# Open Research Online

## The RASH Framework: enabling HTML+RDF submissions in scholarly venues

### Conference or Workshop Item

For guidance on citations see FAQs.

# The RASH Framework: enabling HTML+RDF submissions in scholarly venues

Angelo Di Iorio[1], Andrea Giovanni Nuzzolese[1,2], Francesco Osborne[3],
Silvio Peroni[1,2], Francesco Poggi[1], Michael Smith[4,5], Fabio Vitali[1], and
Jun Zhao[6]

[1] Department of Computer Science and Engineering, University of Bologna,
Bologna, Italy
[2] Semantic Technology Laboratory, Institute of Cognitive Sciences and Technologies,
Italian National Research Council, Rome, Italy
[3] Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom
[4] World Wide Web Consortium, Shinjuku, Tokyo, Japan
[5] Graduate School of Media and Governance, Keio University,
Fujisawa, Kanagawa, Japan
[6] School of Computing and Communications, Lancaster University,
Lancaster, United Kingdom
angelo.diiorio@unibo.it, andrea.nuzzolese@istc.cnr.it,
francesco.osborne@open.ac.uk, silvio.peroni@unibo.it,
francesco.poggi@unibo.it, mike@w3.org, fabio.vitali@unibo.it,
j.zhao5@lancaster.ac.uk

**Abstract.** This paper introduces the *RASH Framework*, i.e., a set of
specifications and tools for writing academic articles in *RASH*, a simpli-
fied version of HTML. RASH focuses strictly on writing the content of
the paper leaving all the issues about its validation, visualisation, conver-
sion, and data extraction to the tools developed within the framework.

**Keywords:** Digital Publishing, RASH, Semantic Publishing, Semantic
Web, XSLT, document conversion

## 1 Introduction

In the last months of 2014, several posts within technical mailing lists of the Web
and Semantic Web community have discussed an evergreen topic in scholarly
communication, i.e., how authors of research papers could submit their works in
HTML rather than, say, PDF, MS Word or LaTeX. Besides the obvious justifica-
tion of simplification and unification of data formats for drafting, submission and
publication, an additional underlying rationale is that the adoption of HTML
in the context of scientific publications would ease the embedding of semantic
annotations, thus making a step towards the improvement of research communi-
cations thanks to already existing W3C standards such as RDFa and Turtle. The
adoption of Web-first formats in scientific literature, i.e., HTML and RDF, is a
necessary step towards the complex (and exciting) scenarios that the Semantic

Publishing has promised us [1] [6]. However, such formats should support the needs of the actors involved in the production/delivery/use of scholarly articles.

Along the lines of other existing works on this topic (e.g., *Linked Research* project [2] and *ScholarlyMarkdown* [5]), in this paper we introduce the *RASH Framework*, i.e., a set of specifications and tools for writing academic articles in *RASH* (an HTML+RDF-based markup language for writing scholarly documents) which aims at addressing all the aforementioned issues.

The rest of the paper is structured as follows. In Section 2 we introduce the rationale for the RASH Framework. In Section 3 we provide a quick overview of all its tools, that are available in the Framework repository. Finally, in Section 4 we conclude the paper sketching out some future developments.

## 2   A "Web-first" framework for research articles

Some works, e.g., Capadisli *et al.* [2], suggest not to force any particular HTML structure for research papers. In this way, the author of a paper is free to use any possible kinds of HTML *linearisations* for writing a scholarly text. This freedom could, however, results in two main kinds of issues:

- *visualisation bottleneck* – it may affect the correct use of existing, well-developed and pretty standard CSSs;
- *less focus on the research content* – the fact that a certain paper is not visualised in a browser very well could bring the author to work on the presentation of the text, rather than on its research content.

A further complication to an already complex scenario comes from the *necessary* involvement of publishers. Leaving the authors free of using their own HTML format could be also counterproductive from a publisher's perspective, in particular when we speaking about the possibility of adopting such HTML formats for regular conference/journal camera-ready submissions.

The RASH Framework[7] has been proposed in order to address all the aforementioned issues. It is a set of specifications and tools for writing academic articles in *RASH* - a summary of the whole framework is introduced in Fig. 1.

The *Research Articles in Simplified HTML* (*RASH*) format is a markup language that restricts the use of HTML elements to only *25 elements* for writing academic research articles, and it is entirely based on a strong theory on *structural patterns* for XML documents [4]. It allows authors to use RDFa annotations within any element of the language. In addition, RASH allows the use of elements `script` (with the attribute `type` set to "text/turtle" or to "application/ld+json") within the element `head` for adding plain Turtle or JSON-LD content. Any RASH documents begins as a simple (X)HTML5 document[8], by

---

[7] The full project is available at `https://github.com/essepuntato/rash/`. Please use the hashtag `#rashfwk` for referring to any of the items defined in the RASH Framework via Twitter or other social platforms.

[8] Please refer to the official RASH documentation, available at `http://cs.unibo.it/save-sd/rash`, for a complete introduction of all the elements and attributes that can be used in RASH documents.

specifying the document element `html` (with the usual namespace) that contains the element `head` for defining metadata of the document, and the element `body` for including the whole content of the document.
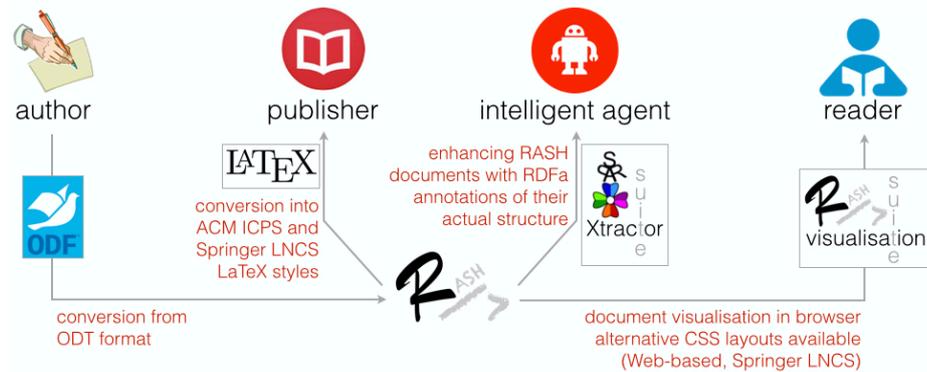


**Fig. 1.** The RASH Framework and its components addressing needs of different users.

## 3   Tools in the Framework

In this section we introduce all the tools shown in Fig. 1 that we have developed in order to support users in adopting RASH - all the tools are distributed under an ISC License or a CC-BY 4.0 International License.

**Validation.** All the markup items in RASH are defined as a RelaxNG grammar and are compatible with HTML5. We have developed a script to enable RASH users to check their documents simultaneously both against the specific requirements in the RelaxNG grammar and also against the full set of HTML checks that the W3C Nu HTML Checker does for all HTML documents.

**Visualisation.** The visualisation of RASH documents is rendered by the browser in the current form by means of appropriate CSS3 stylesheets and javascript scripts developed for this purpose. We are actually using some external libraries, i.e., Bootstrap and JQuery, in order to guarantee the current clear visualisation and for adding additional tools to the user. As an example, the RASH version of this paper is available at `https://rawgit.com/essepuntato/rash/master/papers/rash-demo-iswc2015.html`.

**Conversion**. We have spent some efforts in preparing XSLT 2.0 documents for converting RASH documents into different LaTeX styles, such as ACM ICPS and Springer LNCS. This is, actually, one of the crucial steps to guarantee the use of RASH within international events[9] and to be able to publish RASH documents in the official LaTeX format as required by the organisation committee

---

[9] `https://github.com/essepuntato/rash/#venues-that-have-adopted-rash-as-submission-format`

of such events. In addition, we have already developed another XSLT 2.0 document to perform conversions from OpenOffice documents into RASH documents, which allows authors to write a paper through the OpenOffice editor and then converting the related ODT file into RASH automatically.

**Enhancement.** A recent development of the RASH Framework has concerned the automatic enrichment of RASH documents with RDFa annotations defining the actual structure of such documents in terms of the Document Component Ontology (DoCO) [3]. In particular, a Java application called *SPAR Xtractor suite* has been developed: it takes a RASH document as input and returns a new RASH document where all its markup elements have been annotated with their actual (structural) semantics.

## 4   Conclusions

In this paper we have introduced the *RASH Framework*, i.e., a set of specifications and tools for writing academic articles in *RASH*. We have discussed the rationale behind the development of RASH, and we have presented the language with all the validation/visualisation/conversion/extraction tools we have developed so far. As immediate future developments, we plan to create additional scripts for extracting RDF statements from RASH documents according to SPAR Ontologies (`http://www.sparontologies.net`), and to develop additional XSLT documents in order to convert DOCX documents into RASH and to convert RASH documents into several formats for scholarly communications, such as EPUB, DocBook, and LaTeX IEEE styles.

## References

1. Bourne, P. E., Clark, T., Dale, R., de Waard, A., Herman, I., Hovy, E. H., & Shotton, D. (2011). FORCE11 White Paper: Improving The Future of Research Communications and e-Scholarship. White paper, 28 October 2011. FORCE11. `https://www.force11.org/white_paper`
2. Capadisli, S., Riedl, R., & Auer, S. (2015). Enabling Accessible Knowledge. In Proc. of CeDEM 2015. OA version available at `http://csarven.ca/enabling-accessible-knowledge`
3. Constantin, A., Peroni, S., Pettifer, S., Shotton, D., & Vitali, F. (in press). The Document Component Ontology (DoCO). To appear in Semantic Web. OA version available at `http://www.semantic-web-journal.net/system/files/swj1016.pdf`
4. Di Iorio, A., Peroni, S., Poggi, F., & Vitali, F. (2014). Dealing with structural patterns of XML documents. Journal of the American Society for Information Science and Technology, 65(9): 1884–1900. `http://dx.doi.org/10.1002/asi.23088`
5. Lin, T. T. Y., & Beales, G. (2015). ScholarlyMarkdown Syntax Guide. Guide, 31 January 2015. `http://scholarlymarkdown.com/Scholarly-Markdown-Guide.html`
6. Shotton, D., Portwin, K., Klyne, G., & Miles, A. (2009). Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. PLoS Computational Biology, 5(4): e1000361. `http://dx.doi.org/10.1371/journal.pcbi.1000361`