Chapter 3.24
Metadata Management:
A Requirement for Web Warehousing and Knowledge Management

Gilbert W. Laware
Purdue University, USA

ABSTRACT

This chapter introduces the need for the World Wide Web to provide a standard mechanism so individuals can readily obtain data, reports, research and knowledge about any topic posted to it. Individuals have been frustrated by this process since they are not able to access relevant data and current information. Much of the reason for this lies with metadata, the data about the data that are used in support of Web content. These metadata are non-existent, ill-defined, erroneously labeled, or, if well-defined, continue to be marked by other disparate metadata. With the ever-increasing demand for Web-enabled data mining, warehousing and management of knowledge, an organization has to address the multiple facets of process, standards, technology, data mining, and warehousing management. This requires approaches to provide an integrated interchange of quality metadata that enables individuals to access Web content with the most relevant, contemporary data, information, and knowledge that are both content-rich and practical for decision-making situations.

INTRODUCTION

Today, many of us use computers and the World Wide Web to communicate. We enter a Website name or address (www.informationbydesign. biz, www.ibm.com, www.tech.purdue.edu) into a browser on our desktop computer, where a unique numerical number replaces the words representing the Website name or address. It is analogous to a telephone number. We then are connected immediately to another computer assigned a numerical address somewhere on the World Wide Web. This allows us to access any document (Web page) on that computer. The Internet is capable of connecting us with any computer anywhere in the world. This computer sends the Web page we have requested from its Internet address to our

desktop computer, where it is displayed using our browser. In most cases, the returned Web page is written in English and we are able to understand its content. But, if the Web page is written in another language, we would need an interpreter to understand its content. In a telephone analogy — if a person who responds to our telephone call speaks another language, then what is said may not have any meaning to the caller. If the information that describes the currency, content, and location of the Web page or telephone number is erroneous, it is of little value.

Now, in a different way, let's consider the reason why it is difficult for computer systems to communicate and to share data. First, the data often have been structured differently in one system than in another. This is particularly true with older application systems. Second, the data may not be stored in the same format (i.e., they are in a numerical format rather than in text format). Third, the name for the data may be different, causing a problem in identification or recognition of what they represent between systems. Last, the values of the data stored may be inconsistent between the systems. Technically, the programs in each system can be interconnected if they are designed, defined, and structured logically and physically for that purpose. But, each of the above items has to be evaluated for possible integration and sharing of the data between the systems if that is not the case.

One of the most common problems is that identical data are named differently in different systems. All too often, different names or terms refer to the same data that need to be shared. For example, a human resources system may use the term employee or candidate to refer to a person. An ordering system may refer to a person or an organization as a customer. In a sales system, the term may be prospect, client, or customer. Each system may use different terminology — a different language in a sense — to refer to similar or identical data. But if they use the wrong lan-

guage, again, the systems cannot share the data to provide required information.

The problem can be even worse. Consider terms used in different parts of a business. Accountants use jargon — a technical language — that is difficult for non-accountants to understand. Similar terms used by individuals in engineering, production, sales, or in marketing may have different meanings. Likewise, managers may use another vocabulary. Each speaks a slightly different language and uses the same words in different ways. What is said may have no meaning without a common definition and so they cannot easily share common information. Each organization has its own internal language and jargon that becomes part of the subculture that evolves over time and is a key part of the way individuals communicate. In fact, in some organizations it is a miracle that people manage to communicate meaning at all!

As we saw above, there can be more than one language used in an organization. Metadata, the data about the data, identifies and uses their organization's own language. Where different terms refer to the same thing, a common term is agreed upon by all to use. Then people can communicate clearly. The same is true in the use of systems. Systems and programs intercommunicate when there is understanding and meaning between them. But without a clear definition and without common use of an organization's metadata, information cannot be shared effectively throughout the enterprise.

Previously we discussed how each part of an organization maintains its own version of customer, client, or prospect data. Each defines processes (a series of actions) — and assigns staff (persons) — to add new customers, clients or prospects to their own files and databases. When common details about customers, clients or prospects change, each redundant version of that data also has to be changed, requiring staff to make these changes. But, wait a minute, how

24 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/metadata-management-requirement-web-warehousing/25180

Related Content

Ontology
William Buchholz (2011). *Encyclopedia of Knowledge Management, Second Edition (pp. 1221-1236).*
www.irma-international.org/chapter/ontology/49068

Incorporating Indigenous Knowledge in the Higher Education Sector for the Advancement of African Scholarship
Ndwakhulu Stephen Tshishonga (2019). *The Formation of Intellectual Capital and Its Ability to Transform Higher Education Institutions and the Knowledge Society (pp. 212-231).*
www.irma-international.org/chapter/incorporating-indigenous-knowledge-in-the-higher-education-sector-for-the-advancement-of-african-scholarship/231063

Knowledge Producers and Consumers
Atreyi Kankanhalli, Bernard C.Y. Tanand Kwok-Kee Wei (2011). *Encyclopedia of Knowledge Management, Second Edition (pp. 867-877).*
www.irma-international.org/chapter/knowledge-producers-consumers/49034

Sustaining Organizational Innovativeness: Advancing Knowledge Sharing During the Scenario Process
Hannu Kivijärvi, Kalle Piirainenand Markku Tuominen (2012). *Conceptual Models and Outcomes of Advancing Knowledge Management: New Technologies (pp. 99-117).*
www.irma-international.org/chapter/sustaining-organizational-innovativeness/62419

Book Review: The Rise of Blockchains
Yuan Ren (2023). *International Journal of Knowledge-Based Organizations (pp. 1-6).*
www.irma-international.org/article/book-review/327451