Editorial: Special Issue on Semantic eScience: Methods, tools and applications

Daniel Garijo [a], Natalia Villanueva-Rosales [b] and Tomi Kauppinen [c]

[a] *Information Sciences Institute, University of Southern California, USA. E-mail: dgarijo@isi.edu*
[b] *Department of Computer Science, University of Texas at El Paso, USA. E-mail: nvillanuevarosales@utep.edu*
[c] *Aalto University, Finland. E-mail: tomi.kauppinen@aalto.fi*

Abstract. Openly shared, available, and accessible scientific resources facilitate tackling grand challenges that our society, organizations, communities and individuals are facing today. Pandemics, climate change, environmental modeling, genomics, or space exploration all create open research questions for which Artificial Intelligence – and Semantic Web in particular – have a unique potential to accelerate scientific discoveries. In this editorial, we introduce a special issue on Semantic eScience methods, tools and applications for the Semantic Web Journal and outline five challenges for Semantic eScience in the years to come.

1. Introduction

An increasing number of datasets, software tools and methods reported in scientific publications are shared today to support scientific results and ease their reproducibility. Open initiatives such as Zenodo[1] or the European Open Science Cloud[2] contribute significantly to finding and accessing these scientific resources. However, reusing and understanding datasets, software and methods is still challenging given the effort required to address interoperability issues; and capturing constraints, assumptions, limitations, examples and internal variables associated with data and code.

Computational experiments are becoming more complex due to the heterogeneity of models involved, the amount and type of data required, computing resources needed, and interdisciplinary collaborations. Artificial Intelligence and Semantic Web technologies can uniquely contribute to address these challenges by providing means to create meaningful and well-defined descriptions for scientific resources and en-

abling the automation of tedious tasks that are currently performed manually.

Making scientific resources accessible, semantically described, and interlinked would create an invaluable network for scientific research, increasing transparency of Science, enabling reproducibility and assisting researchers in communicating scientific outcomes to current and new generations. Researchers would be able to build on previous scientific experiments and its results more efficiently; and focus their efforts in tackling new challenges instead of manipulating data in the right format for their analysis. Educators may leverage semantically annotated resources for new students to easily access, understand and reuse experiments openly available worldwide.

The goal of this special issue is to emphasize these benefits by collecting the latest research solutions to bridge the gap between existing scientific communication methods and the vision of a reproducible and accountable open science.

2. Special Issue Overview

This special issue was proposed after a series of community workshops on Semantic Web conferences

---

[1] https://zenodo.org/
[2] https://www.eosc-portal.eu/about/eosc

[1–3] that attracted dozens of researchers working on applying Semantic Web techniques to help describe and capture the context of different domains, including the internet of things, environmental modeling, genomics, neuroscience and scholarly communication.

This domain diversity was reflected on the submissions received in this special issue, which addressed problems such as capturing the context and provenance of Machine Learning pipelines, linking and automatically describing virtual containers; or annotating the context of scripts and their provenance to find and help reproduce scientific experiments. After a careful reviewing process, two papers were accepted, which focus on discoverability and findability of scientific products. In [4] the authors ease exploration of scientific literature by grouping similar papers according to their topic; while in [5] the authors assist other researchers in describing and summarizing their scientific experiments based on publicly available metadata.

3. Five Challenges ahead for Semantic eScience

We believe that the interest from the Semantic Web community on Semantic eScience will continue in the few next years, as there are many active and future areas of research. We summarize the five areas we consider most important below:

1) *Automated model and data integration*: The need for increasingly complex models to understand real-world phenomena requires the retrieval and manipulation of large heterogeneous and decoupled data, easy model integration and interaction, and domain expertise for the interpretation of model results. Scientific workflows facilitate the representation and execution of models. However, annotating and manipulating data; as well as validating, interpreting and communicating the results are still manual tasks. Approaches to automate these manual tasks using Semantic Web technologies can significantly accelerate the generation, and reproducibility of scientific results.

2) *Automated capture of scientific research context*: It is unfeasible to manually explore and analyze the large amount of data generated and needed by scientific research. Capturing and representing the context of scientific research (hypotheses, assumptions, limitations, etc.) would enable intelligent systems to assist when comparing and understanding existing work; as well as to help answering a specific research question.

3) *Improving collaboration between scientists and intelligent systems*: New scientific discoveries require collaboration between research teams across disciplines, organizations, and geographical boundaries that require integrating their knowledge, perspectives, and resources. Intelligent systems can leverage rich semantic descriptions to facilitate communication, knowledge exchange and identify common research approaches, resources and opportunities.

4) *Towards AI-based discoveries*: Knowledge graphs capture an increasing amount of facts about human knowledge and the nature of the world (e.g., Physics and Math equations). Intelligent systems should be able to leverage this knowledge and combine it with the latest scientific results to identify potential research lines and assist scientists with automated hypothesis analysis. Similarly, intelligent systems should be able to validate results and foster trust over existing publications. Achieving these tasks would accelerate scientific discovery significantly.

5) *Better abstraction and generalization of scientific results*: Scientific research is often communicated at different levels of detail, which are hard to conflate automatically in a unique semantic representation. For example, a Geosciences paper on climate modeling would emphasize the academic contributions of the analysis, while its code repository would detail how to set up and execute its associated software; and its computational notebook would emphasize the meaning behind the visualization of results. Learning to conflate these different representations and levels of detail in a machine readable manner would help automatically relate scientific results; and may inform intelligent systems to accurately communicate scientific outcomes to different audiences.

References

[1] D. Garijo, W.R. van Hage, T. Kauppinen, T. Kuhn and J. Zhao (eds), Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci), in: *First Workshop on Enabling Open Semantic Science (SemSci)*, CEUR Workshop Proceedings, Aachen, 2017. ISSN 1613-0073. http://ceur-ws.org/Vol-1931/.

[2] D. Garijo, N. Villanueva-Rosales, T. Kuhn, T. Kauppinen and M. Dumontier (eds), Proceedings of the Second Workshop on Enabling Open Semantic Science (SemSci), in: *Second Workshop on Enabling Open Semantic Science (SemSci)*, CEUR Workshop Proceedings, Aachen, 2018.

[3] C. Badenes, R. Denaux, M. De Vos, D. Garijo, J.M. Gomez-Perez, A. Lawrynowicz, P. Lisena, R. Palma, R. Troncy and D. Vila, K-CAP2017 Satellites: Workshops and Tutorials, in: *Proceedings of the Knowledge Capture Conference*, K-CAP 2017, Association for Computing Machinery, New York, USA, 2017. doi:10.1145/3148011.3188410.

[4] C. Badenes-Olmedo, J.L. Redondo-García and O. Corcho, Large-scale semantic exploration of scientific literature using topic-based hashing algorithms, *Semantic Web* (2020), 1–16. doi:10.3233/SW-200373.

[5] A. Gaignard, H. Skaf-Molli and K. Belhajjame, Findable and reusable workflow data products: A genomic workflow case study, *Semantic Web* (2020), 1–13. doi:10.3233/SW-200374.