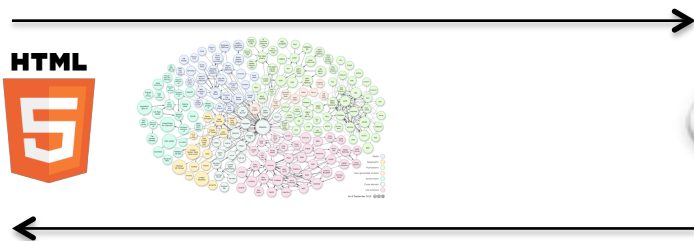


Metadata to the rescue

Create
content & data

Organize
localization

Translate &
adapt content



Answers
via
ITS 2.0
Metadata

“What is a fixed term?”
“What needs to be translated?”
“Maximum size of content area x?”
“ ... ”



Internationalization Tag Set 2.0

- Defining metadata for multilingual processing of Web or other content, e.g. via
 - Machine translation
 - Localization workflows
- Where is the metadata needed – *for example*:
 - In Web content, e.g. HTML5
 - In XML-based formats like DocBook etc.
 - In Localization related formats like XLIFF (XML Localization Interchange File Format)
 - In the Semantic Web

Metadata example: “Translate” in HTML5 and XLIFF

```
<!DOCTYPE html><html> ...  
<span xml:lang="en">Presentation about  
<span translate="no">Internationalization</span> and Semantic  
Web in Moscow, 24 November 2012.</span> ... </html>
```

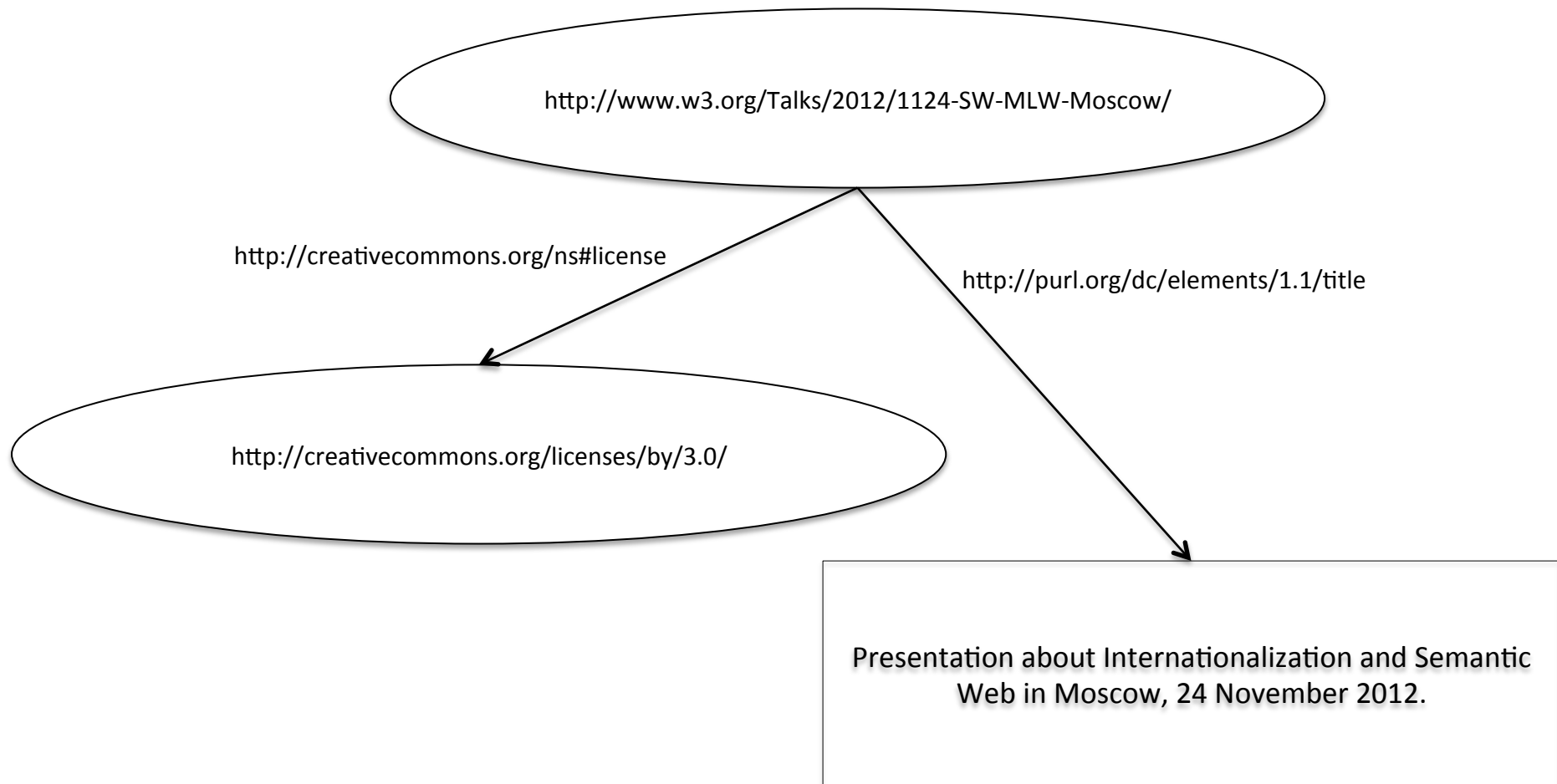
```
<xliff ...> ...  
  <trans-unit id="1">  
    <source xml:lang="en">Presentation about  
    <mrk mtype="protected">Internationalization</mrk> ...</source>  
    <target> ...  
  </xliiff>
```

Metadata example: “Terminology” in HTML5 and XLIFF

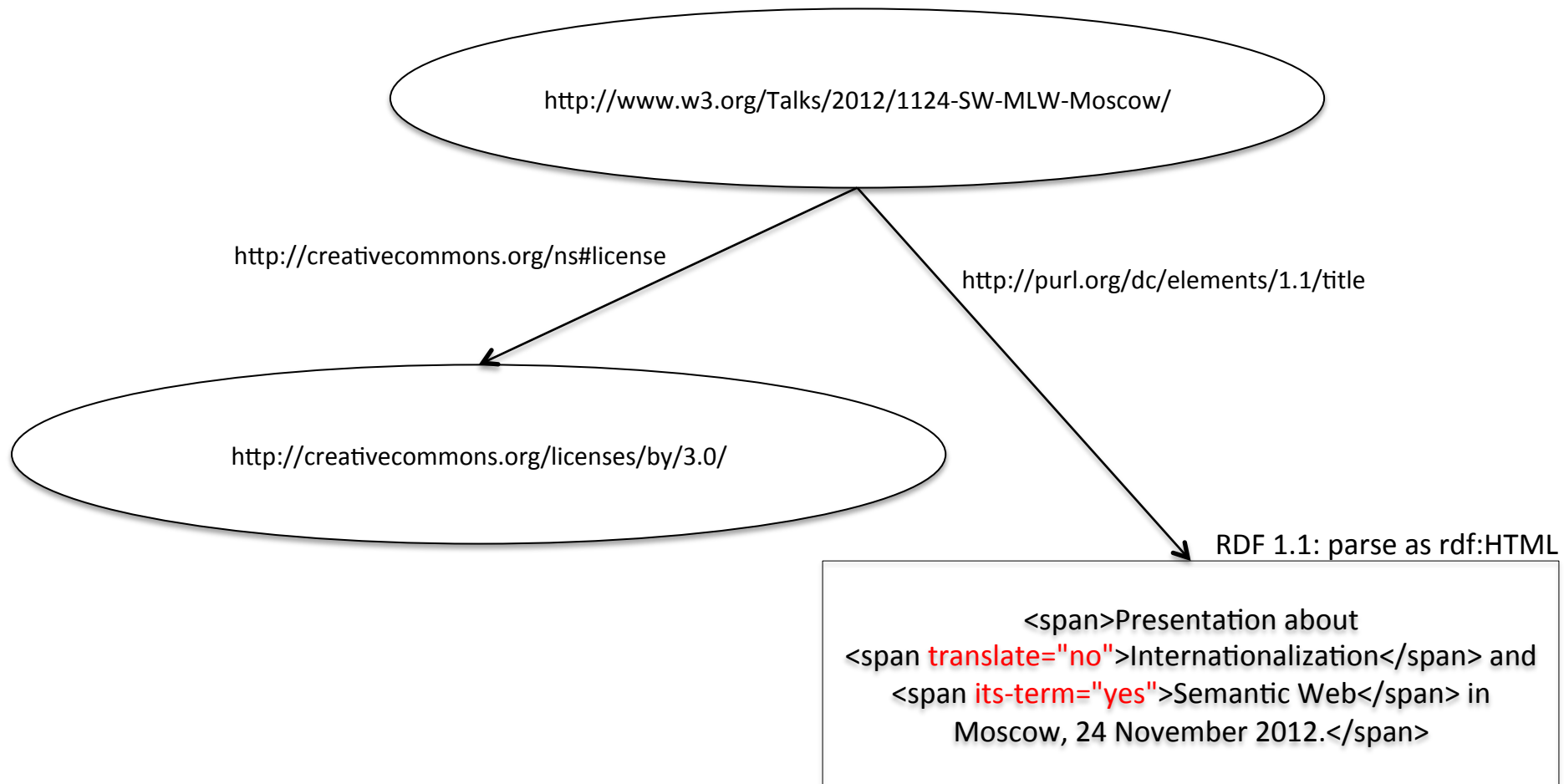
```
<!DOCTYPE html><html> ...  
<span xml:lang="en">Presentation about Internationalization and  
<span its-term="yes">Semantic Web</span> in Moscow, 24  
November 2012.</span> ... </html>
```

```
<xliff ...> ...  
  <trans-unit id="1">  
    <source xml:lang="en">Presentation about ... and  
    <mrk mtype="term">Semantic Web</mrk> ...</source>  
    <target> ...  
  </xliff>
```

ITS 2.0: Metadata for localizing the Semantic Web



ITS 2.0: Metadata for localizing the Semantic Web



Most important: tooling

- No details here (time is out)
- More info at <http://tinyurl.com/its2-uc-impl>
- Covers
 - Content creators: Drupal module for creating metadata <http://tinyurl.com/its2-drupal>
 - Automatic annotation via <http://tinyurl.com/its2-named-entities>
 - Validation via <http://tinyurl.com/its2-validation>
 - XLIFF <http://tinyurl.com/its2-generate-xliff>
 - Much more!

New developments in Semantic Web and Internationalization at the W3C

Felix Sasaki

DFKI / W3C Fellow

Contact: fsasaki@w3.org