Recommendations on safety management systems (SMS) typically address the requirements of implementation but less often the challenges associated with data collection. Inadequate quality of data — "garbage in, garbage out" (GIGO) — can be a problem, as well as too many — or too few — data — which can yield the same net effect, the inability to adequately analyze, understand and act on the organization's safety deficiencies and objectives.

An organization's SMS can be thought of as a data hub, with programs that feed into the SMS as data spokes. Hub-and-spoke data can be derived from a multitude of sources such as flight operational quality assurance, a fatigue risk management system, an aviation safety action program, a line operations safety audit (LOSA), and the analytical results generated by an SMS.

Sometimes all these data become so difficult to manage that their intended benefit is never fully realized. A number of problems may manifest during data collection. The first, GIGO, later can make data interpretation problematic because of a low or undetermined level of confidence. Second, an overabundance of data in relation to the time and tools available can place a severe burden on a safety manager trying to sort through it all and have it make any sense. Effectively, there is so much information that the safety manager may suffer from what I call "data delirium." Conversely, the third problem — a scarcity of data — may not allow management to make actionable decisions because it is unclear whether the data represent reality. This article will address each of these potential problems and offer practical solutions.
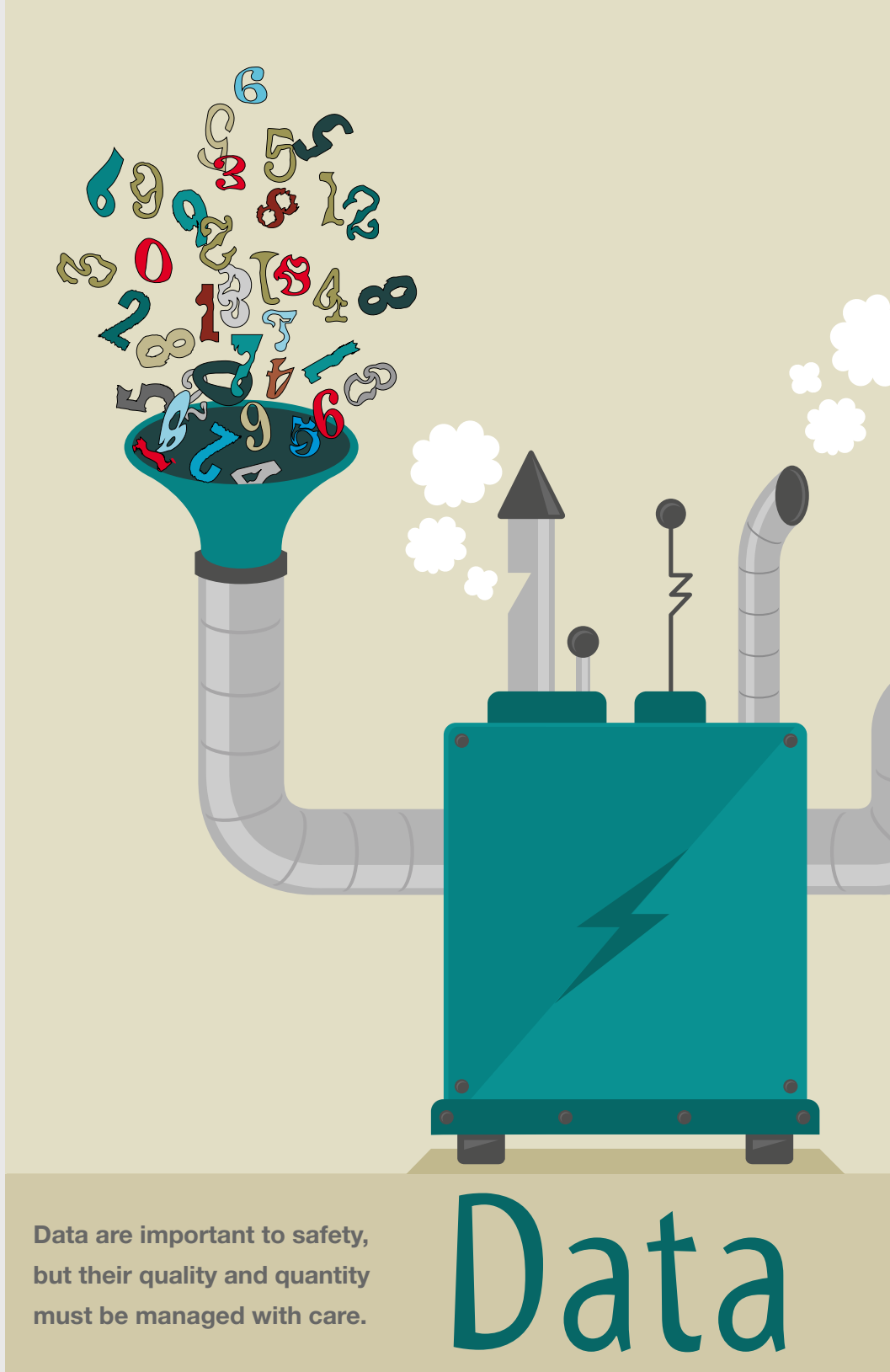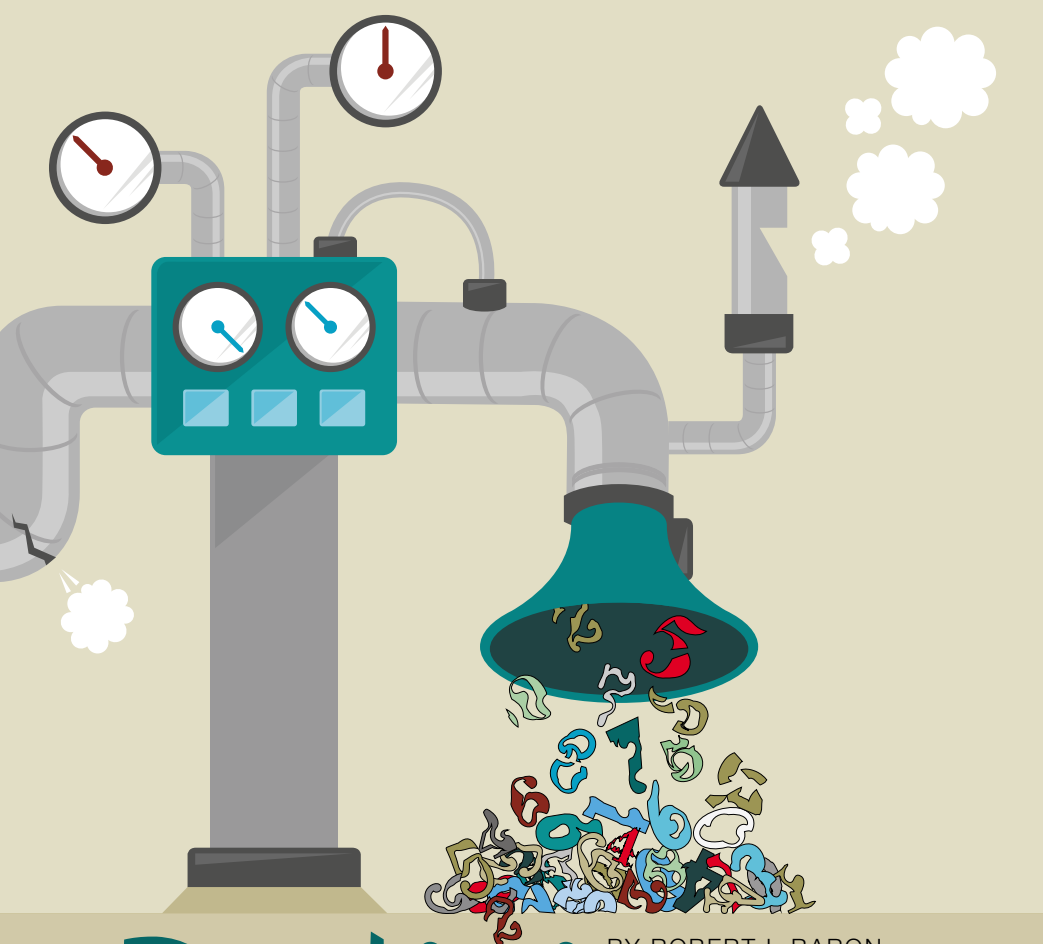


**Data are important to safety, but their quality and quantity must be managed with care.**

## Data Basics

Data can be obtained quantitatively (by focusing on raw numbers), qualitatively (by interpreting text in narrative reports) or a combination of both. Quantitative data are relatively easy to analyze using descriptive and inferential statistics. An example of descriptive analysis of quantitative data is dividing total accidents by a number representing risk exposure (such as total departures) to determine, for example, the accident rate per 100,000 departures in a particular geographic region. This type of data provides useful metrics for

survey. Surveys often use short statements that collect data quantitatively through the use of a Likert scale (a scale typically ranging from 1 to 5, each number representing the strength of the respondent's opinion or attitude toward the corresponding statement). For each statement, there also may be a text area where qualitative data can be collected. This allows the respondent not only to provide the numerical score for an opinion or attitude about each statement, but also to expound on each numerical response with a short explanation.

Regardless of how data are collected, the GIGO principle must be considered for quality control. One of the biggest challenges of collecting data is assuring that the data are valid (measuring what they purport to be measuring) and reliable (consistent when measuring the same thing). Scientific research methodology applied to aviation safety shares theories, statistical concepts and specialized terminology with other fields. Plenty of courses, websites and college textbooks offer further explanations.

Here are a few examples of how GIGO might affect your data. In the first example, let us say an airline conducts a LOSA (a spoke in the SMS hub). LOSA data collection consists of trained observers, riding in the cockpit jumpseat, who fill out quantitative and qualitative checklists related to observed crew performance. Although an observation of every single flight would be highly beneficial, it would obviously not be very practical. Thus, LOSA observations require a series of flights as a sample of the entire flight operation.

A sample should, in theory, very closely resemble the overall flight operation including the flights not



BY ROBERT I. BARON

# Delirium

composite Illustration: Susan Reed; machine: © sammyc/iStockphoto

comparisons but does not tell us much about the actual accidents.

Qualitative analysis of texts, though more unwieldy, time-consuming and potentially subjective, can provide a much more robust understanding of a construct within the accidents under study. An example of qualitative data use

is a brainstorming session directed at identifying airside hazards (e.g., producing a preliminary hazard list). A combination of both methods will offer a much more complete picture of the construct(s) under study.

One data collection instrument that incorporates both methods is the

observed. But what if the sample has been designed incorrectly and therefore does not truly represent the entire flight operation, for example, if the sample is too small? What if the observers are not properly trained or calibrated — *calibration* meaning they have standardized criteria so that there is inter-rater reliability among observers when recording threats, errors and undesired aircraft states. What if the observations are heavily skewed toward one particular fleet or route? Once the LOSA is completed, does the airline have a valid and reliable picture of the entire operation? Probably not.

LOSA data, like any kind of data with safety implications, may require a significant allocation of financial and human resources. If management does not believe your data, it is unlikely that you will be approved for those resources.

For the second example, let us use a survey to understand the GIGO principle. The safety manager at a major airport wants to measure employee morale. Morale can have a very significant impact on safety, because employees with low morale may not be motivated to work as safely as possible. So the safety manager creates a survey using statements that she feels would adequately measure employee attitudes about morale. The survey presents five statements and incorporates a Likert scale (1 — strongly agree, 2 — agree, 3 — neutral, 4 — disagree, 5 — strongly disagree). The statements are worded as follows:

1. Management is never on the same sheet of paper.

2. Low morale seems to be the norm around here.

3. I think low morale is correlated with low self-esteem.

4. Everyone I work with is unhappy most of the time.

5. They don't pay me enough to motivate me.

The safety manager emails a survey link to all airport employees, including contractors (approximately 1,200 total people) and makes the survey available online for 14 days. Upon completion of the data collection period, there are 100 responses and the safety manager emphatically declares that the results are conclusive: Employees are suffering from low morale. But could the results have been affected by the GIGO principle? Yes, and here are a few reasons why:

**Although short surveys are well received, these five statements do not adequately address the full dimensions of a construct such as morale.** The statements are not based on an accepted definition of the construct being studied, but are based on the safety manager's own definition of morale. A review of the extant research literature should be conducted to operationally define the research constructs (or variables).

**The statements include a *neutral* point.** There are mixed opinions about the use of a neutral point. The problem is that respondents use the neutral point as a "safe zone" if they are uncomfortable expressing their genuine feelings, even anonymously. Too many of these neutral answers can work against the purpose of the survey, which is to measure attitudes and opinions about the construct being studied. Some argue that everyone really has an opinion, even if he or she would prefer not to reveal it to the researcher.

**All of the statements are negatively worded.** When all statements have a positive or negative value, it can influence respondents to choose the same response for each. This is called the "straight-line effect."

The actual wording of some of the statements is problematic:

- **Management is never on the same sheet of paper.** Ambiguous. Does this mean lack of agreement or coordination among management personnel, or between management and line employees? Do all respondents understand the expression "on the same sheet of paper"?

- **I think low morale is correlated with low self-esteem.** Confusing. The respondent may not know how to define *low morale* and *low self-esteem*. Additionally, some respondents may not understand the definition of *correlated*. This can become more problematic when English is not the respondent's first language.

- **Everyone I work with is unhappy most of the time.** Double-barrel statement involving *two* criteria. One could be true, the other not. The two problematic words are *everyone* and *most*.

- **They don't pay me enough to motivate me.** A leading question that could suggest a particular answer. Also, this statement has a strong bias, and the word *they* can be ambiguous.

There are problems with the methodology, including:

- **The inclusion of contractors.** Contractors may not be airport employees and thus may come from a very different culture at their own organizations. Contractor responses can skew the results of the resident airport's own personnel.

- **The time allocated for data collection.** Two weeks is insufficient to collect a large number of responses. A better collection period would have been four weeks. After two weeks, a reminder email should have been sent out.

- **Low response rate.** Although response rates for surveys are typically low (in the 20–30 percent range), 8 percent was exceptionally low. This response rate can have implications for the sample, as was discussed earlier. Does this sample adequately represent the other 92 percent of the airport population? Was there something different about the employees who participated in the survey compared with those who did not? Are the results statistically significant (i.e., capable of being extrapolated to the larger population)? Would the results have been different if all 1,200 people had answered the survey?

This was not a very well developed survey and its distribution was problematic (garbage in). Thus, the safety manager may have come to a false conclusion based on the results (garbage out). It would be hard to sell to management on allocating resources to the problem.

## Data Excess

An overabundance of data can become so burdensome that the safety manager may suffer from data delirium. Some safety managers have complained that, while their SMS is a welcome hub for their company's safety processes, paradoxically, sometimes they do not know what to do with all their data. The problem may not be poor data management, but rather a shortage of human resources. Or perhaps the staffing is adequate, but there is so much irrelevant data that it is tying up those limited resources. Whatever the case, I offer the following recommendations.

If the problem is a shortage of human resources, the obvious solution would be to hire more people to assist with data analysis. That may not be feasible these days, where lean is the corporate *modus operandi*. If there is a legitimate need for additional help, consider a temporary

service or a college student intern. Interns are invaluable resources, especially if their study has included research methods and data analysis.

If the staffing is sufficient, but an overabundance of irrelevant data is the issue, then it would be worth taking a look at all the data sources and considering the use of data filters. Which incoming data are relevant and which are not? Prioritize the most-need-to-know data. This does not mean that the other data are irrelevant or useless, just that they will be lower priority.

Are you simply collecting too much data? As a qualitative example, there was a safety manager at an airline who insisted on posting on a bulletin board U.S. National Transportation Safety Board (NTSB) accident reports for operationally similar airlines and environments. That seemed a great idea. However, what he posted was the entire accident report (sometimes hundreds of pages). Pilots are busy. You cannot expect them to read through a complete accident report. This is a case of too much data. A better approach would be to post the NTSB accident summary, a "Causal Factors" story from *AeroSafety World* or, if those are not available, only the most important points, especially causal factors.

Data overload also can be quantitative *or* qualitative. For example, as part of its new SMS, an airline began collecting narrative hazard reports from its large workforce. Before the SMS existed, there were few, if any, reports submitted. For the first year of the SMS, the airline received only 26 hazard reports. Due to the low reporting, the airline, in the second year, decided to put much more emphasis on hazard reporting. In the second year, there was a precipitous spike in reports (267). The safety manger was overwhelmed and was not able to process all the reports, and a large percentage of those reports contained "sneak peek" information — an inside look at the hazards. In this example, the quantitative data were the number of reports received (measurable and comparable), while the qualitative data were the sneak peeks (textual descriptions of hazards).

Was there something different about the employees who participated in the survey compared with those who did not?

composite Illustration: Susan Reed; pizza base: © Jara3000/Vectorstock

| 49
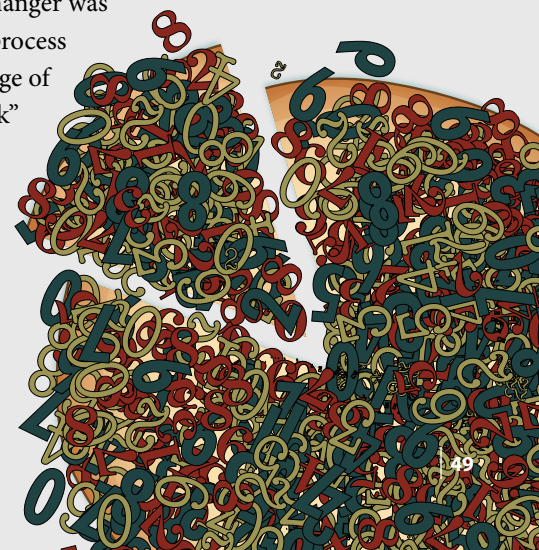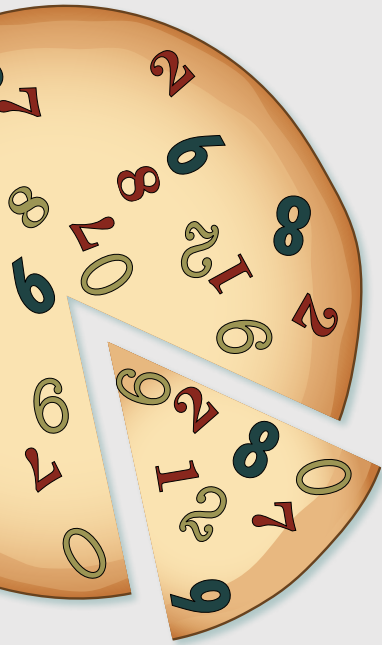
Because of this data overload, reporters quickly lost trust in the system because their reports were not acknowledged. Hence, managing hazard reports should be given high priority.

## Data Shortage

Many times, safety managers and upper management do not see eye to eye about safety expenditures. It is frustrating when upper management disapproves requests to spend financial resources for a safety improvement that you know is needed. This may be due, in part, to the safety manager not having cost-benefit justification for requests. It happens all the time, and because of this, safety may have to be thought of as a "case" or "argument," to persuade management to approve the allocation. You are misled if you think you will be able to walk into the CEO's office and get a quick sign-off on your new safety equipment request simply because you are a good salesperson. The question, then, is why might an astute safety manager lack the necessary data to present a logical case to management?

First, it may be the result of simply not knowing how to mine data. Choosing the right methodology to collect and analyze data (while avoiding GIGO) is imperative. To start, you must ask yourself what type of data you need to collect. Will they be numbers (quantitative), words (qualitative), or both? Will you be collecting data from the entire workforce or a sample? What types of data collection instrument will be used (questionnaires, surveys, test scores, interviews, focus groups, etc.)?

Once the data are collected, how will they be analyzed? Will your quantitative analysis use basic descriptive statistics (which represent a specific study group only) or inferential statistics (which can be generalized to the broader population)? What type of software will you use for the analysis? A standard spreadsheet program will work fine in most cases, but for more complex statistical analyses, you may need a program with more specialized functions.

For qualitative data, how will you sort through the hundreds or thousands of pages of text? Some software programs simplify this process by categorizing responses with keywords. Data collection is a structured process that requires good planning, a proven methodology and effective time management to yield valid results.

Second, the safety manager may not think that data need to be mined. Quite often, people use unstructured, personal observations as data sources. They develop a hunch about something and then try to sell it to management as a verified issue. While this method makes data collection simple, it has little value.

For example, the other day, a ramp worker at a major airport passed by a large paper cup on the apron. He noticed it but did not pick it up. Is that conclusive evidence that lack of foreign object debris awareness or a prevention problem prevails among all or many ramp personnel? Certainly not. But it does lend itself to a hypothesis, which can be tested, and for which the results can be presented to management as a basis for any interventions that might be required.

Third, good data may exist, but the safety manager chooses to ignore them. For example, an airport safety manager is collecting bird strike data as part of a new wildlife risk mitigation program. The manager is comparing bird strike data from before the implementation of the mitigation program (pre-measure) with data from after the implementation (post-measure). However, the data are not incorporated with study results (or data) from other, similar airports that have implemented a similar program. External data are very important not only for reference and comparison but also for benchmarking purposes. Think of it in two ways, "How are we doing?" and "How are we doing compared with other airports?" Use safety metrics to set objectives, goals and targets. Do not ignore relevant, easily obtainable data.

Data delirium can be treated, and the treatment is usually successful! ☕

*Robert I. Baron, Ph.D., is president and chief consultant of The Aviation Consulting Group. He has more than 25 years of experience in the aviation industry. His specializations include human factors, SMS, crew resource management and LOSA training/program development for aviation organizations worldwide.*