

CALL FOR PAPERS

Special issue of the ACM Journal of Data and Information Quality (ACM JDIQ) on Improving the Veracity and Value of Big Data

Guest editors

Floris Geerts, University of Antwerp, Belgium, floris.geerts@uantwerpen.be

Paolo Missier, Newcastle University, UK, paolo.missier@newcastle.ac.uk

Norman Paton, University of Manchester, UK, npaton@manchester.ac.uk

Aim

Big data poses serious challenges with regards to data quality. Indeed, data quality approaches need to be revisited to deal with large quantities of data (volume), to accommodate the dynamic aspects of data (velocity), to be able to uniformly deal with heterogeneous data originating from different sources (variety), to assess and improve the accuracy of data (veracity), and to provide an indication of the impact of data quality, both for decision making and monetary aspects (value).

In this special issue of JDIQ, we aspire to provide an overview of innovative research at the *intersection of data quality and big data*, from theory to practice, with a focus on improvements in veracity and value.

Topics

Specific topics within the scope of the call include, but are not limited to, the following:

- *Scalability of data cleaning techniques* in the context of possibly large quantities of distributed, dynamic and heterogeneous data. Examples of such techniques include the following:
 - Detection of syntactic and semantic errors.
 - Identification of duplicates; blocking of data.
 - Computation of suggested repairs.
 - Discovery of rules to improve data quality.
- *Automated data wrangling and transformation of data* for data analytics, including for instance the formalization of data transformations that re-format and correct values, and automated suggestions of data transformations for cleaning.
- *Budget constrained and value-based data quality* assessment including the following:
 - Methodologies for assessing the cost of cleaning.

- Techniques to add value to data and value-enhancing data cleaning methods.
- Pay-as-you go data cleaning.
- *Responsible cleaning* including guidelines, processes and methods to ensure the following:
 - Prevention of discrimination during preprocessing steps.
 - Fair unbiased repairing of the data.
 - Methods to assure transparent and auditable data quality workflows.
- *Data quality use cases* for big data.

Expected contributions:

We welcome the following two types of contributions:

- Research manuscripts reporting mature results.[25+ pages]
- Experience papers that report on lessons learnt from novel applications of data quality techniques on big data problems. These papers should provide generic insight that is of interest to the broad data quality community. [12+ pages plus an optional appendix.]

If this is an extension of prior published work, then submitted manuscripts must contain at least 30% new material, and the significant new contributions must be clearly identified in the introduction.

Submission guidelines with Latex (preferred) or Word templates are available here:
<http://jdiq.acm.org/authors.cfm#subm>

Important dates:

Initial submission:	Friday March 3, 2017
First review:	Friday June 2, 2017
Revised manuscripts:	Friday August 18, 2017
Second review:	Monday October 2, 2017
Publication:	December 2017