

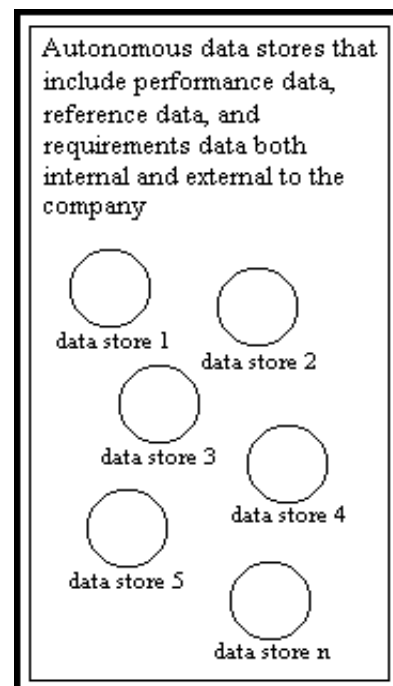
# Information Management through Desktop Integration

## Introduction

The desirability of utilizing a number of different data sources in an integrated environment when analyzing facility operations has long been recognized. Integrated reports that utilize all appropriate information sources are the most effective in supporting the needs of management in evaluating their ES&H situation in a global context. Using this approach, limited resources can be allocated in an effective and proactive manner.

There are, however, a number of obstacles that have proven difficult to overcome in past attempts to perform an integrated analysis. These are primarily related to a number of different characteristics of the data sources such as the data structure and accessibility considerations.

- Data stores are generally autonomous, with each store utilizing its own individual language, logic, and structure. Lack of inter-data source compatibility degrades analysis quality and limits the extent to which information can be integrated (**Figure 1**).
- Data stores have limitations related to reporting timeliness, reporting thresholds, data integrity (e.g., consistency among reporting organization), data granularity, and reporting completeness. This will be seen both between data systems and between organizations within a single data system.
- Many data stores are not directly accessible to the analyst. They reside in local databases that are accessible only through direct connection to the host computer or, in some cases, via internal networks (intranets).



**Figure 1**

Because of these characteristics, the following limitations have often been encountered when trying to perform an integrated analysis.

- The process of locating and extracting relevant information from each individual database is difficult. Many of the sources utilize their own proprietary software to access the data, thus requiring that the analyst not only have access to each software

package but also to be proficient in the use of the numerous different packages.

- It is very time consuming and tedious to correlate and extract data to support proactive response. Much of the correlation of the data from the different sources must be performed manually.

Although the ideal analytical system would include a number of different data stores that have been carefully coordinated during design and implementation, in most cases this will not exist in the DOE environment. Much of the data resides on legacy systems, with historically established structures and order-defined reporting requirements that are not easily changed.

However, advances in technology, coupled with the migration of existing data stores into a new environment, provide an opportunity to develop new methods of accessing information that will help to overcome these limitations. As legacy systems are ported to more modern platforms, many of the interface differences can be resolved through adherence to a few basic design requirements. In addition, although many of the differences in structure and reporting requirements will remain, within the modern computing environment that exists today many of the difficult and time consuming tasks that required manual execution in the past can now be automated and integrated at the desktop.

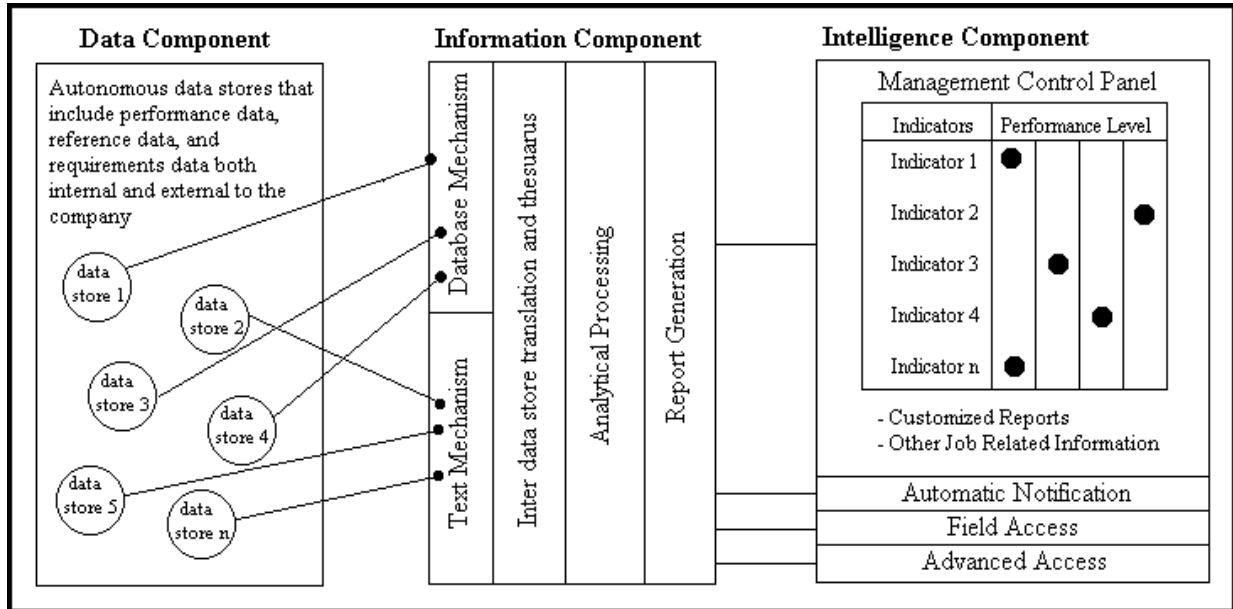
## **Desktop Integration Model**

**Figure 2** displays the desktop integration interface. This interface includes the mechanisms for actually acquiring the data from the individual data stores and the various processing functions that are performed on the data in order to integrate the data sources and provide an output in the desired format. This would include items such as the thesaurus for text based searches, the translation matrices (organization, event categories and binning, etc.) for combining data from multiple stores, and the algorithms for normalizing, weighting, and combining the data into an integrated performance indicator.

Vital characteristics of the desktop integration model are as follows.

- While the specific data may take a variety of different forms, the basic data format must meet minimum access requirements (i.e., it must support querying by a broad based text search engine and/or an ODBC compliant application).
- Addition of accessible data stores should be a seamless process.
- Data sources must be accessible from the desktop. It should not be necessary to create copies of each source in order to meet the geographic and or basic format requirements.
- Although summarized data might be displayed, drill-down to the lowest level of information must be available to provide traceability.
- Dissimilar data stores are related through inter-modular translation devices (e.g., translation matrices). Translation will permit meaningful comparisons between organizational categories, causal factors codes, and other fields.

- The level of detail provided within the data sources must be sufficient to support the desired level of granularity.
- Text based searching should be based on a thesaurus concept.



**Figure 2 - The Desktop Integration Model.**

## User Levels

Effective design requires the recognition that different types of users require different types of data access in order to satisfy their needs. Therefore, the required basic functions of the interface will vary depending on the user type. These types, or levels, of access can be broadly grouped into three categories.

### High Level Access

This is a high level overview of the information. This level is typified by management or support type personnel who need quick and easy access to specific, well defined data. This type of access is characterized by the following features.

- The output is generally highly summarized.
- The system can access data from multiple sources to provide a complete view of an issue.
- Any required translations for integration of the data sources are completely transparent to the user.
- To a limited extent, queries are customizable by the user, or their support staff, to provide the ability to focus on a particular issue or at a particular organizational level.
- The system is easily accessed with no required training in system operation.

- Information displayed includes drill down capability to evaluate major contributors to observed outliers.

Some of the high level requirements for this type of user are summarized as follows.

- A single interface is presented for all data sources. In other words, the data sources are invisible to the user.
- Data retrieval and correlation is based on an inter-data source thesaurus and/or translation matrix.
- The system can be customized to allow the user to select from a number of different issue driven performance indicators.
- The user is allowed to set additional pre-defined performance indicator parameters (organization, classifications/keywords, causal factors, date).
- The system allows customization of default parameters for “single-stroke” command or time-activated command.
- Detail data is tagged for traceability purposes. If trends or other items of interest are noted at a high level, the ability is provided to drill down to the source data in order to obtain additional information.

### **Intermediate Level Access**

This level of access has medium level search capability. This level provides additional, more specific, information using simple, easily learned search techniques and pre-defined output reports. The interface associated with this type of access is envisioned to be similar to that currently being provided in many of the existing data systems as they are ported into a new environment, but with enhanced capability for integrating results from different stores. This type of access is characterized by the following features.

- The output is generally in the form of pre-defined reports that present the data in an easily understandable form. However, actual interpretation of results will depend, to a large extent, on user understanding of the results and system processes.
- The system can access data from multiple sources to provide a complete view of an issue.
- Queries are easily customizable to provide evaluation of a broad range of issues and/or organizational levels.
- In general, translations between data stores should be automated, however, the user should have the capability of specifying additional, lower level, comparisons to the extent that the data permit.
- System operation is straightforward, with only limited training required to utilize all capabilities.
- Information displayed includes drill down capability to evaluate major contributors to observed outliers.

Some of the high level requirements for this type of user are summarized as follows.

- Provisions are made to limit retrieval to a subset of the data sources.
- Data retrieval and correlation is based on an inter-data source thesaurus and/or translation matrix.
- The user is allowed to set performance indicator parameters (organization, classifications/keywords, causal factors, date).
- Detail data is tagged for traceability purposes. If trends or other items of interest are noted at a high level, the ability is provided to drill down to the source data in order to obtain additional information.

### **Low Level Access**

This is an advanced, low level search capability access. Because of the wide variation in requirements for this type of access, it will most likely require direct access to the data sources, independent of any fixed format interface. In general, the overall processes associated with this type of access will not lend themselves to a high degree of automation. This type of access is characterized by the following features.

- This type of access requires that the user possess a high level of knowledge of the systems and processes involved as well as of the structure and content of the specific databases being utilized for the analysis.
- This type of access will generally consist of a highly specific search capability with output to standardized reports or transfer of raw data to external analysis packages for further processing.
- Although translations between data stores will normally not be automated, translation matrices, classification bins, thesaurus information, etc., should be accessible to the analyst.
- This type of access assumes a high level of user knowledge of the systems and processes (software) used to extract the data from the stores. These will vary based on user preferences and standards and, in general, training in the use of these techniques will be the responsibility of the user's organization. Support by organizations maintaining the various data stores is not assumed beyond areas such as the identification of database schemas.

Some of the high level requirements for this type of user are summarized as follows.

- Provisions are made to limit retrieval to a subset of the data sources.
- Data retrieval and correlation is based on an inter-data source thesaurus and/or translation matrix.
- The user is allowed to set pre-defined performance indicator parameters (organization, classifications/keywords, causal factors, date)
- Provisions are made for advanced query capabilities

## **Advantages of Desktop Integration**

Desk top integration has a number of distinct advantages over the more traditional methods of obtaining and analyzing data. These advantages can be divided into those that effect the data access and those that affect the data utilization.

### **Advantages related to data access**

1. The user has access to all necessary and appropriate data stores directly from the desktop. The need for copying or manual entry of data is eliminated.
2. Additional data sources are easily added to the analytical system.
3. The system is platform independent. It will run on a PC, Mac, UNIX, or any other operating system that supports current browser technology.
4. There are no proprietary software requirements for either the client software or the data source. Any client software or data source that meets the minimum connectivity requirements (ODBC and/or Text Search capabilities) can be utilized.

### **Advantages related to data utilization**

1. The system is easy to use. Standard translation matrices, thesaurus entries, etc., are already incorporated into the system, eliminating the need for the user to independently develop them.
2. The system can be configured for multiple levels of access, ranging from higher level, fully automated presentations to low level access with a high degree of user control over the analytical process..
3. In the case of the higher levels of access, utilization of the matrices and other similar tools is automatically integrated into the system. No user interaction is required to correlate data from different sources.
4. The system provides easy drill-down to determine the source data for displayed trends.
5. The system is easily customizable to provide the exact information (organization level, issue, etc.) that a user is interested in.
6. The parameters utilized in trending and analyzing the data are customizable to user specific needs or preferences.