

Big Data Management Challenges and Solutions in the Context of European Projects – Workshop Introduction

Yannis Ioannidis

“Athena” Research Center and University of Athens, Greece

yannis@{athenarc.gr,di.uoa.gr}

The main objective of this workshop¹ has been to share experiences and best practices, discuss challenges and effective solutions adopted, and investigate opportunities for collaboration among European projects (funded by various directorates of the European Commission or other European funding agencies) dealing with big data management. The projects may have ICT as their main focus or, equally well, they may have some other scientific field, industrial application, or societal challenge as their main focus, in the context of which big data issues come up.

The workshop aims at bringing together data management and database researchers and experts, as well as related user groups, designers, developers, and data practitioners. It acts as a broad forum for the exchange of the latest research results in big data management exploring new concepts, techniques, and tools. These showcase how the major big data challenges are being confronted, be they the classical high data volume, great variety, high data velocity, lack of veracity (accuracy or reliability), and difficulty in extracting value from data, or new specialized issues that possibly arise in specific environments and contexts.

More specifically, within the context of European projects, the workshop has the following concrete sub-objectives:

- Bring together active data management researchers, data scientists, and data practitioners from both the private and public sector
- Identify major challenges in big data management
- Exchange experiences and best practices in big data management
- Consider the ethical aspects and societal impact of big data technology
- Discuss the importance of world-wide initiatives such as the Research Data Alliance (<http://www.rd-alliance.org>)
- Clarify the relevance of new roles/job descriptions emerging, such as that of ‘data scientist’
- Initiate a dialogue among seemingly heterogeneous European projects that face similar data management challenges and identify potential concrete actions of collaboration between them
- Connect the data management research community with the European funding scene

Seven such projects, each one addressing one or more of the above sub-objectives has been accepted to the workshop. Four of them deal with big data challenges for particular kinds of data:

- dataCron focuses on spatiotemporal data, either at rest (static data) or in motion (streams). It takes advantage of big heterogeneous data sources to study the trajectories of moving objects and predict their future positions.
- STREAMLINE deals with data at rest and data in motion as well. It studies various techniques to improve performance of big dataflow executions and advance the state of the art in specialized functionality, such as interactive visualization and window aggregation.
- PROTEUS is the third project that focuses on both historical data and streams. It works on developing a software architecture to support online machine learning predictive analytics and real-time interactive visualization on large volumes of such data.
- MyHealthMyData focuses on the issue of data privacy. In particular, it deals with biomedical information in a network of hospitals and aims to provide the necessary technologies, e.g., blockchain, so that anonymised patient data may become available for research, while the patients remain in control of the use of their data.

Furthermore, three of these projects deal with more generic, horizontal issues that may arise in broader contexts.

- SUPERSEDE develops a big data system whose purpose is to analyse large volumes of heterogeneous, user-generated and system-generated data on the Quality of Experience that users have with software services and applications so that decisions about the evolution and adaptation of the latter may be supported.
- TOREADOR offers a Big Data Analytics-as-a-Service environment aiming at helping organizations that lack the proper big data/data science expertise declare their big data analytics goals and have the appropriate big data pipeline be generated for them, ready to use.
- BigDataEurope is somewhat similar in that it aims to develop an infrastructure offering diverse big data computational functionality that may be required in any one of the seven EC Societal Challenges (Health, Food, Energy, Transport, Climate, Social Sciences, and Security).

This uniquely different EDBT/ICDT workshop is a follow-up of a special track of the same flavor that was organized for the first time in EDBT/ICDT 2014 in Athens, Greece.

2017, Copyright is with the authors. Published in the Workshop Proc. of the EDBT/ICDT 2017 Joint Conference (March 21, 2017, Venice, Italy) on CEUR-WS.org (ISSN 1613-0073). Distribution of this paper is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0