

Metadata is likely to be useful in rational, data-driven, analytical decision-making scenarios. Not clear is whether it provides similar benefit in decision processes that are more intuitive or politically charged.

(process metadata) must be mapped to the physical configuration (infrastructure metadata); data delivery information (interface metadata) must be tied to registered users (administration metadata); and quality (content) metadata must relate to the actual data element it describes (model metadata). A consequence of adopting a narrow view of metadata while failing to understand the relationships among metadata components is the creation of fragmented “metadata islands.” Each island includes metadata of a specific functionality, unaware of and unable to communicate with other islands. Even when system designers and developers understand this complexity, implementing an integrated metadata layer is resource-intensive in terms of money, time, and managerial effort, as well as being a technical challenge.

COMMERCIAL SOLUTIONS²

Given the different types of metadata and the complexity of managing them, IT practitioners turn to COTS products for implementing data warehousing and metadata solutions [4, 10]. COTS products in this area are broadly classified into three categories. The first is data storage and management systems (such as Oracle, Sybase, MS-SQL, IBM-UDB, and Hyperion-Essbase). The second includes automated back-end data-processing utilities, commonly known as ETL products (such as Informatica, Oracle Warehouse-Builder, MS-SQL DTS, IBM Warehouse Manager, and Hummingbird). And the third includes reporting, or business intelligence (BI), utilities (such as MicroStrategy, Business Objects, and Cognos). Most leading data warehousing products provide GUI-supported utilities for metadata management.

An examination of the leading COTS products

reveals an ambiguous picture. On one hand, all the vendors acknowledge the importance of metadata and embed it within their products. On the other, these products fail to address several critical issues:

- Comprehensive support for all metadata functionalities. While infrastructure, model, and administration metadata are commonly supported, process metadata is supported only by

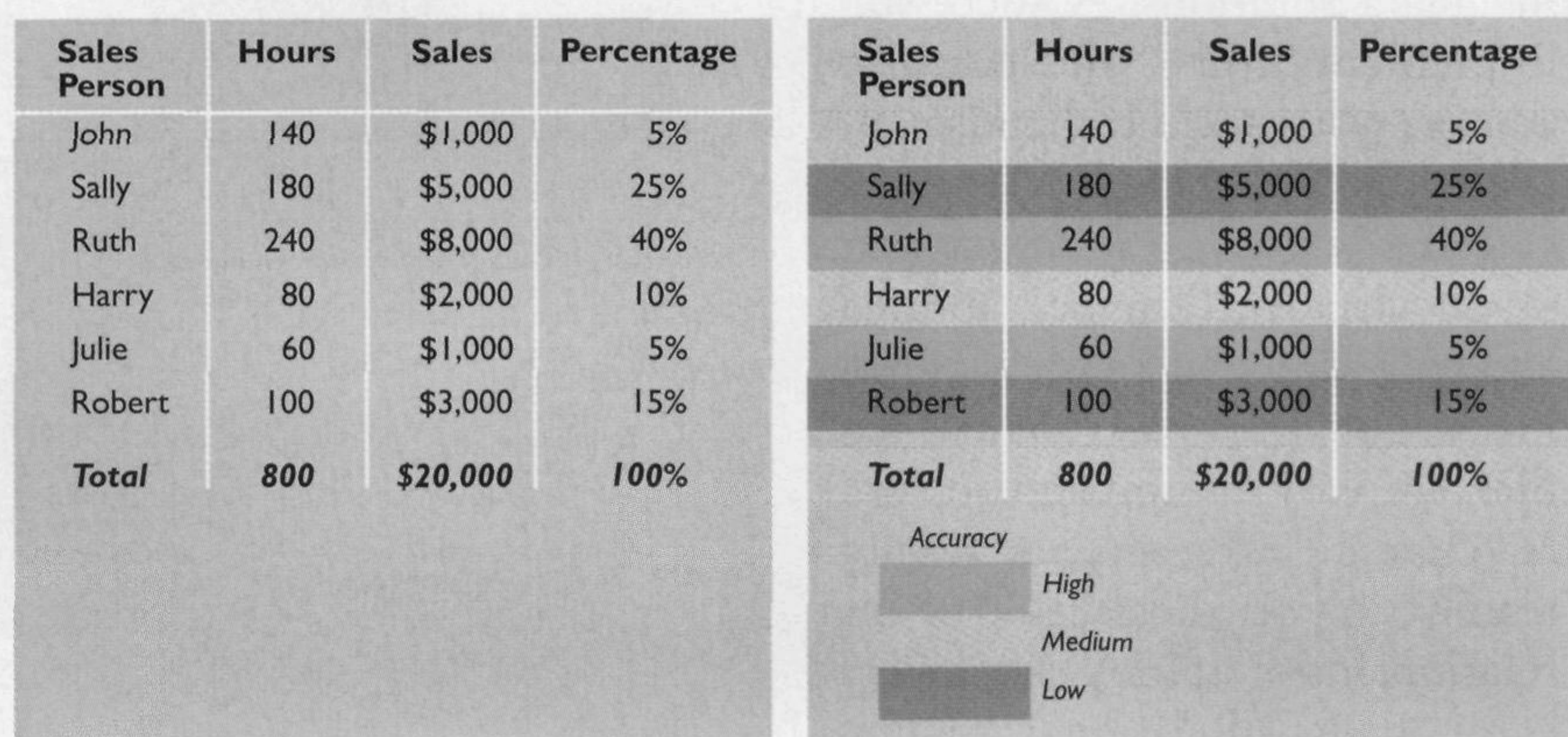


Figure 1. Business reporting with and without data quality metadata.

ETL products, and data-delivery metadata only by BI products. Quality metadata is not yet supported by any of the leading products. This fragmentation necessitates integration across multiple products to create a comprehensive metadata repository;

- Little support for business metadata (and an emphasis on technical metadata). Essential elements (such as those required to interpret data-transformation processes, link user administration to customer relationship management utilities, and track resource use) are minimally provided, if at all. An exception is the business elements of a data dictionary supported by most reporting and BI products;
- Metadata storage and representation formats restricted to relational or complex proprietary structures. This makes products more efficient but preempts integration and/or exchange of metadata with other formats, including textual and graphical (such as ER and DaVinci diagrams); and

²This information is based on a comprehensive review we conducted in February 2004 and revisited in September 2005 [8].