

# Provenance Research Issues and Challenges in the Big Data Era

**Alfredo Cuzzocrea**

## Abstract

Provenance of Big Data is a hot-topic in the database and data mining research communities. Basically, provenance is the process of detecting the lineage and the derivation of data and data objects, and it plays a major role in database management systems as well as in workflow management systems and distributed systems. Despite this, provenance of big data research is still in its embryonic phase, and a lot of efforts must still be done in this area. Inspired by these considerations, in this talk we provide an overview of relevant issues and challenges in the context of big data provenance research, by also highlighting possible future efforts within these research directions.

**Alfredo Cuzzocrea** is Associate Professor in Computer Engineering at the DIA Department, University of Trieste, Italy. He is also habilitated as Full Professor in Computer Engineering by the French National Scientific Habilitation of the National Council of Universities (CUN) under the hegira of Ministry of Higher Education and Research (MESR). He is Research Associate at the Institute of High Performance Computing and Networking of the Italian National Research Council, Italy. Previously, he has been Senior Researcher at the Institute of High Performance Computing and Networking of the Italian National Research Council, Italy, and Adjunct Professor at the University of Calabria, Italy. During the past, he has also been Adjunct Professor at the University of Catanzaro “Magna Graecia”, Italy, Adjunct Professor at the University of Messina, Italy, Adjunct Professor at the University of Naples “Federico II”, Italy, and Adjunct Professor at the University of Naples “Parthenope”, Italy. He holds 50 Visiting Professor positions worldwide (Europe, USA, Asia, Australia). He serves as Springer Fellow Editor and Elsevier Ambassador. He holds several roles in international scientific societies, steering committees for international conferences, and international panels, some of them having directional responsibility. He covers a large number of roles in international journals and international conferences.

His current research interests include multidimensional data modeling and querying, data stream modeling and querying, data warehousing and OLAP, OLAM, XML data management, Web information systems modeling and engineering, knowledge representation and management models and techniques, Grid and P2P computing, privacy and security of very large databases and OLAP data cubes, models and algorithms for managing uncertain and imprecise information and knowledge, models and algorithms for managing complex data on the Web, models and algorithms for high-performance distributed computing and architectures. He is author or co-author of more than 350 papers in international conferences (including EDBT, CIKM, SSDBM, MDM, DaWaK, DOLAP), international journals (including JCSS, IS, KAIS, DKE, INS) and international books. He is also involved in several national and international research projects, where he also covers responsibility roles.