

The Believability of Data

Hugh J. Watson



Hugh J. Watson holds a C. Herman and Mary Virginia Terry Chair of Business Administration in the Terry College of Business at the University of Georgia.

He is a Fellow of TDWI and the senior editor of the *Business Intelligence Journal*.
hwatson@uga.edu

If people do not believe data, they are unlikely to rely on it. People base their beliefs and actions on perceptions, whether or not those perceptions are accurate. You must keep this in mind when providing BI applications and data.

Data believability is related to, but different from, data quality. Data quality focuses on the attributes of data (e.g., accuracy, consistency, completeness) and its suitability for meeting business requirements (Eckerson, 2002). Historically, organizations have struggled to provide high-quality decision-support data, and the need to do so has been a driving force behind data warehousing. Data believability, on the other hand, relates to whether the data is *regarded* as being true, real, and credible (Wang and Strong, 1996). Data believability is based on perceptions.

Although data quality and data believability tend to go hand in hand, this is not always the case. Data can be of high quality and still not be believed, such as when good data does not correspond to users' expectations. In other cases, the data can be bad but users do not know it.

What Makes Data Believable

Many factors influence the believability of data, including:

- **The source of the data.** Some sources are more credible than others. Over time, people tend to learn which sources can be trusted (for specific purposes) and which cannot.
- **The timeliness of the data.** The more recent the data, the more users trust that it accurately reflects current conditions. This is a major reason why users always want fresher data.
- **How well the data matches users' preconceptions.** Users often have expectations about data based on

their business experience, and data is more believable when it is consistent with these expectations. For example, sales figures that seem “out of line” are likely to be questioned.

- **Users’ understanding of and confidence in how the data is processed.** Users want to know “where the numbers come from,” especially in the case of metrics used to evaluate performance. They want to know what data was used, the processing lineage (i.e., history), and the computations performed.

Making Data More Believable

There are a number of ways to make data more believable. I’ll describe three of them here.

At the Application Level

The effort spent and the technologies and approaches used to improve data believability vary by application. Excellent candidates include executive information systems (EISes) and dashboards, which are usually custom-built and can be designed to enhance the believability of the data.

One best practice in screen design for an EIS (and other applications) is to show when the information displayed was last updated. It is also wise to provide the name and contact information for the subject-area expert who is either the data supplier for the screen or is most knowledgeable about the information displayed, so that users can call the person with their questions. It is also possible (and wise) to provide information about the source of the data, how metrics are computed, and term definitions. This information is not usually shown on the display screen, but may be accessed by clicking a metadata button. Each screen should have applicable and possibly unique metadata associated with it.

The testing phase of application development can also be used to enhance data believability. Several years ago I was working on an enterprisewide decision support application, and we had an experienced operating sponsor who was very willing to review any new information screens developed. On more than one occasion, he commented: “These numbers aren’t right,” even though the data came

from the official source. In every case, after further inspection, we found that he was right; something was wrong, either with the data itself or with how it was analyzed or displayed. When the system was ultimately rolled out, we were confident that the data would be believable.

At the Metadata Level

There are multiple ways to think about, categorize, and organize metadata. A common approach is to distinguish between metadata for IT and metadata for business users. IT metadata helps the warehouse staff understand the movement and use of data in the warehouse. Metadata for users provides information about what data is available in the warehouse and how to use it.

People base their beliefs and actions on perceptions. You must keep this in mind when providing BI applications and data. ... It is also possible (and wise) to provide information about the source of the data, how metrics are computed, and term definitions.

In thinking about metadata for users, keep data believability in mind. This may lead to giving users metadata that might initially be considered appropriate only for IT. For example, the source systems used to populate the warehouse, the movement and data transformations as the data flows from source systems to the warehouse, and warehouse refresh cycles are important information for IT, but this information may also be important to users in assessing the believability of the data in the warehouse. For believability purposes, users may need access to more information about the source and processing of warehouse data than was planned.

At the Governance Level

Organizations use committees, people, and processes to govern BI initiatives and activities. There are often committees that span the organizational hierarchy and provide direction commensurate with the organizational level. For example, at the highest level, senior managers set the overall direction and make resources available. At the next level, mid-level managers set schedules and priorities for projects. At the operational level, the BI staff and representatives from the business units (normally power users) work on operational issues, including issues involving data quality.

While all of these committees can contribute to data believability (such as funding data quality initiatives and appointing data stewards), the greatest gains can be realized at the operational level. Blue Cross and Blue Shield of North Carolina, a 2001 TDWI Best Practices Awards winner in the governance category, provides a good example (Watson, Fuller, and Ariyachandra, 2004). BCBSNC has a business-requirements group with representatives from the BI staff and all of the business areas that use the warehouse. This committee meets on a weekly basis and discusses detailed requirements, reaches consensus on data definitions, solves data problems, and performs user acceptance testing. By deeply involving users in data-related tasks and issues, the believability of the data is greatly enhanced.

Conclusion

Although BI professionals (and others in the organization) may work hard to provide high-quality data, that data must be believable. As a statistician would say, “High quality is a necessary but not sufficient condition.” Take the extra steps necessary to ensure that your users trust the decision support data that you work so hard to provide. ■

References

- Eckerson, Wayne W. [2002]. *Data Quality and the Bottom Line*, TDWI Research Report, March.
- Wang, R. and D. Strong [1996]. “Beyond Accuracy: What Data Quality Means to Data Consumers,” *Journal of Management Information Systems*, spring, pp. 5–33.
- Watson, Hugh J., Celia Fuller, and Thilini Ariyachandra [2004]. “Data Warehouse Governance: Best Practices at Blue Cross and Blue Shield of North Carolina,” *Decision Support Systems*, fall, pp. 435–450.