

# On the Role of Data Quality in Improving Web Information Value

Cinzia Cappiello  
Politecnico di Milano  
Dipartimento di Elettronica, Informazione e Bioingegneria  
Piazza Leonardo da Vinci 32  
20133 Milano  
cinzia.cappiello@polimi.it

## ABSTRACT

In today's information era, every day more and more information is generated and people, on the one hand, have advantages due the increasing support in decision processes and, on the other hand, are experiencing difficulties in the selection of the right data to use. That is, users may leverage on more data but at the same time they may not be able to fully value such data since they lack the necessary knowledge about their provenance and quality.

The data quality research area provides quality assessment and improvement methods that can be a valuable support for users that have to deal with the complexity of Web content. In fact, such methods help users to identify the suitability of information for their purposes. Most of the methods and techniques proposed, however, address issues for structured data and/or for defined contexts. Clearly, they cannot be easily used on the Web, where data come from heterogeneous sources and the context of use is most of the times unknown.

In this keynote, the need for new assessment techniques is highlighted together with the importance of tracking data provenance as well as the reputation and trustworthiness of the sources. In fact, it is well known that the increase of data volume often corresponds to an increase of value, but to maximize such value the data sources to be used have to be carefully analyzed, selected and integrated depending on the specific context of use. The talk discusses the data quality dimensions necessary to analyze different Web data sources and provides a set of illustrative examples that show how to maximize the quality of gathered information.

## Categories and Subject Descriptors

H.1.1 [Systems and Information Theory]: Value of Information

## General Terms

Management, Measurement

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.  
ACM 978-1-4503-3473-0/15/05.

## Keywords

Data Quality, Web Quality, Data Quality Assessment

## Short Bio

Cinzia Cappiello is an assistant professor at the Politecnico di Milano (Italy) from which she holds a Ph.D. in Information Technology (2005). Cinzia's research focuses on Data and Information Quality, Adaptive Service-Based Application and Green IT. In particular, she contributes to the Data and Information Quality area with research on the assessment and improvement of data quality in the context of fields like service-based applications, self-healing web services, wireless sensor networks, data warehouses, Web applications and mashups. She is Associate Editor of the ACM Journal of Data and Information Quality. She has been co-chair of the workshops "Quality in Web Engineering" in conjunction with ICWE 2010-2013, "Quality in Databases" in conjunction with VLDB 2010, "Data and Information Quality" in conjunction with CAiSE 2005, and of the tracks "Information Quality Management in Innovative IS" of MCIS 2012 and "Data and Information quality" of ECIS 2008. More information on <http://home.deib.polimi.it/cappiell/>

