

Semantic Publishing of Knowledge about Amino Acids

Robert Stevens¹ and Phillip Lord²

¹ ¹School of Computer Science, University of Manchester, UK

² ²Department of Computing Science, University of Newcastle, Newcastle, UK
`Robert.Stevens@Manchester.ac.uk`

Abstract. We semantically publish knowledge about the amino acids commonly described within biochemistry. We do this as an ontology written in OWL and presented as XML/RDF. The classification of amino acids is based on Taylor's article (PMID:3461222) from 1986 published in the *Journal of Theoretical Biology*. The ontology goes further than the static paper version; it combines many aspects of the physicochemical properties Taylor uses to classify amino acids to give a rich, multi-axial classification of amino acids. Taylor's original description of the amino acid's physicochemical properties are captured with value partitions and restrictions on the amino acid classes themselves. A series of defined classes then establishes the multi-axial classification. The publication, when loaded into an OWL ontology manipulation tool, allows some knowledge about amino acids to be explored and used computationally. By publishing this knowledge about amino acids as a semantic document in the form of an ontology we pursue an agenda of disruptive technology in publishing. It allows us to 'push' at the nature of a semantic publication.

Blogs about the published semantics of amino acids may be found at <http://robertdavidstevens.wordpress.com/2010/12/18/an-update-to-the-amino-acids-ontology/> and links following. The ontology is at <http://www.cs.man.ac.uk/~stevensr/ontology/amino-acids.owl>.

It is, perhaps, an ontologist's question to ask 'what is semantic publishing?'. When is a publication semantic and when does some computational semantic artefact become a publication? A further question is when is a semantic publication a scientific publication? Our submission to *Se publica 2012* was an experiment in this area—or it was the authors 'just trying it on'. Whichever it is, the reviewers have gone along with our game, so here's a narrative around our submission of an ontology of amino acids written in the Web Ontology Language (OWL) as a semantic publication to *se publica 2012*. This narrative is a side-effect of our attempt at semantic publishing—we used our ontology of amino acids as a semantic publication, but does that count?; what is actually published and what can actually be read? this short text is really the front-end to our amino acids semantic publication, but it turns out that the narrative it provides, or something like it, is a necessary part of (scientific) semantic publishing.

A snippet from the Sepublica ‘instructions to authors’³ shows the origins of our submission:

We also invite submissions in XHTML+RDFa or in the format of YOUR semantic publishing tool. However, to ensure a fair review procedure, authors must additionally export them to PDF.

this made us ask ‘what would happen if we submitted an RDF document for one of our ontologies as a submission to Sepublica?’. Our reasoning went something like this:

- Sepublica can have exactly what they’ve asked for. . .
- An ontology in OWL has an RDF syntax, so it matches the representation criterion;
- it has a URI that means it is published on the web, so it matches the publication criterion;
- The ontology captures some knowledge about a field of interest—that is, the semantics of that field, so it matches the semantic criterion;
- The ontology can be argued to be a document. . .

so, that ontology is a semantic publication. Anyway, we decided to ‘try it on’ and, to their credit, both the workshop organisers (after a query to find out if we’d done what we meant to do) and the reviewers went along with what we did. Given that the Amino Acids Ontology, in its RDF form, was accepted as a publication for Sepublica, we can conclude that it is a semantic publication.

Another interesting aspect of the Sepublica process is that the instructions asked for a PDF submission (in addition to any semantic submission) to ease the reviewing process. So, partly because the EasyChair site for sepublica was only set up to submit PDF and to take the organisers at their word, we first submitted a Manchester OWL Syntax version of the ontology and converted it into PDF. MOS is a more or less human readable syntax for OWL. However, the PDF version of the MOS wasn’t especially useful. So, I asked for EasyChair to be set up to allow non-PDF submission; it turns out that a zip file was the only way of achieving submission of an RDF document. If we are going to have semantic scientific publications, then we need a way of handling them; not just in EasyChair’s reviewing process, but in the wider context of the scientific workflow.

The blogs above give sufficient background for the ontology, but here is an outline. The Amino Acid ontology is a simple ontology that captures some basic conceptualisations of amino acids used by biochemists [1]. It has the basic criteria by which biochemists classify amino acids—size, polarity, charge, aromaticity and hydrophobicity. Only the biologically used amino acids are allowed and there are various constraints on the qualities permitted for the amino acids. It works both as an exemplar of the role of automated reasoning in ontology maintenance and as a ‘guide’ to the amino acids. A complex hierarchy of the amino acids is

³ <http://sepublica.mywikipaper.org/drupal/node/23> accessed March 14 2012.

then offered, including some types of amino acid that cannot exist.⁴ Thus the ontology captures the semantics of amino acid entities in some computational form over which reasoning can be performed. The ontology can be browsed using some OWL enabled tool and it can act as an amino acid ‘tutorial’ as well as supply computational semantics about amino acids to applications. This form of publication of Taylor’s classification offers more than the original paper in terms of explicitness, computational manipulation and flexibility. It does, however, lack some, to say the least, of the context and narrative needed for a scientific publication.

We offered our submission to Sepublica as a *disruptive technology*⁵; semantic publishing should be a disruptive technology by creating a new publication market and changing the values by which scientific publishing happens. That Sepublica stil needs to ask for PDF to enable review (though the reviewers of the Amino Acids Ontology managed without) means that publication has not been disrupted enough; data are available with some computational semantics, but we don’t have semantic scientific publication.

What does all of this tell us? The Amino Acids Ontology is a semantic publication, but also that it isn’t really and it isn’t a scientific semantic publication. While the ontology captures the semantics of the domain in a computational form, it lacks the narrative that semantic publication of data needs to make it useful for humans. We don’t want the reverse of the current situation of all narrative and no computation, to be replaced by all computational semantics and no human narative. As others have already said, semantic publication needs human narrative. This would make most of the RDF only link data publications of scientific data only partially a semantic publication; linked data is necessary but not sufficient for semantic publication.

References

1. Taylor, W.R.: The classification of amino acid conservation. *Journal of Theoretical Biology* **119**(2) (1986) 205–218

⁴ Links to supplementary information and the ontology itself are available in the abstract.

⁵ http://en.wikipedia.org/wiki/Disruptive_innovation