HTML5 or messages from beyond the PDF

Martin Fenner, Gobbledygook

January 5, 2011

In 1990 Tim Berners-Lee and others started HTML and the world wide web to facilitate scientific communications at CERN, the world's largest particle physics laboratory.

xmos_basser by .M., on Flickr, November 1993

Although the world wide web profoundly changed scholarly publishing (and of course many other things), HTML did not become the standard document format for scientific papers. In fact, there is no standard document format. We have document formats for authors, for the internal workflow of publishers, and for the distribution and reading of papers.

There are of course many good reasons to use LaTeX for writing, XML for workflows, PDF to print papers or ePub for mobile devices. But reformatting a manuscript into different formats several times is both expensive (in terms of time and costs) and means that the formatting options used will be a compromise of what is available in all formats.

All this would be much easier if we just used HTML. With HTML, authors, publishers and readers can all use the same document format. And they will have an endless number of tools at their hands, including of course WordPress for writing and the web browser of choice for reading. HTML in 2010 is very different from HTML in 1990. HTML5 supports new semantic elements such as <article>, microdata, embedding of video without plugins, geolocation, and offline web applications.

An HTML-based scholarly publishing workflow will make it

- faster and cheaper to publish a paper,
- easier to create rich interactive documents,
- easier to add additional steps, such as integration of data from lab notebooks, publishing of pre-prints, etc.
- easier to integrate additional services, from data visualisation to language editing.
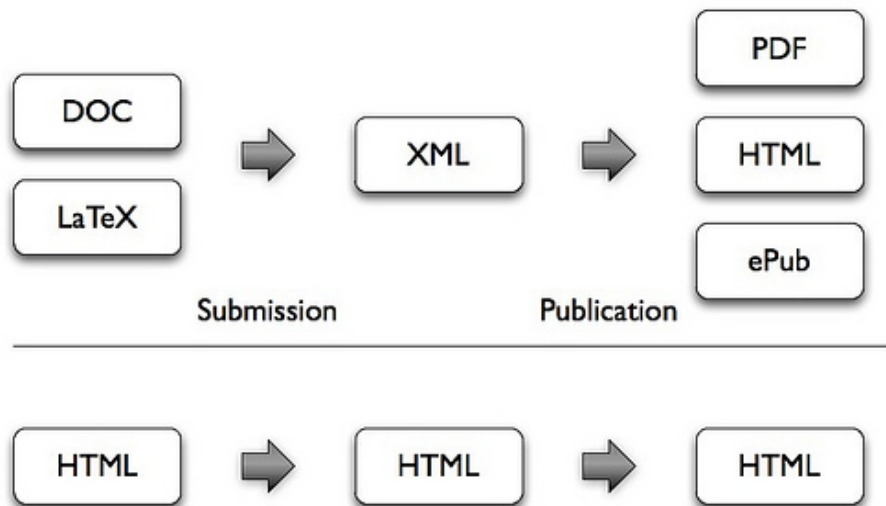
Scholarly Publishing Workflow

DOC

LaTeX

XML

PDF

HTML

ePub

Submission   Publication

HTML   HTML   HTML

Figure 1: