2nd Workshop on Enabling Open Semantic Science (SemSci 2018) —

Workshop Type: established

Abstract

In the past few years, a push for open reproducible research has led to a proliferation of community efforts for publishing datasets, software and methods, described in scientific publications. These efforts underpin research outcomes much more explicitly accessible. However, the actual time and effort required to achieve this new form of scientific communication remains a key barrier to reproducibility. Furthermore, scientific experiments are becoming increasingly complex, and ensuring that research outcomes become understandable, interpretable, reusable and reproducible is still a challenge. The goal of this workshop is to incentivise practical solutions and fundamental thinking to bridge the gap between existing scientific communication methods and the vision of a reproducible and accountable open science. Semantic Web technologies provide a promising means for achieving this goal, enabling more transparent and well-defined descriptions for all scientific objects required for this envisioned form of science and communication. We are particularly interested in four kinds of contributions: 1) novel approaches to analyze scientific publications in order to explicitly describe the relationship between their methods and research outputs; 2) novel approaches that use the research outputs of a scientific publication to facilitate its understanding and reuse (e.g., by generating explanations of results, interactive visualizations or linking datasets and methods); 3) novel approaches that help comparing and relating software, datasets and methods used in different publications; and (4) novel approaches to apply Semantic Web and Linked Data techniques to scientific workflows used in research.

Topics of Interest

Topics for submissions include, but are not limited to:

· Tools, methods and use cases for linking existing papers to their research products: data, software, methods and execution traces.

· New methods for helping linking scientific papers to other relevant papers (e.g., papers that use similar approaches, similar methods, common software, common data, etc.)

· New methods for helping visualizing and presenting scientific information to scientists (e.g., provenance-based visualizations, summaries, presenting results at different levels of granularity, etc.)

· New approaches for extracting the specific steps used in a method described expressed in a scientific paper.

· New methods for automatically generating explanations of scientific results.

· New approaches for comparing methods, protocols and methodologies expressed in scientific papers.

· New methods to highlight the differences between execution runs of a scientific experiment (based on their configuration, performance, results, etc.)

· Tools and methods for discovering data and software used in similar publications or to

address similar problems.

- Vocabularies and ontologies that help relate and describe software, data, methods and provenance used in a scientific publication.
- Vocabularies and ontologies that help capturing and presenting experiment information to scientists.
- Automatic annotation of scientific research
- Provenance, quality, privacy and trust of scientific information
- Novel visualizations of scientific data
- Novel approaches to apply Linked Data and Semantic Web techniques to scientific workflows

Continuation

Open Science is highly diverse and applicable to a wide variety of domain sciences, ranging from geology and climate sciences to chemistry or genomics. The Semantic Web community has researched on several of the challenges of interest proposed in this workshop, but many are far from being addressed. We believe that these challenges make the topic of the workshop relevant for the conference and give the workshop a unique identity (focused on domain sciences). We are confident in attracting submissions and new ideas to discuss. In addition, the interest from the community is currently high. After the previous edition of the SemSci workshop, the second edition of SciKnow (http://sciknow.github.io/sciknow2017) was celebrated co-located with K-CAP17, with more than 20 participants. A special issue in the Semantic Web Journal is currently being discussed, so as to crystallize the most mature contributions submitted to the previous workshops.

Past Workshops

SemSci2018 will be the second workshop edition. Last workshop, SemSci 2017, received 12 submissions, with a total of 10 papers accepted. The workshop was attended by more than 45 participants, who actively participated in the discussions. It must be noted that SemSci has been heavily built on the success of Linked Science (LISC) workshop series (from year 2011 to 2015) at ISWC. Past LISC workshops raised significant interest and discussions, thus we would like to continue building the community now with wider topics on Semantic Science.

Workshop Format

SemSci2018 will combine paper and demo presentations with break-out sessions. Accepted papers will be published at CEUR workshop series. SemSci2018 explicitly encourages alternative and enhanced submission formats such as HTML or communicative online materials. Authors who are preparing such a submission should contact the workshop organizers in advance to make sure we can accommodate for them in the submission and review process. Papers submitted to the workshop are also encouraged to share their research products online, assigning a DOI when necessary. Workshop organizers will provide pointers and guidelines for this purpose, based on the ISWC Resources Track submission guidelines.

In addition to regular research papers, we will also invite short papers of up to 4 pages on tools and demos. Participants presenting such a paper will be expected to demonstrate the tool in the context of Semantic Science. We look both for innovative software solutions as well as tool that have proven useful in enabling Open Semantic Science. Papers will be reviewed based on potential impact of the tool for Open Semantic Science, usability, and documentation.
Workshop participants will be expected to contribute informally to discussions as well as bring up their own topics of interest. In addition, we expect to have break-out sessions to identify and discuss new topics and their related challenges.

Chairs

Short bios of the chairs:
Daniel Garijo (https://w3id.org/people/dgarijo) is a postdoctoral researcher at the Information Sciences Institute of the University of Southern California. He got his PhD at the Ontology Engineering Group of the Universidad Politecnica de Madrid. His research is focused on using Semantic Web and Linked Data to facilitate the reuse and understanding of scientific workflows.
Tobias Kuhn (http://www.tkuhn.org/) is an Assistant Professor in the Web & Media group of the Computer Science department of the VU University Amsterdam. He got his PhD at the Institute of Computational Linguistics of the University of Zurich in 2010. After that, he was a lecturer and researcher at the University of Malta, postdoctoral associate at Yale University, and a postdoc at ETH Zurich, with stays as visiting researcher at the University of Chile and Stanford University in between. His research interests span diverse fields including knowledge representation, user interfaces, controlled natural languages, social systems, bioinformatics, and scholarly communication. His recent work has focused on the approach of nanopublications and how cryptographic methods and provenance modelling can support trust and reliability.
Tomi Kauppinen (http://kauppinen.net/tomi) is a project leader and docent at the Aalto University School of Science in Finland and a Privatdozent at the University of Muenster in Germany. He holds a habilitation (2014) in geoinformatics from the University of Muenster and a Ph.D. (2010) in media technology from the Aalto University. From April 2014 to September 2014 he was appointed as the Cognitive Systems Substitute Professor at the University of Bremen in Germany. A central theme in his work and teaching is data science and information visualization applied to spatio-temporal phenomena, and supporting understanding of related cognitive processes. He has co-organised international workshops on linked data, information visualization, spatial thinking and linked science. He serves also as a co-editor for the research topic Culturomics: Interdisciplinary Path Towards Quantitative Study of Human Culture of the Frontiers in Physics.
Natalia Villanueva Rosales (http://natalia-villanueva.com) is an Assistant Professor in Computer Science and co-Investigator in the Cyber-ShARE Center of Excellence at the University of Texas at El Paso. She holds a Ph.D. in Computer Science from Carleton University. Her long-term research goal is to improve the efficiency and effectiveness of the discovery, integration, and reuse of scientific data and research. Her current approaches enables knowledge negotiation in interdisciplinary collaborations, where a key challenge is the integration of data and reusability of models initially created for a different purpose. Her recent work has been applied in the areas of sustainability of water resources and Smart

Cities.
Michel Dumontier (https://about.me/micheldumontier) is a Distinguished Professor of Data Science at Maastricht University. He is an expert in computational methods for knowledge discovery. Dr. Dumontier was previously a professor at the Stanford Center for Biomedical Informatics Research at Stanford University and a professor of Bioinformatics at Carleton University. His research focuses on the development of computational methods for scalable integration and reproducible analysis of FAIR (Findable, Accessible, Interoperable and Reusable) data across scales - from molecules, tissues, organs, individuals, populations to the environment. His group combines semantic web technologies with effective indexing, machine learning and network analysis towards accelerating discovery science, improving health and well being, and strengthening communities.

Program Committee

Members who have already confirmed their participation in the program committee are marked with an *:

- Marieke van Erp, VU University Amsterdam *****
- Carsten Keßler, Aalborg University Copenhagen **(pending)**
- Yolanda Gil, University of Southern California, USA *
- Oscar Corcho, Universidad Politécnica de Madrid, Spain *****
- Idafen Santana Pérez, Universidad Politécnica de Madrid, Spain *****
- Mark Wilkinson Universidad Politécnica de Madrid, Spain *****
- Craig A. Knoblock, University of Southern California, USA *****
- Gully Burns, University of Southern California, USA *
- Khalid Belhajjame, University Paris-Dauphine *****
- Amrapali Zaveri, Maastricht University, Netherland *****
- Alasdair Gray, Heriot-Watt University, UK *****
- Paul Groth, Elsevier Labs, the Netherlands *****
- Anita de Waard, Elsevier Labs *****
- Jeff Pan, University of Aberdeen, UK *****
- Jun Zhao, Oxford University, UK *****
- Willem Robert van Hage, Netherlands eScience Center **(pending)**
- Alexander Garcia Castro, Universidad Politécnica de Madrid, Spain *****
- Tim Clark, Harvard University *
- Derek Sleeman, University of Aberdeen **(pending)**
- Olga Ximena Giraldo, Universidad Politecnica de Madrid *****
- Silvio Peroni, University of Bologna *****
- Richard Boyce, University of Pittsburgh *****

* means the PC member has confirmed

Length

SemSci2018 will be a full day workshop of peer-reviewed presentations and demos selected among author submissions, along with breakout group discussions.