# M6.15 - Sample papers testing the XML-based editorial workflow elaborated and published

**Leading partners:** Pensoft, NHM

**Compiled by:**

Lyubomir Penev, Teodor Georgiev, Pavel Stoev, Jordan Biserkov, Margarita Grudova

November 2012

# Refinement of the XML schemas for submission of manuscript from Scratchpads to ZooKeys

A paper by Blagoderov et al. Published in Zookeys 2010 (doi: 10.3897/zookeys.50.539) sets out a workflow that describes the assembly of elements from a Scratchpad taxon page (http://scratchpads.eu) to export a structured XML file (Fig 1). The publisher receives the submission, automatically renders the file into the journal‘s layout style as a PDF and transmits it to a selection of referees, based on the key words in the manuscript and the publisher’s database. Several steps, from the author’s decision to submit the manuscript to final publication and dissemination, are automatic or semi-automatic. A journal editor first spends time on the submission when the referees’ reports are received, making the decision to publish, modify or reject the manuscript. If the decision is to publish, then PDF proofs are sent back to the author and, when verified, the paper is published both on paper and on-line, in PDF, HTML and XML formats. The original information is also preserved on the original Scratchpad where it may, in due course, be updated. A visitor arriving at the web site by tracing the original publication will be able to jump forward to the current version of the taxon page.

The publication module of Scratchpads and the export of structured content in XML format from Scratchpads to ZooKeys has been elaborated and exemplified by Faulwetter et al. in a ZooKeys paper published in 2011. The paper deals with a newly described species of marine flat worm found in the Mediterranean sea. The XML schema used in the manuscript was TaxPub, an extension to the Document Type Definitions (DTD) of the US National Library of Medicine Journal Archiving and Interchange Tag Suite (NLM). Since July 2009, TaxPub has been routinely implemented in everyday publishing practice of Pensoft, to provide: (1) Semantically enhanced, domain-specific XML versions of articles for archiving in PubMedCentral (PMC); (2) Visualisation of taxon treatments on PMC; (3) Export of taxon treatments to various aggregators, such as Encyclopedia of Life, Plazi Treatment Repository, and the Wiki Species-ID.net. However, with the development of the Biodiversity Data Journal and Pensoft’s Writing Tool and after consultations with a number of leading bioinformaticians, a number of flaws in the use of TaxPub schema become apparent, especially its limited flexibility for server-to-server XML transactions, including the export of manuscripts from Scratchpads to ZooKeys. At first place, TaxPub was originally designed for taxonomic treatments and, from its conformity to the overall complex structure of a research articles stored in PubMedCentral, it has a quite rigid hierarchical structure. Besides simple taxonomic treatments several other types of manuscripts, such as data papers, software descriptions and checklists, are envisaged for acceptance and publication in BDJ, so the overall XML schema for export from Scratchpads through the PWT API were significantly expanded to reflect the specifics of each of the different publication templates. In two subsequent meetings held in Sofia (Vibrant meeting, 16-18 April 2012) and Beijing (annual TDWG conference, 22-26 October 2012) a number of important decisions were taken in this connection and a new workflow and XML schema developed (Fig. 2).

The main results of the meetings are summarised below:

1. Scratchpads should export full files not just references to files, because if users can delete files in Scratchpads so there is the potential for the file references to become invalid.
2. Rename of element names to be self-describing using the existing attribute values. This will make XSLT and XPATH access to the data much easier, and will make the data self-documenting.
3. Overview of XML suggests that data exchange looks promising with a large amount of overlap with near identical atomisation. Areas of difference seem principally around authoring information, such as contributors’ roles. In particular co-ordination of the submitting author and the different role-based permissions in the applications. Scratchpads will need a list of the PWT fields, and which are mandatory.
4. PWT needs to map the Drupal Biblio fields.

Although PWT is fully operational in beta version at the time of writing this explanation note, the public test is expected to happen in February-March 2013 when not only PWT but also BDJ the publishing workflow will be fully operational (month 27). The test will be based on a automated export from Scratchpads to PWT API and published in Biodiversity data Journal. The following papers, among many others expected, are planned to be submitted and published:

1. Golovatch and Stoev. The millipede family Paradoxosomatidae in the Philippines, with the description of the first, new, Philippine species of Eustrongylosoma Silvestri, 1896, and notes on Anoplodesmus anthracinus Pocock, 1895, recorded in Malaysia for the first time (Diplopoda, Polydesmida).
2. Arvanitidis, Faulwetter et al. Data paper describing data collected through ViBRANT’s SCOMBER project.



Fig. 1. A chart showing the four stages of an XML-based editorial, publication and dissemination workflow applied in ZooKeys from 2008 to 2012. Forms in blue are either implemented or prototyped, forms in red are in a process of development.



Fig. 2. A chart showing the newly developed Biodiversity Data Journal XML work flow, from creation of manuscript in PWT or export from Scratchpads, to PJS and integration of the published content into major data aggregators.