

# Data Freshness Evaluation in Different Application Scenarios<sup>1</sup>

Verónica Peralta, Mokrane Bouzeghoub

Laboratoire PRISM, Université de Versailles  
45, avenue des États-Unis  
78035, Versailles cedex, FRANCE  
{ Veronika.Peralta, Mokrane.Bouzeghoub }@prism.uvsq.fr

**Abstract.** Data freshness has been identified as one of the most important data quality attributes in information systems. This importance increases especially in the context of systems that integrate a large set of autonomous data sources. In this paper we describe a quality evaluation framework which allows evaluation of data freshness in different architectural contexts. We also show how this quality factor may impact the reconfiguration of a data integration system to fulfill user expectations.

## 1 Introduction

Data freshness has been identified as one of the most important attributes of data quality for data consumers (Shin 2003) (Wang et al. 1996). Specifically, the increasing need to access to information which is available in several data sources introduces the problem of choosing between alternative data providers and of combining data having different freshness values (Naumann et al. 1999). This paper deals with data freshness evaluation in the context of a Data Integration System (DIS) that integrates data from different independent data sources and provides the users a uniform access to this data.

Data freshness represents a family of quality factors among which currency and timeliness are representative examples: *currency* describes how *stale* is data with respect to the sources and *timeliness* describes how *old* is data. In (Bouzeghoub et al. 2004) we analyze these factors and several metrics proposed to measure them. In (Peralta et al. 2004), we proposed a framework for analyzing and evaluating data freshness based on a calculation dag which abstracts a workflow of integration activities. After a brief recall of this framework, this paper shows how it can practically be used in different application scenarios and how the data integration system can be improved in order to fulfill user requirements in terms of data freshness.

The rest of the document is organized as follows: Section 2 briefly describes the data quality evaluation framework and discusses how to use it through different application scenarios. Section 3 focuses on the possible improvement actions to put on the DIS workflow to achieve user requirements. Finally, section 4 concludes with our general remarks.

---

<sup>1</sup> This research was partially supported by the French Ministry of Research and New Technologies under the ACI program devoted to Data Masses (ACI-MD), project #MD-33.