Session No. 663

Creating and Utilizing a Functional Root Cause Analysis Database

Capt. Norman H. Black C.S.P.
ESH and Marine Compliance Manager
Bechtel Australia
Gladstone, Queensland

Brian J. Vernetti C.S.P.
ESH Specialist
Tri Alpha Energy
Foothill Ranch, California

Introduction

The ultimate purpose of an incident investigation is to determine who is going to do what by when so the incident is not repeated. In order to arrive at the specific action items, factual information is needed to determine the required corrective actions. The more complete the information, the easier it will be to determine the corrective actions. However, the root cause of the incident is typically not found by simply ending the investigation after the first cause is identified. Five whys must be asked in order to drill down to the true cause. Determining exactly what information is needed in order to identify a root cause and effectively managing the large amount of data produced is greatly simplified using today's software. This paper will focus on using Microsoft Access to manage incident data as a means of data mining.

Data mining is the process of analyzing data from various perspectives and developing useful information. This information can then be used to clearly understand the cause of the data, in this case an incident. Whatever software is used to perform the best data mining job possible, the end user needs to be able to examine the accumulated data from several dimensions, categorize it, and conclude with a table of relationships directing the user to causes or patterns of the incidents.

The Concept of Data Mining

Although data mining is a relatively new term, the technology is not. Companies have used powerful computers to sift through volumes of supermarket scanner data and analyze market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.

For example, one Midwest grocery chain used the data mining capacity of Oracle software to analyze local buying patterns. They discovered that when men bought diapers on Thursdays and Saturdays, they also tended to buy beer. Further analysis showed that these shoppers typically did their weekly grocery shopping on Saturdays. On Thursdays, however, they only bought a few items. The retailer concluded that they purchased the beer to have it available for the upcoming

weekend. The grocery chain could use this newly discovered information in various ways to increase revenue. For example, they could move the beer display closer to the diaper display, and they could make sure beer and diapers were sold at full price on Thursdays.

How does this apply to root cause analysis? As it relates to the information gathered concerning the causes of an accident, the related data needs to be sorted and categorized to develop trends that would indicate repetitions of a possible fault in a system. With any database, the quality must be high in order to have confidence in the results of the analysis. The best and highest quality databases have the following traits: data, information, and knowledge:

Data - Data are any facts, numbers, or text that can be processed by a computer. Today, organizations are accumulating vast and growing amounts of data in different formats and different databases. This includes operational or transitional incidents, such as planned production or work patterns; nonoperational data, such as outages or maintenance periods, or perhaps supply chain incidents; and, Meta data - data about the data itself, such as logical database design or data dictionary definitions.

Information - The patterns, associations, or relationships among all this *data* can provide *information*. For example, analysis of an incident can yield information on which processes tend to be involved in more of the incidents and when.

Knowledge - Information can be converted into *knowledge* about historical patterns and future trends. For example, summary information on incident trends can be analyzed in light of safety initiatives to provide knowledge of what initiatives have negative and positive returns in order to prevent an incident in the first place. Thus, a safety professional could determine which process steps are most susceptible to prevention of incidents.

Defining the Data

While large-scale information technology has been evolving, worker's compensation databases and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in incident causes based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought.

Classes: Stored data is used to locate data in predetermined groups. For example, a company location or plant's incident data could mine employee incident by day of the week or product line being produced. This information could be used to produce targeted corrective actions.

Clusters: Data items are grouped according to logical relationships, employee habits, or work patterns. For example, data can be mined to identify employee titles/grades or incident type by causal factors.

Associations: Data can be mined to identify associations. The Monday morning back injury trend is an example of associative mining.

Sequential patterns: Data is mined to anticipate behavior patterns and trends. For example, a particular weather pattern could predict the likelihood of certain injuries where colder temperatures may bring on more muscle strains and sprains.

Data mining consists of five major elements:

1. Extract, transform, and load transaction data onto the data warehouse system (PC based)

2. Store and manage the data in a multidimensional database system

3. Provide data access to analysts and safety professionals to scrutinize

4. Analyze the data by application software

5. Present the data in a useful format, such as a graph or table

What is the purpose?

An incident may be viewed or at least considered a quality fault; a fault in a system that has allowed an injury, property damage, or near hit. Although quality is not necessarily a concept used in economic theories, much research has been published describing the direct and indirect cost of incidents. If one speaks of “operational excellence,” it is considered the excellence can be measured in one of several descriptors, one being “a safe operation.” Surprisingly, in today’s business climate, as much as an organizational incident rate is compared against its peers, operational effectiveness is also measured by the rates of safety improvement as defined by fewer incidents. This measurement is one of continuous improvement indicating a company’s ability to adapt to change, also considered *innovation.* By utilizing an approach to improving quality by means of data mining incident occurrence data, a user is exercising innovation, provided something is done with the data to prevent future incidents. Therefore, seeking patterns of incidents should be the aim of any data mining exercise.

Although two incidents rarely happen in exactly the same manner, incidents do follow a general pattern. These patterns display common features among select groups and allow the SHE professional to use a statistical approach to incident analysis.

For example, knowing that 40% of an organization's incidents involve ladders is not useful. Furthermore, knowing that 80% of the ladder incidents involved broken rungs still does not provide long-term "correctable" actions. However, the ultimate reason for broken ladder rungs can be determined if after asking five why’s it was found that employees are being forced to routinely carry or lift 75 pounds while positioned on the ladders. This root cause information allows for an accurate correction of the issue contributing to an accident.

There are numerous investigation techniques and software packages available to the SHE professional. Based on the SHE professional’s particular needs, choosing the right one to use can be confusing. Some are technically complex and others are nothing more than a means to document a possible worker's compensation case. However, drawing from ANZI Z16.2, *Information Management for Occupational Safety and Health,* standardization of the incident queries is possible that lead to a systems approach to accident investigation. Coupling this standardization with Microsoft Access allows the SHE professional the ability to easily write queries for the database being maintained.

By combining the essential elements to be examined in a root cause analysis with a manageable database management program, even the novice computer user is able to identify those crucial incident causes that must be corrected. Once these corrective actions have been implemented, similar incidents are far less likely to occur moving forward.

Data Mining and Incident Root Cause Analysis

To build an Access database, first you need to create a plan. Planning? “I thought Access would do all the planning for me!” Well, no, in fact without sketching out a game plan on paper, your tables can get messy. Planning the setup of a database before actually building it will save you tons of time. It will also keep you from wanting to curse the bloody soul who created the dreaded Access program to begin with. A good sketch and brainstorm session prior to plugging away at data entry will help tremendously.

Access is the same as real life - bittersweet. How so? We all desire fulfilling and successful real life relationships, however from time to time they can seem distant and difficult to bring together. Managing large amounts of data presents the same problem. That is where Access offers its unique purpose and separates itself from Excel. It is a relational database that allows you to tie different groups of data together with the click and drag of a cursor. Now, only if our own lives were that easy!

To begin solving our problem, tap your inner Van Gogh and draw boxes on paper (you can also type them out in Word or Power Point, whichever is easier) for each table you want in your database. Fill in the boxes with lists of fields that you want in each table. Then simply draw arrows from a field in one box to a field in another that it needs to relate to. This involves some designing, determining what data you want to store, and how you want to structure it. Ask brilliant questions like: What is the goal of the database, who will use it, and what tasks should they be able to perform? Once you think about who may be using the database at your organization, you will know how user friendly it needs to be. Please see attachment 1, “Planning Sketch for an Incident Database” for an example of how to sketch boxes and fill them in with fields.

Now that your boxes are sketched on paper or typed out, we will want to formulate some questions that we want our database to be able to answer. These, in effect, will be the queries that we will create. For our root cause analysis database at XYZ Company, below are some of the questions we will want our database to answer.

1. What are the immediate and contributing causal factors of our incidents?
2. What departments are suffering the most incidents?
3. What are the corrective actions for each incident, who are they assigned to, and have they been completed?
4. How many reportable vs. recordable incidents have we had?

We have now done all the legwork necessary to build our database without ever opening Access. We can now take our artistic masterpiece from paper and plug it into our not so scary friend, Access. Simply open a new file in Access and create a container that will hold all your tables. Then create all your tables and plug in the data fields from your boxes.

These tables are also called Normal Forms. Your database will have multiple normal forms and they must all follow the structure of the First Normal Form. To create the First Normal Form, you must first have a Primary Key, which is similar to a column header in Excel. Then below the primary key, you have attributes of that column broken down into the smallest meaningful value. It is important to remember not to have repeating groups of data as these can lead to null values in

your tables. All of the following forms must follow the same structure, and their respective fields must describe the Primary Key. If you have fields that are not describing the Primary Key of that form, simply move that field to another table. To relate the tables to each other add the Foreign Keys, which are Primary Keys from another table, to the table that you want to relate.

Before you start entering or importing data into your table fields, an important step is to create relationships between the tables. These relationships can be one-to-one or one-to-many. Access is not exactly the same as real life where MOST relationships are one-to-one. In fact, most relationships in Access are one-to-many. In one-to-one relationships, you join the exact same fields in two different tables that you want to relate to each other. In one-to-many, you join a primary field from one table to a field in another table that can have multiple values. To create relationships between your tables, simply click on the Database Tools tab on the top ribbon. Access has already created some relationships for us by connecting common fields in different tables. We can modify or add to these relationships by highlighting a field in one table and dragging it over to the field we want to connect it to in another table. The result will look very similar to attachment 1, since that is where we first started forming our relationships.

Now that our tables and relationships are established, we can enter data into the fields of those tables. This can prove to be rather cumbersome and time consuming. Not to worry, Microsoft uniformity to the rescue! You can import data from Excel spreadsheets, Outlook, Text Files, HTML documents, or SharePoint lists into Access with ease. You can even add a field to your tables with attachments as the file type. For instance, this would allow you to attach incident reports if they are written using Word or some other program besides Access. You also have the ability to create a recurring import. If your organization updates and shares files in Excel, Access allows you to set a recurring update reminder in Outlook to prompt you to import the updated Excel spreadsheet into your database. For instance, if completed corrective actions are recorded by department managers in a shared Excel file, Access can set a recurring task in Outlook to trigger you to perform an updated import.

The final set-up is creating queries, forms, and reports based on the relationships of our tables. Queries locate and retrieve data from the tables for viewing and analyzing. They also make data entry easier by limiting movement between tables to add or modify records to each table. A query can add that record to all the associated tables for you. A query will connect and display information from the tables you desire and allow you to build a data entry form off that query. Forms allow the users to enter information into the tables already created. They can be customized to be user friendly. Furthermore, if you want to print a report based on the same information, you can take that query and build a beautiful customizable document with which you can “wow” the senior executives. In short, queries allow you to build forms and reports that can be uniquely formatted to meet your organizational needs.

The queries that we will create for XYZ Company will allow us to answer the four questions we posed above. To do this, you need to click on the Query Wizard button under the Create tab on the top ribbon. Using your wizardry skills, select the tables you need to draw from to answer question one. Then select and drag the fields you want to appear in your answer to the query to the construction space below. Once you have done this, click Run and watch out, because you just performed some magic! Do the same thing for each of the questions and you are well on your way to better managing your Incident Program.

Conclusion

An investigation that produces tangible results that culminate in actions that eliminate the root causes of incidents is a product of multi-layered statistical analysis. The SHE professional must first understand and define the classification and system steps that lead to an incident. The professional may effectively manage the data produced using today's commercially available data management tools. This will allow the SHE professional to develop their own action plans that drive organizations to zero incidents.

References

Ackmann, Heather. 2010. *Access 2007 Training*. Schaumburg, IL: Train Signal, Inc.

Adèr, H. J. & G. J. Mellenbergh. (with contributions by D.J. Hand). 2008. *Advising on Research Methods: A consultant's companion.* Huizen, the Netherlands: Johannes van Kessel Publishing.

Asian Development Bank. February 2009. “Five Whys Technique" http://www.adb.org/publications/five-whys-technique. (Retrieved 26 March 2012)

ASTM International. 2002. *Manual on Presentation of Data and Control Chart Analysis*, MNL 7A. West Conshohoken, PA: ASTM.

Bureau of Labor Statistics (BLS). 2012. “Total recordable nonfatal occupational injury and illness incidence rates by employment size, private industry, 2007-2011.” http://www.bls.gov/news.release/pdf/osh.pdf (Retrieved 31 December 2012)

Cook, Ken & Ulrich Fuller, Laurie. 2010. *Microsoft Access 2010 For Dummies*. Hoboken, NJ: John Wiley & Sons, Inc.

Ivanov, K. 1972. "Quality-control of information: On the concept of accuracy of information in data banks and in management information systems." Stockholm: The University of Stockholm and The Royal Institute of Technology. Doctoral dissertation.

Juran, Joseph M. & A. Godfrey. Blanton (1999). *Juran's Quality Handbook*. 5th ed. New York: McGraw Hill.

Lewis-Beck, Michael S. 1995. *Data Analysis: An Introduction*. Thousand Oaks, CA: Sage Publications, Inc.

NIST/SEMATEK. 2008. *Handbook of Statistical Methods*. http://www.itl.nist.gov/div898/handbook/2008

Pyzdek, T. 2003. *Quality Engineering Handbook.* Boca Raton, FL: CRC Press.

Tabachnick, B.G. & Fidell, L.S. 2007. *Using Multivariate Statistics*. Fifth Edition. Boston: Pearson Education, Inc. /Allyn and Bacon.

Veryard, Richard . 1984. *Pragmatic data analysis*. Oxford : Blackwell Scientific Publications.