
# MLW-LT and Representation Formats: Suggestions.

**Maxime Lefrançois**
**Maxime.Lefrancois@inria.fr**

**Inria – wimmics**
**inria.fr – wimmics.inria.fr**

**The Multilingual Web – Linked Open Data and MultilingualWeb-LT Requirements, 11 - 13 June 2012, Dublin**


# Ph.D. student
# Explanatory Combinatorial Lexicology[1] and the Semantic Web

Maxime Lefrançois
Maxime.Lefrancois@inria.fr

Inria – wimmics
inria.fr – wimmics.inria.fr

[1]: http://olst.ling.umontreal.ca/pdf/ECD.pdf

# Outline

1. “Dropping RDFa as a requirement”
2. CURIEs
3. Provenance – XG
4. HTML: local vs. global ITS annotations ?
5. Publication of schemas and vocabularies for ITS 2.0

”Dropping RDFa as a requirement ?”

# 1. RELATED ISSUE-18

answer: NO, it’s in the charter

https://www.w3.org/International/multilingualweb/lt/track/issues/18

# ITS and RDF - RDFa

**Core issues:**

ITS and RDF seem conceptually incompatible,

ITS 1.0: one annotates *à-priori* fragments of text

in RDF literals can't be subject of a triple

***Different conceptualizations !***

**Suggestion for ITS 2.0**

by Sebastian Hellmann:
**use the NIF String ontology[1] elements for mapping ITS 2.0 Data Models to RDF**

[1]: http://nlp2rdf.lod2.eu

# ITS and RDF - RDFa

## The `str:String` Class - NIF receipes

### For any text file (HTML -> source code)

**Offset-based URIs**

doc.html#*offset_14406_14418_Semantic%20Web*

**Context-Hash-based URIs**

doc.html#*hash_4_12_79edde636fac847c006605f82d4c5c4d_Semantic%20Web*

### For XML documents

**XPointer based URIs**

*In the future NIF 2.0 ?*

**example:** **Dublin** is a great city:

doc.html#*xpointer(string-range(id("myId"), "",1,7)[1]* -> this « Dublin » string in doc.html

doc.html#*xpointer(string-range(//, "Dublin",1,7)* -> every « Dublin » string in doc.html

# ITS and RDF - RDFa

| | NIF Recipe: offset http://ex.org#offset_... | NIF Recipe: hash http://ex.org#hash_... | http://ex.org r http://ex.org#myId | CSS selector 1/2/3 | XPath 1.0 | XPath 2.0 | XPointer 1.0 ex.org#xpointer(...) |
|---|---|---|---|---|---|---|---|
| Ranges in HTML source | one ~ | ? | | | | | |
| Elements | / | / | one ~ | list of ~ | list of ~ | list of ~ | list of ~ |
| Attributes | / | / | | | list of ~ | list of ~ | list of ~ |
| Ranges in DOM | / | / | | | | ? | list of ~ |
| Valid URI | | | | | | | 2 |

# ITS and RDF - RDFa

| | NIF Recipe: offset http://ex.org#offset_... | NIF Recipe: hash http://ex.org#hash_... | http://ex.org r http://ex.org#myId | CSS selector 1/2/3 | XPath 1.0 | XPath 2.0 | XPointer 1.0 ex.org#xpointer(...) |
|---|---|---|---|---|---|---|---|
| Ranges in HTML source | one ~ | ? | | | | | |
| Elements | / | / | one ~ | list of ~ | list of ~ | list of ~ | list of ~ |
| Attributes | / | / | | | list of ~ | list of ~ | list of ~ |
| Ranges in DOM | / | / | | | | ? | list of ~ |
| Valid URI | | | | | | | 2 |

reduces verbosity (get rid of lots of spans)
**but** ITS annotations for range can't be added inline

# ITS and RDF - RDFa

| | NIF Recipe: offset http://ex.org#offset_... | NIF Recipe: hash http://ex.org#hash_... | http://ex.org r http://ex.org#myId | CSS selector 1/2/3 | XPath 1.0 | XPath 2.0 | XPointer 1.0 ex.org#xpointer(...) |
|---|---|---|---|---|---|---|---|
| Ranges in HTML source | one ~ | ? | | | | | |
| Elements | / | / | one ~ | list of ~ | list of ~ | list of ~ | list of ~ |
| Attributes | / | / | | | list of ~ | list of ~ | list of ~ |
| Ranges in DOM | / | / | | | | ? | list of ~ |
| Valid URI | | | | | | | [2] |

**XPointer 1.0:**

small extension to XPath but hard to implement ?

http://www.w3.org/XML/2002/10/LinkingImplementations.html

[2]: For a XPointer to be a valid URI, characters [ ] / ? # @ need to be escaped

http://www.w3.org/TR/xptr-framework/#escapingModel

# ITS and RDF - RDFa

## Suggestion for ITS 2.0

1. **use XPointer 1.0** in selector attribute, and in new attributes
2. **"the resulting locations MUST be either element node or attribute node or range nodes."** (c.f. ITS 1.0 REC.)
3. **"ITS 2.0 implementations MUST implement XPointer"** (may use NIF's ?)
4. **use `str:StringSet`** and **`str:String`** in the mappings to RDFa

## Suggestion of requirement for NIF 2.0

introduce **`str:StringSet`** for the class of a XPointer URI evaluation
+ other requirements to be discussed

# ITS and RDF – RDFa

## Still one big issue with RDF / RDFa

How to deal with attribute inheritance / overriding ?

# 2. CURIES:
# USE URIS WITH LESS VERBOSITY

# CURIEs: use URIs with less verbosity

**CURIE[1] = 'compact URI' expressions** (e.g., rdfs:label)

**Suggestion for ITS 2.0**

reuse these interesting features of RDFa : **`@vocab`, `@prefix`, CURIE Datatype**

limit the verbosity of a (X)HTML + ITS 2.0 document
ease the transformation to RDFa.

**Example**:

less verbose, e.g.:
- one/multiple XPointers in a single `@its-selector` for a global rule
- no need for `@its-terminology`, just use CURIE(s) in `@its-conceptReference`
- ... e.g., its-conceptReference="ex:SemanticWeb"

[1]: http://www.w3.org/TR/rdfa-core/#s_curies

# DRAW OUR INSPIRATION FROM THE PROVENANCE – XG

# Draw our inspiration from the Provenance – XG

PROV Data Model[1]:
PROV-XML, an XML schema for the PROV data model
PROV-O, the PROV ontology, an OWL-RL ontology allowing the mapping of PROV to RDF
+ other...

## Suggestion for ITS 2.0 : multiple facets

1. **ITS Data Model:** "Conceptual, prose definitions of data categories“
2. **ITS-XML**, an XML schema for the PROV data model
   Global rule = `its:Rule` element with `@selector="`<a XPointer>"
3. **ITS-O**, the ITS ontology allowing the mapping of ITS to RDF
   Global rules = simple `its:*` properties on <a XPointer> `rdf:type its:Rule`
4. **ITS-HTML**,
   `@its-*` attributes on elements
5. **ITS-HTML-RDFa**, `its:*` properties that can be used in a HTML document.
6. **ITS-HTML-Microdata**, nested groups of name-value pairs that can be added to a HTML doc.

**for 4., 5., 6., XML or RDF companion document to store :**
**global rules, annotations, older versions, annotations that don’t fit in the HTML...**

[1]: http://www.w3.org/TR/prov-dm

# Draw our inspiration from the Provenance – XG

PROV Data Model[1]:

**Agents** lead **Activities** on **Entities**

in MLW-LT:

**Translators** lead **LT-activities** on **fragments of text**

## Suggestion for ITS 2.0

Extend the Provenance Data Model

$\Rightarrow 7^{th}$ facet:

**7. ITS-PROV-Mapping**, a mapping from ITS Data Model to PROV Data Model

[1]: http://www.w3.org/TR/prov-dm

# Draw our inspiration from the Provenance – XG

PROV Data Model[1]:

**Agents** lead **Activities** on **Entities**

in MLW-LT:

**Translators** lead **LT-activities** on **fragments of text**

## Suggestion for ITS 2.0

Re-read users, activities, ... in terms of Provenance Entities, Activities, Agents

| Agents | Activities | Entities |
|---|---|---|
| **prov:Organization**<br>*ex:myLSP, ...* | **its:HumanTranslation ?** | **its:QAResult**<br>*informations on a QA* |
| **prov:Person**<br>*ex:John, ...* | **its:MachineTranslation ?** | **str:String* ?**<br>*a document, a span ...* |
| **prov:SoftwareAgent**<br>*ex:BingTranslator102* | **its:QualityAssessment** | |

**subClasses**
*instances*

[1]: http://www.w3.org/TR/prov-dm

# Draw our inspiration from the Provenance – XG

PROV Data Model[1]:

**Agents** lead **Activities** on **Entities**

in MLW-LT:

**Translators** lead **LT-activities** on **fragments of text**

Table 3: PROV-DM Relations At a Glance[1]

| | Entity | Activity | Agent | Note |
|---|---|---|---|---|
| Entity | wasDerivedFrom<br>wasRevisionOf<br>hadOriginalSource | wasGeneratedBy | wasAttributedTo | hasAnnotation |
| Activity | | wasStartedByActivity<br>wasInformedBy | wasAssociatedWith | hasAnnotation |
| Agent | | | | hasAnnotation |
| Note | | | | hasAnnotation |

**Suggestion for ITS 2.0**

Introduce our relations and annotations

[1]: http://www.w3.org/TR/prov-dm/#data-model-components

# RESTRICT LOCAL ITS ANNOTATIONS FOR HTML

# Restrict local ITS annotations for HTML

**3 combined Issues for local HTML annotations**

1. Can't express complex set of ITS attributes: (can't introduce `its-*` element, attributes)
2. Can't annotate inline `str:String` that are not DOM elements (NIF Receipe: XPointer)
3. DOM elements are `str:String`, not Activities (QAResults, ...), Agents (SoftwareAgent, ...), ...

**Suggestion for ITS 2.0**

in HTML, restrict local ITS annotations to only a subset of ITS data categories: those that apply directly on DOM elements `str:String` entities.

Other annotations must be made global.

# Restrict local ITS annotations for HTML

## Suggestion for ITS 2.0

**Possible solutions to make annotations global:** as simple as for javascript

1. write directly ITS-XML or ITS-RDF in a `script` element under the `head` element

2. link to a ITS-XML or ITS-RDF companion ITS file through a `link` element under the `head` element

=> Keep the HTML facet of the recommendation very light

http://www.w3.org/TR/html-markup/script.html
http://www.w3.org/TR/html-markup/link.html
http://lists.w3.org/Archives/Public/public-rdf-comments/2012Jun/0007.html

# PUBLICATION OF SCHEMAS AND VOCABULARIES

# Publication of schemas and vocabularies: Content negociation


**recommendation:** http://www.w3.org/TR/skos-reference/
**namespace** : http://www.w3.org/2004/02/skos/core

Look at HTTP requests/responses for SKOS namespace

http://www.w3.org/2004/02/skos/core
"HTTP 303 See other" to http://www.w3.org/2009/08/skos-reference/skos
+ content negociation:
http://www.w3.org/2004/02/skos/core.html - human readable description of vocab
http://www.w3.org/2004/02/skos/core.rdf - application/rdf+xml description of vocab

# Publication of schemas and vocabularies: Content negociation

## Suggestion for ITS 2.0

**use the same ITS 1.0 namespace + redirection + content negociation**

**ITS 2.0 recommendation:** http://www.w3.org/TR/skos-reference/

**ITS 2.0 namespace** : http://www.w3.org/2005/11/its

1. **"HTTP 303 See other" to** http://www.w3.org/TR/its-2.0/its
2. **“HTTP 301 Moved Permanently” to:**

**human readable description of the data model**

http://www.w3.org/TR/its-2.0/its.html (when HTTP accept:text/html)

**application/rdf+xml description of the schema**

http://www.w3.org/TR/its-2.0/its.rdf (default otherwise)
http://www.w3.org/TR/its-2.0/its.n3 text/n3
http://www.w3.org/TR/its-2.0/its.ttl text/turtle

**non-normative description of XML based ITS 2.0**

http://www.w3.org/TR/its-2.0/its.dtd DTD
http://www.w3.org/TR/its-2.0/its.ttl XSD

# Recap of suggestions for ITS 2.0

Make extensive use of **XPointer**, **NIF** String ontology, **CURIEs**

Re-read users, activities, data categories
in terms of **Provenance** Entities, Activities, Agents, relations.

For HTML
Define what data-categories can be kept local on the HTML elements,
Define what data-categories must be defined globally, in `script` or `link`

Clearly separate facets of the recommandation (stubs in the wiki)
Publication of schemas and vocabularies: redirection + content negociation


# MLW-LT and Representation Formats: Suggestions.

**Maxime Lefrançois** **INRIA – WIMMICS**
**maxime.lefrancois@inria.fr** **inria.fr – wimmics.inria.fr**

**The Multilingual Web – Linked Open Data and MultilingualWeb-LT Requirements, 11 - 13 June 2012, Dublin**