

# Workshop Publishing using the Open Web Platform

## Styles are not cascading!

Luc Audrain  
Head of Digitalization, Hachette Livre, Paris



### Participant's interest

As Head of Digitalization in Hachette Livre<sup>1</sup>, publishing group number one in France with 60 different brands, I have been involved for years in many production processes based on content creation tools that have largely evolved since the late 1990s.

In one part at least these tools so widely spread among authors and publishers have not changed, making it still difficult to reuse author's content: styles are not cascading.

### Mainstream text tools in publishing

In the upstream of the publishing chain, authors use the most largely available tools to input text, namely word processors. All text they produce at home for any kind of book is delivered to publishers as Word or LibreOffice files.

Then editorial teams prepare texts still using these specialized products and pages are built with desktop publishing tools where the text blocks are still managed as in word processors with paragraphs and styles.

### Print publishing workflows are not structured

One major limit of these text tools is the lack of semantic context. Mainly presentation oriented, those software don't enable easy semantic qualification of content.

To get anyway some semantics from word processors files, authors and publishers have setup lists of named styles to be used to qualify paragraphs and text zones. Paragraph style names and character style names are still today the best and only way to give name to text content when edited with those most widely spread authoring tools.

But this still brings a poor result as several limitations impair the goal:

- Styles can be easily bypassed using the wrong style and then editing the presentation of the paragraph to make it look as the correct style
- And mainly, there is no contextual heritage in these styles: a paragraph name persists as long as another paragraph comes, and is then forgotten.

Thus a chapter title paragraph cannot tell where the chapter ends, and if you want to enable semantics, you have to name it as "chaptertitle", this forbidding you to separate chapter style from title style.

## Digital need structure

In our digital era, content reusability is a must to produce any digital format. These new channels need structured content, not only for accessibility reason which a good one in first place, but also for navigation (TOC, indexes), quality display and all sorts of new usages on mobile devices.

## Facing those digital needs

Trying to enable all these new digital functions on word processor styles only is a mess and that's why when publishers decide to build apps, developers ask for the book content in XML.

Also, when publishers plan to build web sites to bring readers interactivity and much more user friendly access, they had to face a huge content preparation as their assets were for print (PDFs, InDesign or word files).

Not to speak of ebooks where the simplest novel need to be structured to enable digital accessibility, and any other more semantically rich content cannot benefit from a linear sequence of paragraphs neither as a suite of fixed layout pages.

## Existing solutions are still niches

Starting content creation in a structured context is obviously possible as new tools have been made available.

Examples of database input for specialized information which is then exported in a structured way to composition systems can be found.

XML first workflows tend to appear also where the input tools are XML or xHTML editors with a user friendly interface "à la Word".

But I have to say that it still stays in niches and it always needs heavy investments in software development, production setup and training. It is still reserved for specialized content and small dedicated teams with strong managerial vision.

Though these practices are most of the time profitable hopefully, they never went mainstream to replace the basic standard style oriented word processor everybody is using at the very start of content input.

## Web content is structured but not semantic enough

The hierarchal concept of content is well implemented in Web standards. HTML/xHTML is a tree model where main division encompass smaller one. A chapter title is in a section that has a precisely positioned end tag.

Then the styles benefit from heritage with cascading effects from the top of the tree to the smallest inner portion of text. Cascading style sheets allow for a clear separation between for instance, chapter style and title style.

This is well done for the web but editorial content need more semantics. As soon as publishers try to structure their content, they are facing the need of semantic tagging and they design XML vocabularies much more rich and specialized as what HTML even in its 5<sup>th</sup> version enable.

Let's think of cooking books, travel or wine guides, school and university books, dictionaries and other highly structure content. None of these can be tagged with HTML.

### **Open Web platform: structure to everyone?**

The Open Web platform is aimed at all Internet users, and authors who are not Internet users are seldom these days.

How could the Open Web platform bring structured content input to everyone?

Hopefully, this workshop will bring some ideas to this recurring question in publishing environment.

---

<sup>i</sup> Hachette Livre : <http://www.hachette.com/en/>