

# DATA AND INFORMATION QUALITY - WILL THE NEW YORK *TIMES* EVER "GET IT"?"

James Hurysz

Quality Consultant, Alexandria, Virginia

## 1. Introduction

The New York *Times* is generally considered to be one of the world's great newspapers. The *Times's* editorial and journalistic philosophies are a hundred years old. They were implemented by Adolph S. Ochs, who purchased the *Times* in 1896 and published it until his death in 1935. Ochs's descendants publish the *Times* today.

Ochs aimed for the *Times* to give the news concisely, clearly, promptly, and impartially, without fear or favor. Ochs also aimed for a large audience of patriotic, affluent, and educated readers. His goals clearly differed from the journalism of the day. William Randolph Hearst and Joseph Pulitzer made fortunes from "yellow journalism" in their respective papers, the New York *Journal* and New York *World* [Frankel, 1996]. The renowned journalist H.L. Mencken noted in his autobiography that at the turn of the century even journalists who worked for reputable newspapers routinely made up "facts" [Mencken, 1941].

The New York *Times* prospered under the leadership of Adolph Ochs and his descendants. The New York Times Corporation had revenues of \$2 billion in 1995 and is a multi-media conglomerate. It owns television and radio stations and is expanding both in print and electronic publishing. In 1993, the New York Times Corporation bought the Boston *Globe*. In 1995, an electronic edition of the *Times* began to appear on the America Online electronic network. Despite such valuable goals as impartial reporting, many sources show that the *Times* pays insufficient attention to information and data quality. By expanding its markets the *Times* increases the need for data quality techniques in both its editorial policies and information content.

### 1.1. Sources of Information

There are several sources of information about the *Times's* history, operations, people, personalities, corporate culture, editorial decisions, reporting, and other factors that affect information quality and data quality. Well-known sources include the widely-available autobiographies of *Times* journalists and editors [Talese, 1969; Salisbury, 1980; Reston, 1991]. Other sources include books by journalists outside the *Times* [Diamond, 1993]. The New York *Times* is an excellent source of information about itself, when analyzed statistically and for information and data quality. A printed index to the *Times* has been available for decades; an electronic index has been available for almost twenty years. Both the index and *Times* articles for the past several years recently become available on CD-ROM. This allows fast Boolean and topic searches; and subsequent retrieval of an article. These are

important tools, as they can help not only a researcher but reporters and editors produce more accurate articles more quickly.

Other sources of information about the New York *Times* can include social, economic, and demographic data about journalists, editors, managers, and executives - for example real estate directories that show where journalists and editors live. Of course, one can also interview *Times* journalists, editors, and other employees. And because the New York Times Company stock is publicly traded, there is significant financial and legal information available to researchers from public sources, such as the U.S. Securities and Exchange Commission.

### 1.2. Characteristics of the New York Times: 1995-1996

This paper explores information and data quality in the New York *Times* during a twelve month period - September 1995 through August 1996. The author lived in New York City and the New York metropolitan area for six years (1979-1985) and has subscribed to both the daily and Sunday editions of the *Times* for ten years. The author has either contacted or attempted to contact *Times* journalists on a dozen occasions during the past decade - most recently through the *Times's* America Online electronic bulletin board - regarding what appeared to be serious data quality and information quality errors and omissions. (For example, maps that have appeared in the *Times* have sometimes been so ambiguous as to be useless). Typically, the *Times's* response was no response. When the author did speak to a *Times* journalist or editor, that person invariably claimed that a "deadline" loomed, and that he would "take care of that when I have time."

Today, weekday editions of the *Times* cost sixty cents within the New York metropolitan area and a dollar outside the New York metropolitan area. The Sunday edition costs \$2.50 within metro New York, and \$3.00 outside. The cost of a year's subscription in the near-suburbs of Washington D.C is \$442. The *Times's* dimensions and pages have been steadily "downsized." During the past thirty years, the number of pages have been reduced from over a hundred to about eighty for weekday editions; and from over two hundred to about a hundred and fifty on Sunday. Holiday (weekday) editions are about forty pages. Physical dimensions of the pages have also shrunk. Today, a *Times* page measures 13 1/2 inches by 22 1/4 inches. A page in the Sunday *Times* Magazine measures 9 1/2 by 11 1/2 inches. These are significant changes from how the *Times* appeared on newsstands 35 years ago.

## 2. Analysis

The New York *Times* is run as a profitable business owned (as mentioned earlier) by a media conglomerate. It is not a federal statistical agency, nor a non-profit organization. Therefore, the primary oversight of *Times* operations is by corporate management, shareholders, and federal agencies (like the EPA, and IRS) that regulate corporations for environmental, tax and other compliance. Content is regulated by the marketplace, by libel and national security statutes, by the *Times's* "corporate culture," and by the ideas, opinions and beliefs of its journalists, editors, managers, and owners. Libel laws have become much less onerous for the media during the past thirty years. The

*Times*'s editors and executives have relentlessly sought to broaden First Amendment free speech guarantees, both in the courts and through *Times* editorials. Indeed, the present American legal standard for libel (willfully malicious reporting with reckless disregard for the truth) was the result of a libel case (i.e., *New York Times vs. Sullivan*) that the *Times* won in 1964 before the U.S. Supreme Court. In the area of "national security," however, the *Times* is extremely circumspect [Salisbury, 1980]. The 1971 publication of the "Pentagon Papers" was an aberration. The *Times* has, to date, not seriously attempted to enlighten its readers about the veracity of the "facts" it reported during the Cold War, though many new sources of information about the Cold War now exist [Reston, 1991].

### **2.1. Information Quality and Data Quality in the New York Times**

The purpose of publishing newspapers is to sell information and data. The New York *Times* is entirely information and data, in various forms. The *Times* uses prose, photographs, and graphics to describe and illustrate events; and uses data (individual facts, statistics, or items of information) to bolster prose arguments, to lend authenticity to statements, to make hypotheses credible, and as succinct statements of fact themselves (e.g., "Princeton 17 - Cornell 7").

Obviously, there is a large amount of information and data in each issue of the *Times*. In addition to pages of subjective reports and advertising, there are photographs, graphs, tables, and thousands of individual data items (for example, stock quotes, sports scores, and racing statistics). There is also information about data, and data about information. A search of the UMI/NEXIS New York *Times* CD-ROM database between January, 1995 and June, 1996 discovered no instances of "information quality" or "data quality" appearing as topics, and only a half-dozen occurrences of the phrases "data quality," "quality of data," "information quality," and "quality of information."

### **2.2. Objective and Subjective Information and Data**

As in all newspapers, there is wide variability in data and information quality in the New York *Times*. This is by design. To exist and be profitable newspapers must provide readers with a wide variety of interesting topics. Few would purchase a newspaper that consisted of verbatim interviews transcribed from magnetic tape to print, sports scores, and stock quotations. But newspapers must guard against the appearance of objectivity when publishing data and information that is subjective, sometimes very subjective. Newspapers are published by human beings. Journalists, editors, and newspaper executives are subject to human emotions, and the stress related to daily reporting, editing, and publishing. Finally, though the *Times* may print an interesting hypothesis, survey, report, dataset, or a pronouncement by an esteemed scientist may be the subject of a half-page article, it does not necessarily "prove" anything!

### **2.3. Information and Data Quality Problems and the Times**

.....From the viewpoint of information and data quality, the New York *Times* is seriously remiss. Data and Information Quality simply don't appear as topics. The only recent use of the phrase "data quality" was in a pejorative quote [Bryant, 1995]. A Boolean search of the UMI/NEXIS *Times* CD-ROM

database for 1995-1996 indicates that when "data" and "quality" appear in the same article, the context is almost always data about the quality of a product or service. Information quality and data quality *per se* are only part of the *Times's* problems in both areas. Other IQ/DQ problems are longstanding. A graduate journalism student could (in the author's opinion) easily write a doctoral thesis about each problem.

### 2.3.1. Corporate Culture and the *Times*

One idea that clearly comes through in the memoirs of the great "*Timesmen*" (i.e., Reston, Salisbury, and Talese) is the extent to which *Times* journalists, editors, and executives are (and have been) "wired" into a system that is the antithesis of information and data quality. All too often "getting ahead" for an aspiring *Times* journalist involved reporting quotations and other self-serving information from foreign and domestic politicians which were obtained by beginning "friendships" with them [Reston, 1991]. As stated earlier, we still don't know much about what American and Soviet leaders were doing during the Cold War *versus* what the New York *Times* and other newspapers told us they were doing.

### 2.3.2. Pride, Prejudice and the *Times*

One has only to peruse the New York *Times's* Metro Section for a few months to determine which topics are either "off limits" entirely or covered at a low priority. For example, metropolitan news about boroughs outside Manhattan appears to be a low priority while suburban news seems to be a high priority [Diamond, 1993a]. Unfortunately, the same prejudice appears to be at work when the *Times* reports on other topics. The *Times* does not like guns or people who commit adultery. It does not favor the legalization of marijuana, and, until recently, its coverage of gay and lesbian issues was nil. Although today over a third of the *Times's* staff writers are women, in the recent past the *Times* was almost entirely staffed by white men. Ochs and his descendants had (and have) their own prejudices [Diamond, 1993b]. Unfortunately, *Times* readers would not know what those are without a careful reading of the paper and background biographies and autobiographies. Obviously, when the *Times* is prejudiced about a group, an issue, or a topic information doesn't appear or the information that does appear is negative. A careful reading of the *Times* indicates that *Times* journalists typically represent the values of their day. And many of those values don't hold up well after a decade of socioeconomic, demographic, economic, cultural, and political changes.

### 2.3.3. Surveys, Hypotheses, Data, Reports, and Pronouncements

The most ubiquitous information quality and data quality problems in the *Times* do not involve data *per se*. Stock quotations, sports scores, racing results and similar data have been compiled and published for decades by people who know and understand what they are doing. But it appears that *Times* journalists, editors, and executives often know little about the subjects of their articles, about how research is conducted, about peer review, about the scientific method, about quality, and about science.

This ignorance is evidenced both by what *Times* articles say and how they say it. A sampling

of recent article titles indicates a reliance on data and surveys. An article citing a study by a child advocacy group claims that the number of children of the working poor is up sharply [Holmes, 1996]. A professor theorizes that there is no evidence that global warming is serious, but presents no data to prove his hypothesis [Stevens, 1996]. Atmospheric data point to the ultimate recovery of the ozone layer [Associated Press, 1996]. A report indicates that health costs are growing more slowly in the United States, but the time series is too short to predict a long-term trend [Pear, 1996]. Economic data are sending conflicting signals [Uchitelle, 1996]. A survey shows that nonvoters are more alienated than voters [Berke, 1996]. And data seem to show a solar system nearly in the neighborhood [Wilford, 1996]. These and similar articles are the staple of the *Times* today. Presumably the *Times* journalists who wrote these articles took the time and effort to determine the quality of the data and information the researchers relied upon - the reader almost always can't. At a minimum, a short evaluation of the data source should be included.

### 3. Conclusion

The New York *Times* has a "world class" reputation for quality in journalism. But the *Times* is often biased in its coverage of people and events. *Times* journalists write articles with insufficient attention to information and data quality; and *Times* editors and executives publish such articles frequently.

#### References

- ASSOCIATED PRESS. 1996. New Data Point to the Ultimate Recovery of the Ozone Layer. *New York Times* (May 31), A14.
- BERKE, R.L. 1996. Nonvoters Are No More Alienated Than Voters, a Survey Shows. *NYT* (May 30), A21.
- BRYANT, A. 1995. Looking to Software for Help. *NYT* (October 15), 18.
- DIAMOND, E. 1993. *Behind the Times*. Random House, New York.
- DIAMOND, E. 1993a. *Behind the Times*, pp. 390-391.
- DIAMOND, E. 1993b. *Behind the Times*, pp. 382-386.
- FRANKEL, M. 1996. 'Earnest' Goals: First Times Under Ochs. *NYT* (August 19), A1.
- HOLMES, S.A. 1996. Public Cost of Teen-Age Pregnancy Is Put at \$7 Billion This Year. *NYT* (June 13), A19.
- MENCKEN, H. L. 1941. *Newspaper Days, 1899-1906*, Alfred A Knopf, New York, pp. 260-275.
- PEAR, R. 1996. Health Costs Are Growing More Slowly, Report Says. *NYT* (May 28), A14.
- RESTON, J. 1991. *Deadline, A Memoir*, Random House, New York.
- RESTON, J. 1991a. *Deadline, A Memoir*, pp 443-450.
- RESTON, J. 1991b. *Deadline, A Memoir*, pp. 199-244.
- SALISBURY, H.E. 1980. *Without Fear or Favor*, Times Books, New York.
- SALISBURY, H.E. 1980. *Without Fear or Favor*, pp. 118-148.
- STEVENS, W.K., 1996. A Skeptic Asks, Is it Getting Hotter, Or Is it Just the Computer Model? *NYT* (June 18), C1.
- TALESE, G. 1966. *The Kingdom and the Power*, World Publishing Company, New York.
- UCHITELLE, L. 1996. Economic Data sending Conflicting Signals. *NYT* (June 12), D1.
- WILFORD, J.N. 1996. Data Seem to Show a Solar System Nearly in the Neighborhood. *NYT* (June 12), A24.