

## Preface

These Proceedings contain the papers accepted for publication and presentation at the first International Workshop on Semantic Web for Scientific Heritage (SW4SH) held in conjunction with the 12<sup>th</sup> ESWC 2015 Conference on June 1, in Portoroz, Slovenia. This workshop aimed at providing a leading international and interdisciplinary forum to disseminate the latest research in the field of Semantic Web for the study of pre-modern scientific texts and of the history of ideas and their transmission.

The program committee members are all involved in this interdisciplinary synergy. They have accepted nine papers (ratio 75%) and invited two keynote speakers. The four workshop organizers belong to the Zoomathia<sup>1</sup> international research network funded by the French National Scientific Research Center (CNRS). This network gathers French, Italian, German and English researchers and aims to study the formation and transmission of ancient zoological knowledge over a long period, with an historical, literary and epistemological approach, and create open knowledge sources on classical zoology to be published on the Web of Linked Data. This workshop was also planned as an opportunity to present the activity of the network, to enlarge it with interested participants of the workshop, and to benefit from the results of related research projects.

This encounter takes place within the general context of Digital Humanities, a research area at the intersection of Humanities and Computer Science which is gaining an ever-increasing momentum and where the Linked Open Data is playing an increasingly prominent role. The opportunity of the workshop was to provide a forum for discussion about the methodological approaches to the specificity of annotating “scientific” texts (in a wide meaning, including disciplines such as history, architecture, or rhetoric), and to support a collaborative reflection on possible guidelines or specific models for building historical ontologies. The iconographic data are also relevant in history of science and arise similar problematics; they offer suggestive insights for a global methodology for diverse media.

The opportunity for a fruitful encounter of knowledge engineers with computer-savvy historians and classicists has come. Since the mid-1970s, classicists and historians have developed textual databases, intending mostly to gather and explore large amounts of primary source materials. For a long time, they mainly focused on text digitization and markup. They only recently decided to try to explore the possibility of transferring some analytical processes they previously thought incompatible with automation to knowledge engineering systems, thus taking advantage of the growing set of tools and techniques based on the languages and standards of the semantic Web, such as linked data, ontologies, and automated reasoning. On the other hand, Semantic Web researchers are willing to take up more ambitious challenges than those arising in the native context of the Web in terms of anthropological complexity, addressing meta-semantic problems of flexible, pluralist or evolutionary ontologies, sources heterogeneity, hermeneutic and rhetoric dimensions.

A key goal of the workshop, focusing on research issues related to pre-modern scientific texts, was to emphasize, through precise projects and up-to-date investigation in Digital Humanities, the benefit of a multidisciplinary research to create interoperable semantic data and reason on them. One of the main interests of the very topic of pre-modern historical data management lies in *historical semantics*, and the opportunity to jointly consider how to identify and express lexical, theoretical and material evolutions. Dealing with historical texts, a major problem is indeed to handle the discrepancy of the historical terminology compared to the modern one, and, in the case of massive, diachronic data, to take into account the contextual and theoretical meaning of words and sentences and their semantics.

Three papers are interconnected to the ZOOMATHIA project. They develop three problematics: extracting knowledge from literary data, linking historical data with Web available information, and assessing theoretical conflicts in the zoological tradition. [1] addresses the problem of extracting zoological knowledge in a text using

---

<sup>1</sup> <http://www.cepam.cnrs.fr/spip.php?rubrique229>.

Natural Language Processing (NLP) methods and to publish it as an RDF dataset. [2] focuses on the linking of historical documentation to a reference taxonomical database (TAXREF) and presents a SKOS thesaurus enabling multi-disciplinary studies and approaches. [3] intends to give a general outline of a method combining argumentation theory, Semantic Web languages and techniques to formalize theoretical controversies in scientific texts in an argumentation framework.

Two papers also deal with Medieval knowledge. [4] deals with historical semantics and the epistemological problems arisen by the online *Dictionary of Medieval Scientific French*, and considers how Semantic Web can help to represent and save a semantic complexity and evolution. [5], related to the BIBLISSIMA (*Bibliotheca Bibliothecarum*) project, presents a prototype using open source solutions developed to index and allow complex searches on iconographic databases.

[6], related to the BIBLIMOS project, addresses the problem of digital processing of ancient Arabic manuscripts and present a semantic virtual infrastructure to operate on distributed and heterogeneous sources of digitized manuscripts. [7] reflects on the archaeological knowledge modeling and the methodological issues involved in the description of artefacts, suggesting design perspectives for computer models and tools to address human semiotics.

[8] presents the HPST interdisciplinary project on history and philosophy of science and technology and focuses on the epistemological issues raised by the development of new tools based on Semantic Web and used in historical research. Paper [9] addresses the problem of automatically disambiguating authors' mentions in a corpus of French literary criticism and propose a method based on named-entity linking.

Arnaud Zucker  
Isabelle Draelants  
Catherine Faron Zucker  
Alexandre Monnin