

HAWTIN, STEVE, Schlumberger Information Systems, Houston TX;
NAJIB ABUSALBI, Schlumberger Information Systems, Stavanger, Norway;
LESTER BAYNE, Schlumberger Information Systems, Stavanger, Norway;
MARK CHIDWICK, Schlumberger Information Systems, Stavanger, Norway

The Spectrum of Data Integration Solutions: Why You Should Have Them All

Data Integration is a key element in the operation of today's Exploration and Production (E&P) companies. This paper argues that no single approach can address all the real integration needs of an E&P organisation, but that a range of technology solutions must be adopted in a data integration strategy. It describes one way to classify integration approaches and explains how this can help when solving a particular integration issue.

Why Integrate?

The first question to focus on is a fundamental one, "Why Integrate?". Since proper integration is expensive what are the benefits that justify these cost?

Increasing exposure of the data increases the level of expertise viewing and manipulating the data

More errors in the data are detected and corrected

Utilization of data in the workflow increases

Quality and trust increase

Business risk decreases

The process of integrating a particular category of data can be thought of as a progression. Initially the organisation has only a single source version the data. Over time more sources become available, these are brought together until perhaps a single comprehensive system is created that relates these multiple "variants" of the information and resolves the inconsistencies.

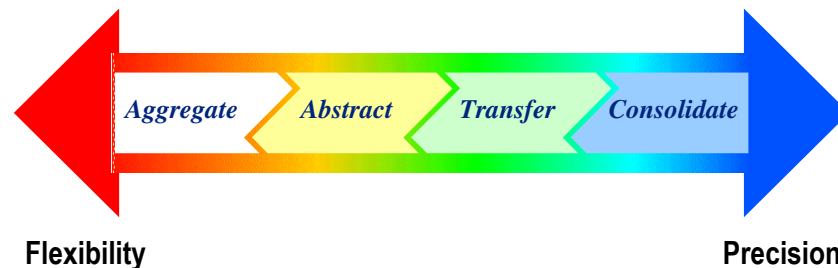


As the data moves from left to right, the "value" of the information to the organisation increases, but so does the cost of preparing it. This cost includes both the time as well as the effort taken to prepare the data. For many types of data, having the information available soon is much more important than checking consistency with other data sets. Often the cost of going all the way to a single integrated data source cannot be justified.

The benefits of some level of integration almost always outweigh the costs, so why is there an issue? Because there are various approaches to integration, each with conflicting claims made by their various supporting vendors. Each group has its own "integration" story. One may claim that GIS is the best integration tool while another may say that data is only "really integrated" when it is all placed in a single database. How can such different approaches to "integration" be compared?

The Integration Spectrum

There are a range of approaches to integration and also a number of ways to compare them. One useful classification scheme is to think of integration solutions as occurring along a “spectrum”.



On the left side there are approaches that merely display information from multiple sources, performing the integration only visually. On the right side, the data is integrated by the process of migration into a single consistent structure, simplifying further data manipulation by automated processes, and helping to ensure data integrity through business rules that are now enforceable.

This paper highlights the four groups of approaches along this spectrum that are in common use today: Aggregation; Abstraction; Transfer and Consolidation.

Visual Aggregation

Visual Aggregation brings the data together on the user’s display without attempting to validate it through business rules. For example GIS and “web portal” solutions allow the user to relate information from a range of sources that have no well defined underlying connection.

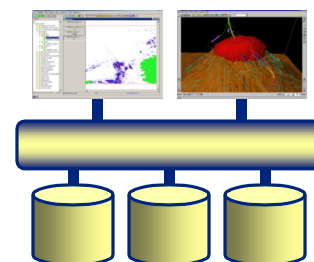


This approach has the advantage that it does not require complex processes to confirm data consistency before it can be applied. This usually results in a more up-to-date set of information being available, often a crucial element in the success of the solution. In addition it can provide a good mechanism for relating “higher level” business objects, without being slowed down by the “technical details” that can often make less flexible approaches more expensive to implement.

This flexibility comes at a price in that data that is integrated this way does not lend itself to further processing. While visual integration helps provide a high level summary it does not assist in enforcing consistency or in later automated processing, such as quality checks.

Abstraction

Abstraction provides an intermediate layer that isolates the applications from the details of information accessible from a variety of locations. This is achieved by summarising the actual data using a more abstract set of “business objects”.



In Visual Aggregation there are no constraints placed on the data that is accessible. Under Abstraction, however, it is important that there is consistency in the way

entities are identified in the various data sources. The system must be able to recognize two different renditions of “the same” real world business object.

Abstraction leads to integration solutions that are powerful and can be easy to install. Some processing on the data can be carried out, although normally performance requirements constrain this to the most commonly used attributes.

It is often complex to create the “Data Adaptors” that mediate between the abstraction framework and the external data sources. This is especially true where local conventions change the way that data is stored. In addition the generic object definition provided by the framework can never meet the needs of all possible applications.

Even when it is possible to run applications that can extract information from many possible sources there will normally be good business reasons for restricting how the abstraction system is used. For example, if the company’s business processes define the location where approved well headers are to be found then the abstraction system will be prevented from obtaining them from alternate locations.

One of the advantages of this approach is that the management of data is kept closer to the data sources. This implies that other tactics are required to fulfil the more traditional data management tasks.

Transfer

Transfer performs the classic extract, transform, load (ETL) sequence to copy data from where it is to where it needs to be. At the moment this is the most widely adopted approach to Data Integration and can support any complexity of data flow.



The main advantage of this approach is that it can apply any transformation to the data as part of the copying process. This is just as well, since the data must conform exactly with the restraints of the target before it can be successfully inserted.

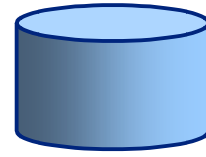
Transfer approaches can bring together data from a wide variety of locations, making this the only viable way to integrate certain legacy and proprietary data sources. For some kinds of data, such as off-line information, there is no option but to duplicate the data via a transfer.

The data being transferred can also be captured in a format-independent form for archiving and transmission to remote locations. The view of the transfer can be easily tailored to the end-user’s requirements, showing only those aspects that are important.

However, even with all these advantages, the transfer approach does of course lead to data duplication, with all the well-known difficulties that this implies.

Consolidation

The most precise form of integration is to consolidate all the data into a single repository. If information is available from a single consistent location it simplifies the task of subsequent processing as well as making logical groupings of data readily available for a wide range of applications.



This approach makes the later task of data management much easier. For example with a single trusted location the management of entitlements is simplified. By having a single “master” location the difficulties of data duplication are also eliminated.

The difficulty with this approach is primarily the cost due to: the effort it takes to initially implement; the complexity of transforming the data; and the time involved in carrying out the necessary data quality checks. In addition, there are categories of information that cannot readily be made to fit into a single repository in an effective way, leading to the holding of data in a form that does not readily meet the needs of applications or without some essential related or contextual information.

Integration as a partnership

A number of factors determine the success of an integration solution, not just the underlying approach being used, but also the degree of partnership between the solution provider and the end users.

Traditionally vendors have selected an integration technology and then attempted to discover the customer’s business processes that it addresses. This “push” approach requires a detailed understanding of the vendor’s tools, features and functions. It can be effective when customers are unfamiliar with the advantages of using a newly available technique.

A better alternative has been for customers to “pull” technology by asking vendors to advise on the best technology to address particular business issues. This relies more on a detailed understanding of the customer’s business processes and constraints, but vendor “bias” is normally an issue.

An even more effective way to arrive at successful solutions is to treat data integration as a partnership between the E&P company and a well versed solution provider. This requires a melding of talents to deploy the most effective tools for the challenges ahead. It is critical that you select a partner that is familiar with the range of technologies covering the whole integration spectrum. Such a vendor will be able to propose and help implement solutions to fit your needs rather than fitting their preferred technology to your requirements.

Conclusion

Integration improves exposure, and by extension, the value and quality of information to facilitate workflow and reduce business risk. It is an important element of the way that the E&P business process operates.

The various approaches that are commonly used today to provide integration can be classified within the “integration spectrum”. No single one of them addresses all the business needs of a customer, a complete solution requires a combination of all of

them. Each of these different approaches has different requirements. No single technology can hope to address the whole spectrum.

Trying to force fit an inappropriate favourite technology to an integration issue normally leads to an unsuccessful implementation. It is important to be aware of the complete range of options so that appropriate technologies can be selected to fit your integration needs.