

SemData Manifesto

The growth in volume and diversity of data continues to create new challenges for data management. RDF and linked data practices are emerging as the main model for publishing structured data. In this new culture of publishing and consuming data, we find increasing heterogeneity in content and schema and a consequently greater need for ad hoc integration.

The proliferation of RDF data has to date resulted in the creation of numerous specialized RDF databases as well as in RDF-oriented extensions to existing general purpose databases. However, the question shall be raised: is management of RDF a discipline distinct from database research?

The history of computer science and of software development culture at large contains numerous examples of factionalization and repetition of effort, i.e. the Not Invented Here (NIH) syndrome. On the other hand, when a general purpose solution does not work well in a specific use case, one is fully justified in making a case-specific solution.

To date, the RDF and mainstream database communities have been relatively separate, with limited cross-over between conferences of either camp. In order to realize RDF's promise of becoming the lingua franca of data integration and structured publishing, a qualitative threshold needs to be crossed. To bring this about, the cutting edge of mainstream database, semantic technologies and reasoning need to come together in a spirit of synergy.

The SemData initiative aims at building a platform for this in the form of a series of workshops and events. This will provide a venue for publishing research at the crossroads of RDF, database and reasoning. We believe that existing conferences and publications fully cover each area separately, yet we also see that these domains in isolation are not enough for addressing the challenges that arise with the Web-scale integration, online serving and analysis of structured content.

The SemData initiative seeks to bring together researchers and practitioners from the following domains in order to foster cooperation:

- Semantic Web: RDF and Linked Data
- Web technology and cloud computing
- Database research
- Information retrieval / information extraction
- Knowledge management and reasoning
- Visualization
- Machine learning
- Vertical markets

There is much promising work in each area separately. The SemData initiative believes that the functions of core database, inference engine and information retrieval server are related and need to be integrated in order to deliver the full promise of semantic data management. For this integration to become a matter of course, much as full text is a matter of course in most RDBMS's, more work is needed.

The SemData initiative aims at providing a forum for publishing high quality original work pertaining to these topics. We specially invite submissions concerning:

- Integration of RDF and IR techniques
- Advances in RDF query languages, enhancing SPARQL with inference
- Scalable reasoning close to the data
- Use cases and experiences involving RDF databases, from online to analytics
- Scalability of RDF storage
- Core database research for RDF including index structures, storage density and query optimization
- RDF storage and query evaluation and benchmarking