

European and International Statistical Linked Dataspaces and Analysis

Contributors

* Sarven Capadisli <info@csarven.ca> is an open Web technologist primarily researching and developing Statistical Linked Dataspaces and Analysis (University of Leipzig) - *Corresponding author*. <http://csarven.ca/>

* Dr. Sören Auer <auer@cs.uni-bonn.de> (University of Bonn): <http://www.iai.uni-bonn.de/~auer/>

* Dr. Reinhard Riedl <reinhard.riedl@bfh.ch> (Bern University of Applied Sciences): <http://www.e-government.bfh.ch/>

Type of the presentation proposed

Research contribution

Summary

Linked Data principles are increasingly employed to publish high-fidelity, heterogeneous statistical datasets in a distributed way. Currently, there exists no simple way for researchers, journalists and interested people to compare statistical data retrieved from different data stores on the Web. The data gathered from European and international statistical agencies like European Central Bank, Eurostat, IMF, OECD, and the World Bank, to name a few, are re-deployed as Linked Data, where common statistical concepts across these agencies are interlinked with one another. The Statistical Analysis application and service presented here utilizes federated queries to gather statistical data from distributed data stores. Statistical computing software is employed to perform statistical analyses and visualizations. Generated analysis is stored for future research, and as a result, distributed linked statistical data can be more easily explored and analysed.

Description

Statistical data artifacts and the analyses conducted on the data are fundamental to testing scientific theories about our societies. As statistics are often used to add credibility to an argument or advice, they influence the decisions we make. The decisions are, however, complex beings on their own with multiple variables based on facts, cognitive processes, social demands, and maybe even factors that are unknown to us. Regardless of uncontrollable forces, in order for the society to tract and learn from its own vast knowledge about events and things, it needs to be able to gather statistical information from heterogeneous and distributed sources.

This is to uncover insights, make predictions, or build smarter systems that the society needs to progress. This brings us to the core of our research challenge; how do we reliably acquire statistical data in a uniform way and conduct well-formed analyses that is accessible to different types of data consumers and users?

We present Statistical Linked Data Analyses (<http://stats.270a.info/>) towards this challenge with our contributions. In a nutshell, it takes advantage of Linked Data design principles that are widely accepted as a way to publish and consume data without central coordination on the Web (<http://270a.info/>) The work herein offers a Web based user-interface for researchers, journalists, or interested people to compare statistical data from different sources against each other without having any knowledge of the technology underneath or the expertise to develop themselves. The service, which we built, proceeds with running decentralized (federated) structured queries to retrieve data from various endpoints, runs an analysis on the data, and provides the analysis back to the user. For future research, analysis is stored so that it can be searched for and reused.

This brings us to an outlook for Linked Statistical Data Analyses. The reuse of Linked analyses artifacts as well as the approach to collect data from different sources can help us build smarter systems. It can be employed in fact-checking scenarios as well as uncovering decision-making processes, where knowledge from different sources is put to their potential use when combined.

We hope that the research presented herein serves as the feasibility and potentials of a decentralized Linked Statistical Dataspaces with an analysis platform, and encourages European and international statistical agencies to publish their data using the Linked Data design principles for better collaboration.