Self-describing fiscal data

How can you ease reuse of fiscal data and avoid misinterpretation?

Jindřich Mynarz, Jakub Klímek, Jan Kučera

Faculty of Informatics and Statistics
University of Economics, Prague
{jindrich.mynarz, jakub.klimek, jan.kucera}@vse.cz

Introduction

Fiscal data released by public sector bodies looks like impenetrable fog of accounting terms to most lay users without expertise in public finance. Relying on column labels to convey what data is about is a recipe for misinterpretation. Fiscal data tends to be poorly documented and lacking schemas that would guide its users. We propose to publish self-describing fiscal data to help resolve these issues. Machine-readable descriptions of data increase the degree to which data processing can be automated. Data descriptions can guide human reusers and improve their understanding of the described datasets. Self-describing data enable some processing without reaching for out-of-band information by reading documentation or contacting dataset maintainers. We describe 2 complementary approaches for self-describing data from the domain of public finance: data model based on the RDF Data Cube Vocabulary and Fiscal Data Package that is based on JSON descriptors.

Session description

Problem addressed by the proposed session

Fiscal data such as data about public budgets and spendings is an example of a domain that requires significant domain-specific knowledge to properly interpret the data. Moreover, fiscal data tends to be poorly documented and lacking schemas that would guide users of data. In order to resolve these issues we propose to publish self-describing fiscal data. Possible benefits of publish self-describing fiscal data include increase of degree to which fiscal data can be processed automatically and it might also help human users to better understand the data and thus to mitigate the risk of misinterpretation of the data.

Expected outcomes

The aim of the session is to identify a set of key elements that make up core of the fiscal datasets and to discuss the possible approaches to making data in this domain self-describing. We strive to find a balance between data description that gives leverage to data reusers and description that is burdensome to produce for data publishers. Two complementary approaches for self-describing data from the domain of public finance are going to be presented at the sessions: data model based on the RDF Data Cube Vocabulary standardized by the W3C and Fiscal Data Package that is based on JSON descriptors. Thus this session should also serve to facilitate feedback on these approaches under development in order to make them more suitable for real-world scenarios.

Expected outcomes of the session are:

- a set of key elements that make up core of the fiscal datasets and
- a set of recommendations for publication of self-describing fiscal data.

Intended audience

We would like to invite anyone interested in fiscal transparency and data interoperability to come to our session. Intended audience should involve, but is not limited to:

- public sector representatives, especially those responsible for fiscal transparency and publication of fiscal data;
- data reusers, especially those interested in fiscal data;
- data experts, especially those interested in semantic interoperability of data.

Acknowledgements

This session is supported by the EU ICT PSP project Share-PSI 2.0, Grant Agreement no. 621012.