

# The Need for Data Quality

Blake Patterson, Member of Technical Staff

Information Quality Group  
AT&T Bell Laboratories  
Holmdel, NJ USA

Data is a key asset for most companies.  
But most data is flawed.

These are some reasons for flawed data:

- Lack of management understanding
- Flawed data-acquisition processes
- Too many databases
- "Duplicate" data produced by different methods, organizations, companies
- Flawed data-update processes

The Information Quality group of AT&T Bell Laboratories has developed and applied methods to improve the quality of data in some AT&T databases, and our goal is to reduce data error rates by another 97%. We believe that the following areas of future work will help achieve that improvement:

- **Data Management**

We need to develop comprehensive theories that manage data as an asset and a resource. We need models that describe what value data contribute and how. Data is usually not bought and sold as a valued asset. Rather, companies bundle it with other products and give it little attention.

- **Measurement Systems**

Standard methods for reporting data quality do not exist, despite the importance of accurate measurement to any discipline. Standards bodies exist for most physical properties, but there is no agreed-on metric for data quality.

---

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, requires a fee and/or special permission from the Endowment.

Proceedings of the 19th VLDB Conference  
Dublin, Ireland, 1993

- **Data Architectures**

Lower costs for data storage and distributed computing have led to multiple databases and poorer data quality. Most data resides in several databases to make it more accessible to users. When data changes, how do companies update all databases quickly? How many copies of data should a company keep? Companies need data architectures that balance user's needs for access, security, and low cost.

- **Process Simulation**

Computer simulation is an important tool for optimizing processes like manufacturing. We believe that computer models for processes that create data values could help improve those processes and thus improve data quality.

- **Information Processes and Systems Designed for Data Quality**

Usually, process engineers begin with existing information systems and design processes to best use those systems. Similarly, system designers create information systems that best fit existing processes. Both have little regard for data. We need the joint efforts of process engineers and system designers to increase data quality.

As a consultant to an organization that analyzes data-intensive processes to help owners improve those processes, I have observed that even after analysis shows obvious data problems, it is hard to convince anyone to take steps to resolve the problem.