Preface

These Proceedings edit the papers accepted for publication and presentation at the second International Workshop on Semantic Web for Scientific Heritage (SW4SH 2016) held in conjunction with the 13th ESWC Conference on May 30, in Heraklion, Creta, Greece. 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, history and transmission of ideas.

This encounter takes place within the general context of Digital Humanities, a research area crossing Humanities and Computer Science which is gaining an ever-increasing momentum. Beyond digitalization projects, many classical scholars feel the necessity to get involved in a real conceptualization and ergonomic reflection on the scientific value of technologies based on semantic web models. The scholarly literature and documentation from Antiquity to Modern Times could be fruitfully reassessed by means of reasoning on large-scale data. Many classical scholars are now fully aware of the qualitative benefit of digital production and the formalization of data. They increasingly 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. The Semantic Web has indeed already proved to be helpful in enhancing interpretation practices and comprehension, and in discovering new knowledge. A wide range of initiatives in Digital Humanities emerge in a more or less institutional way: this workshop aims at supporting the impulse towards a closer cooperation and conceptual partnership between scholars interested in History of Science, to shape new fields of research and cross-inquiries.

The four workshop organizers belong to the Zoomathia[1] international research network funded by the French National Scientific Research Center (CNRS). The network gathers French, Italian, German and English scholars in order to study the formation and transmission of ancient zoological knowledge across historical times, combining an historical, literary and epistemological approach with the creation of open knowledge sources and the publishing of linked data

---

[1] http://www.cepam.cnrs.fr/zoomathia/

on classical zoology. All the program committee members share the same spirit of interdisciplinary synergy described above.

Like was the first, this second edition of SW4SH was also planned as an opportunity to introduce the networks activity, to enlarge it with interested participants of the workshop, and to benefit from the results of related research projects. SW4SH aims at providing a forum for discussion of the methodological approaches specific to annotate scientific texts (in every sense including disciplines such as history, architecture, or rhetoric). It wishes to support a collaborative reflection on possible guidelines or specific models for building historical ontologies. In this workshop focusing on research issues related to pre-modern scientific texts, a key goal is to emphasize the benefit of a multidisciplinary research to create interoperable semantic data and to reason on them. One of the main issues of the very topic of pre-modern historical data management lies in historical semantics, i.e. in the opportunity to consider together how to identify and express lexical, theoretical and material evolutions. Dealing with historical texts, a major problem is indeed to handle the discrepancy of an 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.

Thirteen papers were submitted; the committee accepted eight of them and invited a keynote speaker. In this second workshops edition, all papers, except one, come from contributors external to the Zoomathia network. Papers (LANA *et al.* : 7-18) and (GROS : 19-24) deal with geographical data. Paper (LANA *et al.* : 7-18) describes an ambitious project on building a thesaurus of ancient Latin geographical knowledge, using GIS technologies, and taking into account a wide range of methodological aspects required when searching on ancient documentation, and specifically in the field of geography, which often results in disseminated and loosely scientific information. Paper (GROS : 19-24) presents a united toponymical database called Ata, based on the archaeological corpora of the British School at Athens which aims at providing a common open-access geo-referencing tool. Papers (WEINGART, GIOVANNETTI : 25-36) and (KHAN *et al.* : 37-46) hit on extensions of the lemon model for complex thesaurus. Paper (WEINGART, GIOVANNETTI : 25-36) comments the relevancy of the lemon model for constructing a multilingual and multialphabetical lexicon of Old Occitan medico-botanical terminology; the DITMOA project also addresses a linguistic challenge (since the terminology appears in Latin, Hebrew and Arabic scripts), dealing with polysemy, historical semantics, complex multialphabetical lemmatization with graphic and grapho-phonetic variants, and offering interesting epistemological issues. Testing two ancient Greek and medieval English corpora, paper (KHAN *et al.* : 37-46) offers another insight into the potential of the lemon model and aims at expressing how scientific terminologies evolve, focusing on two kinds of problems in ontology engineering: polysemy and diachronic evolution. Papers (COECKELBERGS, VAN HOOLAND : 47-52 ; FABIO VALSECCHI *et al.* : 53-58 ; FARON-ZUCKER *et al.* : 59-68 ; GUILLON : 69-76) are related to semantic annotation. Paper (COECKELBERGS, VAN HOOLAND : 47-

52) relies on topic modelling techniques for annotating the Hebrew Bible, known for having lots of ambiguous lexemes and textual variants. Paper (VALSECCHI *et al.* : 53-58) recommends to join text transcribing and annotating processes; it shows the benefit of this approach when applied to the corpus of the works and letters of the astronomer C. Clavius (16-17 cent.). Papers (COECKELBERGS, VAN HOOLAND : 47-52) and (VALSECCHI *et al.* : 53-58) both argue for stand-off markup. Paper (FARON-ZUCKER *et al.* : 59-68) deals with automatic annotation of Ancient text and presents a methodology developed within the Zoomathia project, for extracting zoological knowledge from Plinys Historia Naturalis. Paper (Guillon : 69-76) outlines the project of a semantic wiki on contemporary research about analytic philosophy, in order to structure critical and historical annotation on theories and argumentation.

We wish that you will find this workshop programme exciting and stimulating!

Isabelle Draelants
Catherine Faron Zucker
Alexandre Monnin
Arnaud Zucker