

Towards a Harmonized Dataset Model

Ahmad Assaf^{1,2}, Aline Senart² and Raphaël Troncy¹

¹ EURECOM, Sophia Antipolis, France. <firstName.lastName@eurecom.fr>

² SAP Labs France. <firstName.lastName@sap.com>

Abstract. The Open Data movement triggered an unprecedented amount of data published in a wide range of domains. Governments and corporations around the world are encouraged to publish, share, use and integrate Open Data. There are many areas where we can see the value of Open Data, from transparency and self-empowerment to improving efficiency, effectiveness and decision making. The growing amount of data constitutes the need for rich metadata attached to it. This metadata enables dataset discovery, comprehension and maintenance. Data portals, which are considered to be datasets' access points, present their metadata in various models. In this paper, we propose a harmonized dataset based on the analysis of seven prominent dataset models (CKAN, DKAT, Public Open Data, Socrata, VoID, DCAT, Schema.org). We further present use cases that show the benefits of providing rich metadata to enable dataset discovery, search and spam detection.

Keywords: Dataset, Dataset Profile, Metadata, Dataset Model

1 Introduction

2 Dataset Models

3 Towards A Harmonised Model

Acknowledgments

This research has been partially funded by the European Union's 7th Framework Programme via the project Apps4EU (GA No. 325090).

References