



ERCIM "ALAIN BENSOUSSAN"
FELLOWSHIP PROGRAMME



Scientific Report

First name / Family name	Fabrizio Orlandi
Nationality	Italy
Name of the <i>Host Organisation</i>	Fraunhofer Institute for Applied Information Technology
First Name / family name of the <i>Scientific Coordinator</i>	Sören Auer
Period of the fellowship	01/05/2014 to 30/04/2015

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

A. *Linked Data for SMEs (ongoing)*

Linked Data refers to a data representation methodology that is machine readable and semantically enriched, and hence apt for semantic interlinking and automated processing of semantic queries. The goal of our research effort is to advance the current state of the art for exploration and consumption tools for Linked Data. In particular by constructing a useful integrated tool chain while focusing on the non-expert end users, and thus enabling them to make their first steps in this field. Non-expert end-users are regarded as those that are new to the realm of Linked Data but do possess sufficient knowledge of data science in their respective domain.

Essentially, our approach consists in a number of tools, integrated in a workbench, that are accessed from a set of prefabricated minimalistic but holistic workflows. Thus, the tool chain that we developed covers both data provisioning and data consumption, including publishing and importing from external data sources. The pivotal point is the Linked Data format. For the data provisioning, there are tools for

data renovation i.e., conversion from various formats towards Linked Data (in this case, specifically, the conversion to the RDF format); for the data consumption side, there are tools for querying, analytics, and visualizations. Conceptual and technical simplicity are ranked higher than feature completeness. For instance, the data formats handled are confined to “semi-structured data”, i.e., data from relational databases, CSV files and similar, as opposed to “unstructured data”, which is the term commonly used to refer to natural-language data. The reason is that the respective complexities differ vastly, and Natural Language Processing (NLP) is an active research field in its own right, making it particularly unsuitable for toolchains that are supposed to subsume results rather than to support research.

In particular, during the fellowship, these efforts were supported by other two Ph.D. students, under the fellow’s supervision, who focused on an automated visualization approach for Linked Data and a framework for transformation and consumption of RDF data respectively (described in the following subsections).

B. Linked Data Publication and Consumption

The Linked Data Publication and Consumption Framework aims to assist SMEs, data publishers and consumers in analysing and interlinking public sector information with enterprise data. The novel implemented workflow starts with dataset exploration. Here a user can explore existing open datasets, both with regards to their content and also with regards to the underlying schema. The user then has two options, either directly executing a SPARQL query using a SPARQL Query Tool or use the Query Builder Tool to generate the query. Finally, the user can proceed to convert the result set in a number of different formats and export results. This conversion allows users to import data from linked open datasets into their native system. The results can also be exploited further through visualizations, analytics and data mining services.

The implementation of this workflow in a comprehensive framework has been supported by a Ph.D. student, under the fellow’s supervision, in the context of an EU FP7 research project.

C. Linked Data Visualization

The goal of this research effort is to provide a largely automatic visualization workflow that enables SMEs (and Linked Data inexperienced users) to visualize data in different formats and modalities. In order to achieve this goal, a generic web application has been developed based on state-of-the-art Linked Data approaches to allow for visualizing different categories of data, e.g. statistical, geographical, temporal, arbitrary data, and a largely automatic visualization workflow for matching and binding data to visualizations. The implemented Visualization tool consists of two main components: (i) The Explore and Select Data component and (ii) the Visualization component. The Explore and Select Data component allows users to pre-select the data to be processed and provides a concise preview of the data through a tabular representation. The input data formats supported are RDF

(any serialization and vocabulary) and CSV. The selection performed with this component is used as an input of the Visualization component. The Visualization component is responsible for the creation of various plots, charts and maps and offers opportunities to customize visualization options or save and share the graphical results. The visualization is powered by a recommendation algorithm which aims at suggesting the most suitable visualization type according to different features of the input data.

The implementation of this software prototype tool has been supported by a Ph.D. student, under the fellow's supervision, in the context of an EU FP7 research project.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

- Thellmann, K., Orlandi, F., & Auer, S. (2014). LinDA-Visualising and Exploring Linked Data. In proceedings of *SEMANTiCS 2014*.
- Orlandi, F. (2014). Profiling user interests on the social semantic web. *ACM SIGWEB Newsletter*, (Spring), 3.
- Tygel, A. F., Attard, J., Orlandi, F., Campos, M. L. M., & Auer, S. (2015). “How much?” Is Not Enough-An Analysis of Open Budget Initiatives. *arXiv Preprint arXiv:1504.01563*. (Submitted to WebScience 2015).
- Endris, K. M., Faisal, S., Orlandi, F., Auer, S., & Scerri, S. (2015). Interest-based RDF Update Propagation. *arXiv Preprint arXiv:1505.07130*. (Submitted to ISWC 2015, under review).
- Thellmann, K., Galkin, M., Orlandi, F., & Auer, S. (2015). LinkDaViz – Automatic Binding of Linked Data to Visualizations. (Submitted to ISWC 2015, under review).
- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). Open Government Data - A Systematic Review of Open Government Data Initiatives. *Government Information Quarterly Journal*. (Under review).
- Attard, J., Orlandi, F., & Auer, S. (2015). Softening RDF Data to Enhance Data Re-Use. (Submitted to SEMANTiCS 2015, under review).

III – ATTENDED SEMINARS, WORKSHOPS, CONFERENCES

- SEMANTiCS Conference, September 2014, Leipzig, Germany (poster and demo paper presentation).
- In addition, during the course of the fellowship, the fellow was involved in as many as 4 project proposals targeting DFG (German) and H2020 (EU) calls, one of which was accepted. The fellow has been involved in several European research projects (1 FP7: LinDA - 2 H2020: ODINE and OpenBudgets - 1 MSCA ITN: WDAqua) and participated to many of their meetings. As part of the experience in the EIS group at Fraunhofer IAIS, the fellow managed teaching for the Lab (practical group work), Seminar (presentations prepared by students) and

Exercises modules offered by the group at the affiliated Computer Science department of the University of Bonn.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

During the course of the fellowship, the fellow visited the OEG¹, Ontology Engineering Group, based at the Computer Science School at Universidad Politécnica de Madrid (UPM), Spain. Headed by Asunción Gómez Pérez and Oscar Corcho, it ranks eighth among the two hundred research groups from UPM and it is widely recognised in Europe in the areas of Ontology Engineering, Semantic Infrastructure, Linked Data, and Data Integration. The group is part of the SPARCIM research network.

Over the course of two weeks, the fellow exchanged research ideas and objectives with a large number of researchers in this team, learning about many diverse topics and new methods that were previously not considered for research contributions. The main points of contact identified for collaboration are especially focused on the fields of Open Government Data and Linked Data Quality. Concrete plans for future additional research exchanges and research project proposals have been made.

¹ <http://www.oeg-upm.net/>