

Title: Semantics in Big Data Management

Chairs:

Paolo Ceravolo: (<https://orcid.org/0000-0002-4519-0173>), Universita' degli Studi di Milano, Italy

Paolo.Ceravolo@unimi.it

Florence Sedes: Toulouse Institute of Computer Science Research, France

Florence.Sedes@irit.fr

Estimated length:

We plan to organize a workshop of 1/2 day length. We, however, would like to remark that we are in overlapping with another workshop proposal of Paolo Ceravolo (SIMPDA).

If feasible, it could be nice to avoid overlappings.

Aims:

The complexity of Big Data technologies and the variety of competencies required to design applications relying on them has emphasized the relevance of systems for managing and documenting Big Data architectures. Documentation, reconfiguration, and verification are for example crucial tasks for a solid design of technological solutions but are not easily supported in the current landscape of Big Data technologies. Rethinking data and information management in the context of Big Data technologies is then a primary goal for future research.

Methods, principles, and perspectives developed by the Data Semantics community can significantly contribute to this goal. Solutions for integrating and querying schema-less data, for example, have received much attention. Standards for metadata management have been proposed to improve data integration among silos and to make data more discoverable and accessible through heterogeneous infrastructures. A further level of application of Data Semantics principles into Big Data technologies involves Representing Processes, i.e. representing the entire pipeline of technologies connected to achieve a specific solution and make this representation shareable and verifiable to support a mature implementation of the Big Data production cycle.

Following the manifesto paper "Big Data Semantics" (J. Data Semantics 7(2), 2018), which identified the important and yet unsolved issues of managing semantics of big data, we continue research on this topic.

Topics of interest for submission include, but are not limited to:

Big Data Management

Metadata Management

Big Data Persistence and Preservation

Big Data Quality and Provenance Control

Big Data Storage and Retrieval

Big Data Integration Architectures and Techniques

- Data Source Discovery
- Big Data Profiling and Semantics Discovery
- Querying Heterogeneous Big Data Repositories
- Caching and Materializing Query Results

Quality of Big Data Services

- Big Data Service Performance Evaluation
- Big Data Service Reliability and Availability
- Reproducibility of Big Data Services
- Verifiability of Big Data Services
- Assurance in Big Data Services

Big Data Visualization

- Real Time Visualisation
- Visualization Analytics for Big Data
- Big Social Media Mining

Big Data Security and Privacy

- Big Data System Security and Integrity
- Big Data Information Security
- Privacy-Preserving Big Data Analytics
- Usable Security and Privacy for Big Data

Performance of Big Data Architectures

- Query Optimization
- Optimal Selection of Analytics
- Physical Structures

Tentative list of PC members:

This workshop was born by the initiative of the IFIP WG 2.6. In 2018 it has been organized in conjunction with the IFIP WCC 2018. We will invite as PC members researchers and practitioners participating in the previous edition of this workshop (<http://wcc2018.org/sembdm>).