

Chapter 11

Principled Reference Data Management for Business Intelligence

Ivan Milman
IBM Corporation, USA

Martin Oberhofer
IBM Corporation, USA

Sushain Pandit
IBM Corporation, USA

Yinle Zhou
IBM Corporation, USA

ABSTRACT

Most large enterprises requiring operational business processes (e.g., call center, human resources, order fulfillments, billing, etc.) utilize anywhere from a few hundred to several thousand instances of legacy, upgraded, cloud-based, and/or acquired information management applications. Due to this vastly heterogeneous information landscape, Business Intelligence (BI) systems (e.g., enterprise data warehouses) receive unconsolidated data from a wide-range of data sources with no overarching governance procedures to ensure quality, consistency, or appropriateness. Although different applications deal with their own flavor of data (e.g., master data, metadata, unstructured and structured data, etc.), reference data (residing in code tables) is found invariably in all of them. Given the critical role that BI plays in ensuring business success, the fact that BI relies heavily on the quality of data to ensure that the intelligence being provided is trustworthy and the prevalence of reference data in the information integration landscape, a principled approach towards management, stewardship, and governance of reference data becomes necessary to ensure quality and operational excellence across BI systems. In this chapter, the authors discuss this approach in the domain of typical reference data management concepts and features, leading to a comprehensive solution architecture for BI integration.

DOI: 10.4018/978-1-4666-4892-0.ch011

INTRODUCTION

An organization's approach to reference data management (RDM) can have a profound influence on the effectiveness of that organization's business intelligence processes, reference data are those data values that support consistent representation (coding) of key information within an organization or across groups of organizations. Reference data are usually sets of permissible values for an attribute (Xu, et al., 2012) or classification schemas that are referred to by system applications, data stores, processes, and reports, as well as by transactional and master records (McGilvray, 2008). These may be standard codes defined external to the organization such as currency codes and country codes (such as those defined by ISO – the International Organization for Standards), domain specific standards subscribed to by the organization such as the Logical Observation Identifiers Names and Codes (LOINC 1999, 2013) standard for coding medical laboratory observations, or internal standards such as internally developed product codes. Such reference data sets, comprising a range of permissible values for entity attributes, usually reside within specialized tables known as look-up, code, check, or domain tables (Xu, et al., 2012).

Although there are many classification schemes that address the various types of data, almost all recognize the categories of master data, metadata, reference data, and transactional data (Dreibelbis, et al., 2008) (McGilvray, 2008). Of these, reference data is most related to master data and metadata, but reference data is distinguished from them. Master data represents the common business objects (such as people, places, and things) that need to be agreed on and shared throughout an enterprise (Dreibelbis, et al., 2008). Reference data, while used to consistently author master data instances by governing the permissible set of values for a given master data attribute, is not the same as master data. Reference data is also different from metadata which literally means “data about data”.

Metadata labels, describes, and characterizes other data and make it easier to retrieve, interpret, or use information (McGilvray, 2008).

Figure 1 illustrates the concepts of master data, metadata and reference data by using an example. Assume international retailer ABC Inc. who sells luxury products around the world and whose customers tend to travel and shop across the countries. A global customer information integration project is conducted to improve customer information and relationship management. The master data table in Figure 1 has the customer records which are captured when customers are shopping around the world. For example, customer Carrie Lee is a Chinese citizen who lives in Beijing. She shops for ABC's products in Beijing, but due to her job, she often travels to United States and South Korea, and purchases ABC's products there, too. In different regions, ABC uses the local language for product information and customer information. Therefore, Carrie Lee's information in the integrated system is in three languages: Chinese, English, and Korean. To properly represent this information as master data entities, a well-defined set of reference data tables is required. The reference data table shown in Figure 1 gives the range of permissible values for the attribute *language* and follows the ISO 639-1 standard (ISO 639, 2013). The metadata table describes the data usage and data structure specifications for each attribute in the master data table and gives the requirement for the reference data table as well. For example, the definition of *language* in the metadata table links back to the *language* reference data table, which follows the ISO 639-1 standard.

Although reference data tends to change less frequently compared to master data or transactional data, it is not totally static. For example, in the United States, postal codes (zip codes) are generally classified as reference data, but are subject to change over time as older codes are divided and new codes added to reflect changes in the population. Master data tends to change more quickly by regularly updating and merging

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/principled-reference-data-management-for-business-intelligence/96152

Related Content

Enterprise Information System and Data Mining

Kenneth D. Lawrence, Dinesh R. Pai, Ronald Klimbergand Sheila M. Lawrence (2010). *International Journal of Business Intelligence Research* (pp. 34-41).

www.irma-international.org/article/enterprise-information-system-data-mining/45725

Dynamic Enterprise Modeling for Knowledge Worker Industries

I. McKeachie (2007). *Adaptive Technologies and Business Integration: Social, Managerial and Organizational Dimensions* (pp. 59-81).

www.irma-international.org/chapter/dynamic-enterprise-modeling-knowledge-worker/4229

Design Methods for Decision Support Systems: A Perspective of Small Business B2C E-Commerce Environment

Madhury Khatunand Shah J. Miah (2018). *International Journal of Business Intelligence Research* (pp. 38-52).

www.irma-international.org/article/design-methods-for-decision-support-systems/203656

A Unified Approach for Taxonomy-Based Technology Forecasting

Andreas Henschel, Erik Casagrande, Wei Lee Woon, Isam Janajrehand Stuart Madnick (2012). *Business Intelligence Applications and the Web: Models, Systems and Technologies* (pp. 178-197).

www.irma-international.org/chapter/unified-approach-taxonomy-based-technology/58416

Discovering Data and Information Quality Research Insights Gained through Latent Semantic Analysis

Roger Blakeand Ganesan Shankaranarayanan (2012). *International Journal of Business Intelligence Research* (pp. 1-16).

www.irma-international.org/article/discovering-data-information-quality-research/62019