Conference Abstract

# Entry Life-Cycle with automatic Change-History & Provenance Tracking in collaborative Semantic Web Content Management Systems as implemented in SOCCOMAS

Roman Baum[‡], Peter Grobe[§], Christian Köhler[‡], Sandra Meid[‡], Björn Quast[§], Lars Vogt[‡]

‡ Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany
§ Zoological Research Museum Alexander Koenig, Bonn, Germany

Corresponding author: Lars Vogt (lars.m.vogt@googlemail.com)

## Abstract

SOCCOMAS is a ready-to-use Semantic Ontology-Controlled Content Management System (http://escience.biowikifarm.net/wiki/SOCCOMAS). Each web content management system (WCMS) run by SOCCOMAS is controlled by a set of ontologies and an accompanying Java-based middleware with the data housed in a Jena tuple store. The ontologies describe the behavior of the WCMS, including all of its input forms, input controls, data schemes and workflow processes (Fig. 1).
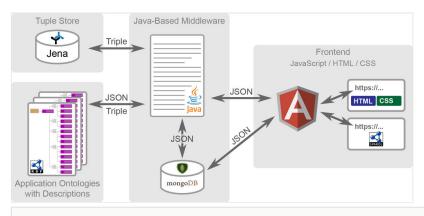
Figure 1.

Overall workflow of SOCCOMAS. Left: Jena tuple store and the descriptions of data views and workflows in the application ontologies. Middle: the Java-based middleware. Right: the frontend based on the JavaScript framework AngularJS with HTML and CSS output for browser requests and access to a SPARQL endpoint for service requests.

Data is organized into different types of data entries, which represent collections of data referring to a particular material entity, for instance an individual specimen. SOCCOMAS implements a suite of general processes, which can be used to manage and organize all data entry types. One category of processes manages the life-cycle of a data entry, including all required for changing between the following possible entry states:

1. current draft version;
2. backup draft version;
3. recycle bin draft version;
4. deleted draft version;
5. current published version;
6. previously published version.

The processes also allow a user to create a revised draft based on the current published version. Another category of processes automatically tracks the overall provenance (i.e. creator, authors, creation and publication date, contributers, relation between different versions, etc.) for each particular data entry. Additionally, on a significantly finer level of granularity, SOCCOMAS also tracks in a detailed change-history log all changes made to a particular data record at the level of individual input fields. All information (data, provenance metadata, change-history metadata) is stored based on Resource Description Framework (RDF) compliant data schemes into different named graphs (i.e. a URI under which triple statements are stored in the tuple store). All recorded information can be accessed through a SPARQL endpoint. All data entries are Linked Open Data and thus provide access to an HTML representation of the data for visualization in a web-browser or as a machine-readable RDF file. The ontology-controlled design of SOCCOMAS allows administrators to easily customize already existing templates for input forms of data entries, define new templates for new types of data entries, and define underlying RDF-compliant

data schemes and apply them to each relevant input field. SOCCOMAS provides an engine for running and developing semantic WCMSs, where only ontology editing, but no middleware and front end programming, are required for adapting the WCMS to one's own specific requirements.

## Keywords

entry life-cycle, provenance tracking, change-history tracking, ontology, knowledge base, CMS, Linked Open Data, SOCCOMAS

## Presenting author

Roman Baum, Christian Köhler

## Funding program

## Grant title

Digitisation / Cataloguing of non-textual objects: eScience-compliant standards for morphology