Upgrading Risk Analysis? Don't Forget The Data!

By **Nick Park, GeoFields, Inc.,** Atlanta, GA

Pipeline-risk assessment has been well explored from the intellectual perspective of understanding relationships among factors that cause — and determine the consequences of — pipeline failure. Historically, this process has involved the collection and integration of a large amount of data, both in terms of the numerous types of information needed and their extent across the entire pipeline system.

Unfortunately, pipeline operators have lacked the detailed data management frameworks necessary to make risk analysis a streamlined process. The significant volumes of data often had to be collected together, reformatted and then loaded into a risk-model application. Any changes or updates required a repeat of this process. In the worst cases, a risk model user had to retype data into a series of screens — often taking considerable amounts of time — and with the likelihood that data translation errors would occur.

Additionally, as first-generation applications, many risk-model software packages hard-coded risk-analysis algorithms and were inflexible in how data should be formatted in an analysis. The results have been analyses that are detached from their true data sources and users swamped with unending reformatting tasks.

The amount of research and expertise invested in pipeline risk modeling is significant. However, what has been missing is that risk models that have been developed need a home - an application framework in which to manage the data and run the model. Modern pipeline data management systems and risk-modeling applications now provide these capabilities.

Rather than asking for data to be preprocessed and moved into an independent store, thus removing it from its source and causing many months of exhausting data manipulation, the trend is toward data-centric risk assessment — for applications to read data from a wide range of existing data stores and in a range of formats. Processing occurs on the fly, both registering the data to the pipeline and creating the necessary segmentation for analysis.

Moving to an integrated data-management approach, users assessing risk at pipeline companies can now start to think about how their data can be organized and standardized, and as a result, better manage data. And, they get back to what really improves risk analysis: better data and a better understanding of the interaction between data and its impact on risk.

Gaining Confidence

Consumers of risk-analysis output must be able to understand the results of the model and have confidence in them. There are three components that lead to this confidence:

1. The first is that the model and its output must be feasible. The model must be realistic and something that can be implemented. Perhaps most importantly, there must be a plan for making the data available.

2. Secondly, the approach to modeling must be systematic. Users must be able to show the process, detailing how the model's inputs were prepared and how the model was run. The algorithm must be developed in a repeatable and justified way. Any two runs of the model with exactly the same data and the same algorithm must produce the same result.

3. And thirdly, the results must be defensible. Any result must be justifiable by looking back to the source data and reviewing the algorithm that processed it.

Feasibility, a systematic approach and defensible results come from the risk assessment being transparent. By basing the modeling process on a data management system, users can run the model with confidence, assured that the data is of known quality and known format.

When an algorithm is open to iterative review as new data is available and when new interactions between data become known, it ensures that confidence in the process will increase. By being transparent and having a dynamic process, more experts can be involved and the way the model works will be easier to follow and understand.

Model Development

Be data-centric in risk-model development. The classic risk-modeling approach called for the development of a model based on observed and predictable interactions between factors in specific situations. The field of risk analysis works on the following theory: if all information is available, an accurate assessment of risk can be made. The less accurate the data, or the fewer inputs we know about, the less confident we are about our analysis of risk.

In the world of nuclear power stations, this approach to risk analysis works because there are vast repositories of data. Many people have conducted expensive data collection efforts, coupled with detailed material analyses and process modeling, to understand the behaviors of systems and to be able to predict failure. The modeling of an absolute probability of failure, coupled with the consequences based on the nature of the environment, and presumably the severity of the incident, allows a real measure of risk.

Pipeline operators, however, do not have access to such vast repositories of data, nor has the industry yet had the resources to understand all the complex interactions between the present threat inputs. Significant work in this area has been done, but often the modeling efforts are thwarted by insufficient data or data of low quality.

Many pipeline operators are still working toward an accurate electronic basemap of pipeline location — a key dataset to establish interaction with the surrounding environment and integrating a range of datasets relating to pipeline threats. As such, the goal of probabilistic risk assessment is real, but for many, a long way off.

The approach to date in the pipeline industry, called a model-centric risk assessment, has been to evaluate risk by producing scores for individual threats and consequences, and then to sum up scores for an indexed ranking of pipe segment risk. The data requirements for this approach were frequently addressed using simple data stores that were managed by basically reformatting the data externally and then reloading the risk application.

Data-Centric Risk Assessment

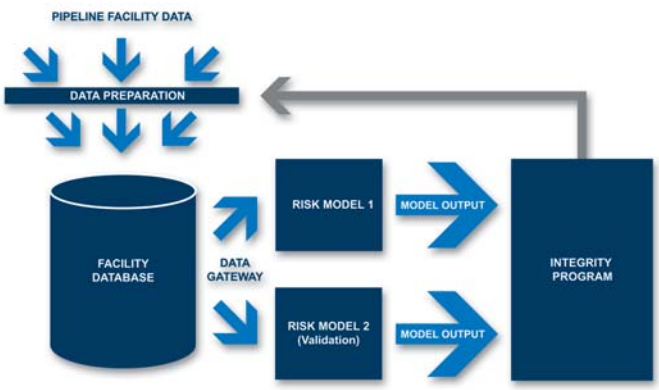

Diagram 1: A data-centric approach to risk assessment.

Today, many operators are moving from an index-based risk model to one that starts to look at how contributing factors interact to influence risk. This approach demands further improvements in data management and requires data that is readily available, both for the research phase where the model's interactions are determined and then available when an analysis is run.

It is important to make sure that data is used in the model determination for two reasons. First, the relationships must reflect real interactions in the field and the nature of relationships must be derived from data coming from field events. Evaluations of the causes of failures that have occurred are key - both at an operator level and across the industry.

Secondly, it is important to make sure that a risk factor is being developed that has at least a modest chance of finding suitable data to allow it to contribute to the model. If a risk factor is developed without regard for the available data, there might be significant mismatches between what the model needs and what the data can deliver. The way to avoid these potential pitfalls is to be data-centric in the modeling approach — to consider the data during model development.

Being data-centric in risk-model development has other benefits too. The data will become better managed for the benefit of the whole organization. When you worry about the data, where it came from, what it looks like and how good it is, you tend to want to manage it better. It is no longer acceptable to receive 30 spreadsheets from all the corrosion engineers. You start to think about how their data can be organized and standardized.

Model-Centric Risk Assessment

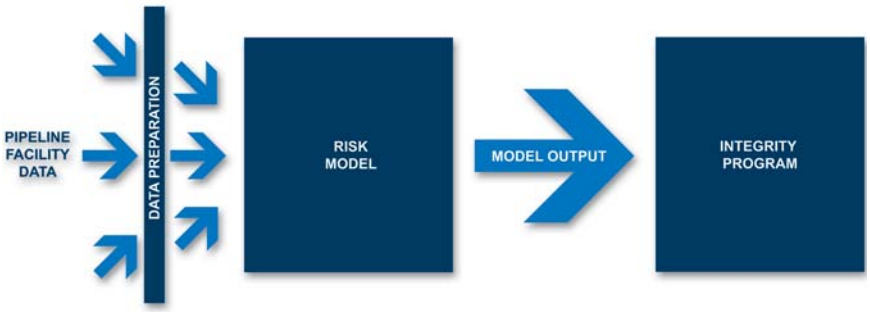

Diagram 2: The classic model centric risk assessment.

Data Management

What does data management mean? Data management is often confused with data storage. This confusion can hamper its proper adoption. Data storage is simply the means by which data is stored. It defines a format for the data and a location, often a database on a server within the company. Storage is important — without it, data management is impossible. But the database server and the data format don't describe the processes through which the data gets into the store, nor what happens during processes such as update, addition and deletion. Those come about with a data management framework.

In data management, the focus is on policies and procedures which are implemented through software. Effective data management software for pipeline companies understands the format of the data — often based on industry-standard data models — as well as the changes and additions that a pipeline company needs to make. Functions of the data-management system should include:

- Adding new pipeline sections,
- Changes in pipeline status (such as planned, active, idle, abandoned, removed),
- Pipeline reroutes, whereby pipe sections are idled and new pipeline sections added in,
- Modifying the pipeline location based on better survey data or a better basemap, and
- Managing the facilities that make up the pipeline system.

The key to successful data management software is that the edits are performed and the software tracks all the necessary changes through any dependant data. For example, if a pipeline section is retired, so are all the components that exist on that section. If a pipeline's location is updated to reflect a new GPS survey, then all the features that lie on the line must remain in their correct locations on the pipeline.

Further, data-management software must ensure that edits are tracked through an approval process audit trail. It is important not only to know when and what edit occurred and by whom, but also what the edit history of a specific feature is. It is important to ensure that errors in the data are identified and managed correctly, providing fixes where appropriate. For example, an edit to a landowner record as the pipeline passes from property to property must ensure that there are no gaps or overlaps in the data. The user must be presented with such errors when they occur and provided with appropriate fixes.

Legacy risk-modeling applications have been strong on algorithms because they evolved from the work of thorough and well-qualified authors. However, solving the algorithm is only half the problem. The data inputs were left to rudimentary desktop databases and often required a lot of data preprocessing to load, or worse, having to type inputs into screens over and over again.

When users have a mature data-management system, confidence in the data expands greatly. A complete data-management system has benefits beyond risk analysis, bringing a positive impact to the workflows of users in a wide range of departments. There is clearly enough justification to implement one. However, the risk analysis team alone — one of the primary users of the vast array of corporate data — can make a powerful case for a corporate data-management system.

Major Benefits

There are five major benefits of using a data-management system and an open, data-centric risk analysis application:

- Gathering and formatting information will be more streamlined. A data management system provides a place and a means for data owners to store and work with their data, often enhancing their own workflows while making the data available to others in a consistent and reliable way;
- The analytical environment improves because data is in a known format, from a known location, and of known quality. This has a measurable impact on the reliability of the output from the model;
- Risk-analysis scenarios become easier to run because of the iterative nature of the process. Data-centric modelers are used to evaluating data while developing risk models, and so when the model changes, or as new data is available, they are able to iterate through existing processes to evolve the model and incorporate the new data;
- Documenting the model becomes easier because extensive written material is not needed to describe the complex pre-

Model Centric vs. Data-Centric Risk Assessment

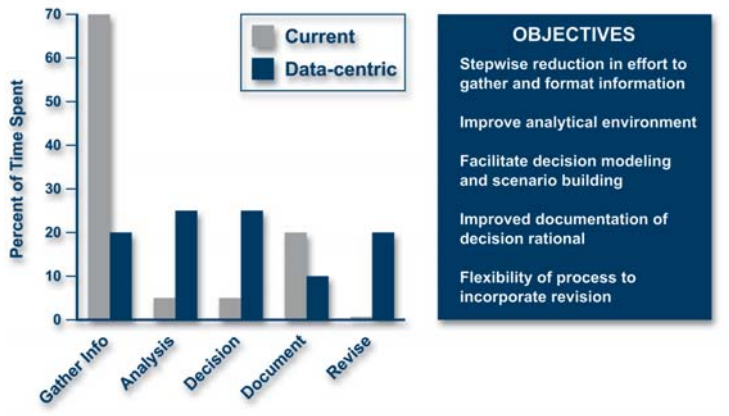

Diagram 3: With data-centric models, valuable time is focused more on critical risk analysis than tedious gathering, formatting and documenting information.

processing of data. The best documentation already exists from the data owner, describing how they store and integrate their data into the corporate system; and

- The process can incorporate new revisions, both to data format and to the algorithm being used. This is important as change is inevitable, both in the nature of the data that supports the model and in our understanding and sophistication in the model itself.

Conclusion

Risk modeling is a difficult and complex process that doesn't end when the baseline assessment is complete. Inevitably, the model will need to evolve and new data will come in from a wide variety of sources. By moving to a data-centric risk assessment that focuses on the data as much as the finer points of the model, significant efficiency benefits can be gained. And, users can reduce uncertainty in the risk-modeling process by taking advantage of modern data-management practices and systems, focused on their needs as pipeline operators. ***P&GJ***

Author: Nick Park *is vice president, product management, for GeoFields, Inc. He holds a Master of Science degree in Geographical Information Systems from the University of Leicester, and a Bachelor of Science degree from the University of Leeds, both in the United Kingdom.*