

TELECOM ParisTech's Interest in the Workshop  
**Semantic enrichment in EPUB3: new ways of reading**

|  |   |   |
|--|---|---|
| <b>Vincent Gros</b><br>Telecom ParisTech<br>Paris<br>Groupe Multimedia/Multimedia Group<br>+33141236098<br>vincent.gros@telecom-paristech.fr | <b>Jean-Claude Moissinac</b><br>Telecom ParisTech<br>Paris<br>Groupe Multimedia/Multimedia Group<br>+33145818088<br>moissinac@ telecom-paristech.fr | <b>Jean-Claude Dufourd</b><br>Telecom ParisTech<br>Paris<br>Groupe Multimedia/Multimedia Group<br>+33145817733<br>dufourd@ telecom-paristech.fr |
|--|---|---|

## 1. Participant's Interest

TELECOM ParisTech<sup>1</sup> is a French “Grande Ecole” in Paris, i.e. a small university dedicated to Telecommunications and teaching at graduate and PhD levels. TELECOM ParisTech laboratories are part of INSTITUT TELECOM Recherche. TELECOM ParisTech today has a faculty of about 150 full-time staff (full professors, associate and assistant professors), over 200 part-time lecturers and a student body of about 1000 students.

The Multimedia group of the Signal and Image Processing Department is:

- Involved with/implanter of W3C widgets, and extensions for widget communication within the home network (MPEG-U: interface with discovery, communication and agent-like mobility across devices);
- Involved with/implanter of SVG and other presentation standards in the open source platform GPAC;
- Involved with W3C and other standards for the past 15 years.

Initially focused on multimedia scene representation, the team expanded its interest toward the use of semantics in multimedia services.

In parallels with his PhD based on the EPUB standard (in partnership with Hachette Livre), the first author is also a contributor into IDPF projects: *Radium*<sup>2</sup>, the *epub3 samples project*<sup>3</sup>, and the *epub conformance test suite*<sup>4</sup>.

The third major release of the EPUB standard includes Web standards: (X)HTML5, CSS, SVG, and opens the door to new researches in electronic publishing. Our interest is how we can define new reading mechanisms into a book or a reader, by adding semantics in EPUB content. RDFa seems to be a great candidate to define this semantics.

## 2. Point of View

<sup>1</sup> <http://www.telecom-paristech.fr/nc/formation-et-innovation-dans-le-numerique.html>

<sup>2</sup> <https://github.com/readium/readium>

<sup>3</sup> <http://code.google.com/p/epub-samples/>

<sup>4</sup> <https://github.com/mgylling/epub-testsuite>

Here is a set of issues we work with:

- What kind of semantics is already defined in EPUB3: we are aware of the evolution of the Publication specification (<http://idpf.org/epub/30/spec/epub30-publications.html>), especially as regards the metadata and the use of the property data type. And we know that the epub:type attribute is a powerful tool, but can only be used to define the structure of the book, not a deep semantic enrichment. What about the *Content Switching* (epub:switch element, <http://idpf.org/epub/30/spec/epub30-contentdocs.html#sec-xhtml-content-switch>)?
- Our focus is on RDFa, because it can be integrated to (X)HTML5 and SVG, academic research is often based on this language to develop tools for extraction, annotation, or reuse content. What is the stability of the W3C standards in the field of the Semantic Web (to eventually consider integration in a future revision of EPUB3)? What about RDFa?
- How we can use semantics enrichment in the new paradigm of EPUB? Indeed, a website is first an online content, then it can be read offline; but one of the most interests of an EPUB file is that it is readable offline, and eventually it can be "connected". Then where store vocabularies: at the reader level? At the book level? Vocabularies could be embed in the EPUB file to be used offline, but what about the file size, or the Reader's performance? Can vocabularies be shared between books (into the same library)? We think that it is important to define both of use cases online and offline, with priority given to read offline whenever possible.
- What drive the new reading mechanisms based on semantics enrichment? The reader, or the book?

We believe the W3C standards are a key to this vision of a coherent family of interoperable standards. Only W3C standards have the necessary breadth and momentum, but there is a need for significant adaptation to really take into account books environments.