Original Article

Business need, data and business intelligence

John Jordan

is a Principal Consultant in the Business Intelligence and Media Data unit within Siemens Global Media Consulting. He has over 25 years of experience in the broadcasting industry and has been involved in data architecture for the last 10 years. He currently leads the Siemens consulting team providing data architectural services to the BBC and has recently been involved with complex business change projects in Germany and South Africa.

Clive Ellen

is a Business Intelligence Consultant for Siemens Global Media Consulting. He has spent the last 10 years designing enterprise strength business intelligence solutions for companies such as British Airways and Aviva. He is now involved in developing similar solutions for the media industry as part of Siemens Broadcast Media BI offering.

ABSTRACT The critical value of an organization's data is becoming widely acknowledged within the media industry. This paper outlines a robust and cost-efficient way to enable the development of an enterprise's data environment – the adoption of an enterprise data architecture. The data architecture is discussed in the context of an overall enterprise architecture. Its content and structure are outlined and examples are given of its use along with the value derived from its adoption. The paper then goes on to describe the derivation of a business intelligence function from the understanding of an organization's data, enabled by the data architecture.

Journal of Digital Asset Management (2009) 5, 10–20. doi:10.1057/dam.2008.53

Keywords: business intelligence; data warehouse; data architecture; semantics; service-orientated architecture; enterprise architecture

INTRODUCTION

Data architecture covers the provision of a structured framework for an organization's data, enabling that organization to develop and evolve its systems and processes in order to support its current business activity and, most importantly, allowing it to change in order to achieve its strategic goals in a cost-effective manner.

This framework can cover areas including the following:

- The definition of the data objects that are of interest and use to the organization and the relationships between them, for example what is a 'program' and how do different parts of the business use the concept; does a 'program' always have to belong to a 'series'; in terms of the data is there any difference between transmission (TX) on

a terrestrial analogue channel compared with a digital satellite channel; what is the difference between on-demand publication (on the web) and broadcast publication as regards rights and TX Runs?

- The definition of the actual values used across systems, for example it is important that the rights, scheduling and TX systems use the same definitions of channels, languages, aspect ratios and so on? This includes attention to the details such as how we handle clock changes and what time standard to use (local time vs GMT).
- Ownership of the data, for example if the indexing terms/keywords have to change to reflect our output, who has the authority to sanction the change? It becomes more complex when the data under discussion cross systems and organizational boundaries and are important in multiple systems.

Correspondence:

John Jordan
Siemens I.T. Solutions and
Services, Stadium House,
68 Wood Lane, London
W12 7TA, UK
E-mail: johnjordan@
siemens.com

An example here might be the genre system in use or the language codes used to define the language(s) of our output.

- Data quality, for example how much re-work is caused by incorrect or incomplete data? Are there measures we can put in place, at appropriate points in the business process, to help us reduce this problem? How do we define 'fit for purpose' across a number a different processes?

Without such a framework it is almost impossible to predict the impact of system change on an organization's data or to plan and scope the evolution or replacement of an organization's systems in a way that maximizes the return on investment. This is particularly important in organizations that are undergoing rapid change in their business activity and/or their system provision.

Business intelligence (BI), the provision of information to support the decision-making process, is only achievable if it is built on data of an assured quality, which is relevant to the business. In order to enable this, a successful data architecture framework is vital.

ENTERPRISE ARCHITECTURES

The approach to data that we will discuss falls within the discipline known as enterprise architecture. What does this mean, enterprise architecture?

It might help to consider the history of the discipline. Some people would say that an enterprise architecture is about 20 years old and they would trace its beginnings (as a unified coherent concept) to an article by John Zachman¹ in 1987. This article was written to address two problems which remain with us today:

- Increasing IT system complexity – organizations develop more and more IT systems of increasing complexity in an effort to maintain and expand their business efficiently.
- Decreasing business benefit – as the new IT systems are created, it is becoming more and more difficult to align them, efficiently, with our other systems in order to satisfy our overarching business strategy.

At its simplest, an enterprise architecture is a formalized way of capturing and communicating an organization's system (and other)

requirements in order to support the current activity and to achieve the tactical and strategic goals of the enterprise as defined by its business plan.

Modeling the enterprise's architecture is a common practice in the US Federal Government and in industry sectors such as petrochemicals (BP), car manufacturing (Volkswagen) and aerospace (British Airways and NASA).

As discussed, an enterprise architecture covers more than just data. It is not common to find architectural services in organizations dealing with business strategy, including logical applications and physical applications. In addition, business processes, systems platforms and, of course, data architecture are found as well.

This paper will focus on logical data architecture. It will outline some of the activities that are required in an enterprise data architecture and then go on to discuss how an organization can use this structured approach to data and the clarity of understanding that comes as a result of this approach in order to realize the benefits of BI.

It should be reiterated that the purpose of this activity is to enable the organization to achieve its goals. This is not an 'academic' activity – it is core to the development and success of any business.

DATA ARCHITECTURE

Principles of data architecture

There are a number of principles that underlie the enterprise data architecture. These principles have been articulated differently by a number of organizations, but the following selective list is representative of several of the main points.

Most importantly, we will see that the thrust of the principles is not data architecture for its own sake but as a tool to improve and strengthen business – to enable business to be more agile and responsive.

Data is valuable

Data is valuable. We spend money and resources collecting and storing data and, like any enterprise resource, we should value and manage it. Data is critical to our functioning as an

organization – it is the basis on which we make decisions and it is the medium by which the enterprise co-ordinates its activity. Unless data is managed as carefully as any other corporate asset, we will make increasingly inappropriate decisions and our processes will become more and more inefficient.

Unless we know what rights we have in an asset we cannot make decisions about how to use or broadcast it and how much this will cost. An even more dangerous scenario involves publishing the wrong material or the wrong version of material at the wrong time on the wrong channel.

Data is a common resource

More and more we find the usefulness of specific data-crossing systems and organizational boundaries. We must, therefore, ensure our data are fit for these various purposes. Failure to achieve this has repercussions on what BI can be gained from the data, as we will see later in this paper. Reuse of shared data can also increase process efficiency and will minimize resources required for re-keying.

If commissioners and schedulers share the same data with the play-out chain, then the likelihood of the wrong material being sent to air is minimized. This is an example of shared data solving an existing problem.

In addition, sharing data also brings new benefits. If data that have been stored, previously, in incompatible silos are shared and integrated across the enterprise, then we can start to derive real BI insight. By making data a shared resource, we enable simpler and more cost-effective decision making.

Data quality is managed

As we share data across the organization and use it to inform business decisions, we must be confident of the quality of the data. As we have noted, like any other valuable resource, its fitness for purpose – its quality – should be an acknowledged responsibility, and processes should be put in place to manage this quality.

Sharing data of undefined and uncontrolled quality is more dangerous than not sharing data at all. As we move toward relying on data more and more to inform our business decisions, the quality of that data must be understood. Buying the wrong films or programs based on incorrect

airtime sales, audience figures and rights positions can be costly.

Semantics should be shared

Without the definition of shared meanings for items of data, it is impossible to integrate systems or communicate efficiently across the organization. All communication relies on a common understanding of the meaning of the data exchanged.

A classic illustration of this issue is the inconsistent use of reference data across the enterprise. For example, the incorrect interpretation of genre and/or version codes can give rise to inappropriate material appearing on childrens' pages in electronic program guides (EPGs) or cause the publication of the wrong program version.

This is a critical issue and will be returned to later.

Data must be secure

As we adopt the policy of openly accessible, shared data, we must ensure data's security. Not all items of data should be universally visible. Nor should everyone have the right to modify whatever piece of data they choose, whenever they choose.

There will also be legislative controls over the availability of certain classes of data.

A lax understanding of the security requirements of library catalogs may allow sensitive information, such as contributor names, sensitive source information, contact details and so on, to be published.

Working in the context of an organization's business process architecture, we can develop robust data security profiles, defining who can create, read, update or delete data items and under what circumstances, and link them to the organization's standardized processes and business roles.

Business drivers

A number of current business trends are increasing the need for the business to take a more structured and robust approach to its data.

One of the most important business drivers is the dematerialization of assets. As we move toward file-based storage of asset instances, we find that the importance of metadata in identifying the correct material increases.

Without the provision of appropriate metadata of correct quality, we cannot reliably identify or retrieve our assets.

In addition, we find that there is a change to the users' concept of what assets actually are, which changes both the value they put on metadata and the metadata they use. For example, when copies of assets were based on tape, a business could comfortably merge the concept of a particular tape with that of a program version. This meant that the management of versions of material could be accomplished by the tracking of physical assets. Once the tapes disappear from general use, the business finds that they must rely much more on the descriptive metadata associated with the assets, such as version number, aspect ratio, synopsis and so on.

As the number of delivery platforms increases (for example, new digital channels and web-based, on-demand publications), we are under increasing pressure to repurpose and reversion our existing assets. This also means that we must commission material in the most cost-effective formats to enable asset reuse going forward.

Several of the contributors to the recent Henry Stewart Digital Asset Management Conference (London, 25–26 June 2008) stressed the fact that one of the major challenges facing organizations in the media industry today was the need to maintain and integrate an increasing number of systems within their environment. There will never be a single system that supports all of the processes our businesses require, and our challenge is to select, develop and integrate an increasing number of packages.

Thus, as the integrated broadcasting environment becomes more complex and the richness and number of system interfaces proliferate, we find that the need for automated exchange increases. It will never be desirable, or even possible, to automate all the system interfaces, but we must strive to minimize human interaction where we can. By reducing system integration, where possible, to mechanistic exchanges, we can automate them, but first we need to understand the business processes involved and the data exchanged; therefore, data architecture is indispensable here.

The pressure to open up our archives to commercial exploitation or to make them available to citizens can only increase. Here, the

importance of the descriptive and rights-related metadata cannot be overestimated. What is, perhaps, less obvious is the role of an enterprise's data architecture in ensuring that the appropriate metadata is collected throughout the production process in order to minimize the effort required to furnish the catalog for the digital library.

An additional issue emerges from these last points. As our requirement for high-quality, appropriate metadata increases and as we rely, increasingly, on fault-free, automated system integration, we discover the need to ensure that not only there is agreement in the structure of data held in the separate systems but also, where relevant, the data values held by the different systems also agree.

Components of data architecture

A logical data architecture is composed of enterprise data standards, policies and processes in some, or all, of the areas illustrated below: this list is not exhaustive, but it does contain most of the major components of a standard data architecture (Figure 1).

Business semantics

This is related to the development of an enterprise-wide semantic standard for the data held by, and exchanged between, business units. An example of this is the definition of exactly what the 'Title' field on a system record for a program actually means and for what purpose it is intended.

This example may seem simple, but it worth remembering that under some circumstances a program may have several titles, for example, episode title, episode sub-title, EPG title, on-demand publication title, descriptive title (as used in playout control). It may even inherit a title from the series of which it is a member. Without a clear understanding of exactly what that specific title field means, it is, for example, difficult to integrate systems without human intervention.

The definition of the corporate semantic standard is of critical importance. Although generally expressed as an information systems standard – the enterprise logical data model – it must be remembered that this is a business standard. It records the meaning of the data to the business and must be owned by the business, not the IT department. A previous paper by one

Figure 1: Components of a logical data architecture.

of the current authors² describes the structure and development of the corporate data model and its relationship with business projects in some detail.

It may seem anomalous but one of the great benefits of having a standardized way to express data meaning, requirements and structure is that it supports diversity within the business communities. Having a common definition of meaning – shared by the entire organization – enables separate business units to focus on their core activity and the data that concerns them – while ensuring their data remain understood within the context of the organisation as a whole.

As well as recording the shared meaning of the data for the enterprise, this work stream would also produce policies to ensure that the data structures and meaning defined by projects would be congruent with the meaning and structure of the enterprise standard.

Enterprise and project data modeling

This strand of work is closely linked to the business semantics strand. It would define the approach taken toward the development and maintenance of the corporate standard and the sorts of data-modeling documents that the enterprise expects to be produced by projects.

It is important to note that these policies do not put any additional burden on a project over and above best practice. A project would normally be expected to produce a number of data artifacts – data requirements specifications, logical and physical data models, interface definitions and so on. These are all part of standard system development lifecycles. This encourages good practice and ensures that the

project produces its deliverables in a style that allows simple comparison with the corporate standard.

Interface definitions and schemas

This strand of work defines the enterprise standard covering the structure and meaning of data exchanged between systems. For example, the enterprise standard might require that all interfaces between systems must be defined using the enterprise model's definitions and fields. This means that the interface data content can be understood by anyone familiar with the enterprise model.

Data quality definition and measurement

As we understand the data and their role supporting the business processes, we can start to define data-quality measures that are appropriate for the business. Data quality is defined by the degree to which data follow the appropriate defined characteristics set.

Quality in its broad sense is 'fitness for purpose'. It is a very broad topic and covers areas such as completeness, accuracy, consistency and timeliness.

We can, for example, define completeness standards that say that a news item can go to air with little or no metadata other than editorial acceptance. Nevertheless, we might want to apply quite different completeness criteria to the submission of a commissioned program to the digital library.

Timeliness criteria could define, for live programs, the period after TX by which the reporting information should have been added to the program record. This would allow us

to mitigate the risk of missing agreed deadlines with the rights societies for post-TX reporting.

Data quality is emerging as a factor of major importance. Gartner, the industry analysis and intelligence group, have estimated the data-quality market as growing to about \$600 million annually by 2011. British Telecom (BT) have estimated \$80 million per year of savings achieved through their data-quality improvement program.

Reference data management

This is another enormous topic and is of huge importance to integration projects. This section can only hope to outline the topic and the issues.

'Reference data' is a common term in the description and design of information systems. It can be defined as data that is used to categorize other data or objects within a system. Examples in use in a typical media organization will include, but not be limited to, the following types:

- Type codes – tables of codes and values used to categorize an object by its type (for example, media types of 'audio' and 'video', format types of 'MPEG', 'AVI' and so on), version types and version languages;
- Status codes – (codes that describe a lifecycle of an object, for example, 'rushes, completed');
- Externally available value sets – these are updated on a regular basis, possibly from a source external to the business (for example, exchange rates);
- Descriptive classifications or vocabularies that are constantly growing — (for example, genres, content classifications, lists of performing artists and detailed locations).

Typically, reference data sets are small, discreet sets of values that are not updated as part of business transactions but are used to impose consistent classification. An important point to note is that reference data classes tend to be relevant across more than one business system. In order for these systems to interoperate, the relationships between the reference data sets in the systems must be understood, or we will be faced with issues such as the wrong material being played out, on the wrong channels, at the wrong time.

According to a study by Reuters/Capco,³ over 30 per cent of all system transaction exceptions can be traced to bad reference data.

Governance

As can be seen from the previous sections, the topic of governance looms large over any attempt to implement data architecture.

The importance of governance is difficult to overstate. Unless responsibility and accountability for the definition and quality of metadata are established, it will prove difficult if not impossible to support an effective data management environment. This governance will also involve developing measures of quality, and delivering to those measures.

The existence of roles with this responsibility and accountability also creates a level of confidence in metadata so that reuse and sharing are promoted and embraced.

How to use an enterprise data architecture

The principles behind a data architecture and its components have been outlined, but how do we use it? There are several areas where managing an organization's data in an architectural manner will provide business benefit.

The architecture will define the agreed business data requirements in a clear, unambiguous and consistent manner, enabling system specification, shortening system development timescales and reducing project risk.

Although individual projects may have a clarity of understanding of their own data requirements, unless there is some device for understanding of the organization's overall data requirements it is difficult, if not impossible, to positively impact the organization as a whole. The development of an enterprise data model will establish a common semantic standard across the organization, enabling unambiguous communication across the enterprise, leading to simpler system integration and increasing the return on investment of system development.

Because the data model is pan-enterprise, it can be used to define project scope. In addition, by mapping the organizational data to the enterprise's systems, applications, processes and business units, it becomes possible to perform impact analysis on suggested changes.

The clarity of the data definitions and the development of the data-quality criteria provide mechanisms by which processes can be defined to increase data quality and decrease re-work.

The understanding of the organizational data, the messages exchanged between systems and the mapping of the data to the system processes, systems and business units will simplify the development of a BI capability.

In summary, the data architecture enables the business to be responsive, efficient and proactive.

Let us now turn our attention to one of the specific business benefits enabled by this structured and rigorous approach to our corporate data – BI.

WHAT IS BUSINESS INTELLIGENCE?

A common definition of BI is information that supports business decision making delivered at the right time, in the right format and to the right people. The important points to pull out of this are as follows:

- It is 'information' not data that is important. The two terms are often used interchangeably; however, data are just 1's and 0's and only becomes information when supplemented by the correct business context (for example, '26' on its own is just a number, but if 26 is the average age of viewers of a sports highlights show, then it becomes powerful information).
- 'At the right time' means both having the information in hand when the decision is made and having the information at the correct currency (for example, last week's information may or may not be good enough, depending on the decision).
- 'In the right format' covers the way the information is presented, for example, does the user need to see high-level trends or are they interested in lower-level information? Also, how do you present the information to the user – is it on regular reports, are they self-servers, do they need to manage by alerts and exceptions?
- 'To the right people' is all about insuring that the right reports are distributed to the people who need that cut of information to make the decisions. It is key to understand user groupings and the different levels of responsibility that they hold.

This paper will also discuss data warehouses and their role in BI. A data warehouse is essentially a

store of all the data, which supports business decision making, pulled together from disparate sources into a central common format. The data warehouse is designed to hold more information than operational systems (as it is able to store historical data to track trends over time) and is tuned to allow that information to be distributed to the 'right people' as efficiently as possible.

Challenges for BI

The challenges with which data architecture is faced also cause problems for the BI world.

Consider a scenario where a production manager queries the production management system to find out the running cost of his current production. The figure returned suggests that he is under budget and so he can make a decision to go back and re-shoot a scene that he was not happy with.

In the meantime, this data is passed to the finance system, where the accountant asks the same question of the data. She gets a much larger figure, over budget for the production, and makes a decision to limit future budget for the production.

So, why has this happened?

One explanation is that they are actually asking slightly different questions of the data; if their definitions of 'cost' are different (maybe they include different cost codes) then they will get different results. This is known as 'different versions of the question' in the BI world and is a result of not having a common understanding of business terms.

The other explanation is that the data is held differently in each of the queried systems, which could apply different rounding logic or apply different types of validation on the data. This is known as 'different versions of the truth' and is a result of not having a common application of business rules.

Data warehousing is all about bringing data together and joining it to derive insight. A third problem encountered in data warehouse construction is being able to join data that has, historically, grown up in separate data silos.

For example, a common problem encountered by audience researchers is to be able to assess audience figures for a particular series. They will have to join data from audience research organizations (such as BARB in the

UK), data from their online broadcasting services as well as data from any mobile services provided. Unless key concepts, such as 'Series Name' and 'Genre', have common business definitions across the three systems, then it can be very difficult to join the data together.

The benefits to BI of a common understanding of data

Data quality is vital, not only to the success of systems interfaces, but also for allowing decision makers to act on the information provided with confidence. Many warehouse initiatives fail, as the end users are not able to trust data to make decisions.

The data warehouse is also a sink of data and so can be used to identify data-quality problems and identify where, at source, the issues can be corrected. For example, programs may be classified into different genres in the archives and rights management systems. If you have not integrated these systems then you may not have spotted this. The warehouse provides this integration and allows you to compare values in

different systems against a set of data-quality criteria (which an organization will have defined as part of the data architecture).

In fact, a data warehouse can only be developed if its foundations lie in a common understanding of data – it is, by definition, a central store of information within an organization for supporting the decision-making process. The core principal is to reduce all the data to a common format so that areas that can be linked in business terms can be linked to answer business questions. This common format is the semantic model, the common business understanding.

A common understanding of data (such as a semantic model) also facilitates BI business requirements gathering. When you are trying to understand the decisions that different areas of the business need to take and the information they need to support those needs, the benefits of a common language that a semantic model gives you is invaluable. A business term has a clear definition and you can avoid making costly assumptions (such as 'What is a program?').

Figure 2: Typical business intelligence deployment using SOA.

It also helps the next stage when the business analyst puts the requirements in front of the system designer, helping to remove another common area of ambiguity. As anyone who has been involved in a business change project will have experienced, ambiguity leads to a delivery that does not meet the actual requirements of the business.

Finally, it promotes reuse of data assets – if you have a common understanding and achieve storing the information in one place then you can easily reuse terms to help satisfy different types of key performance indicators (KPIs). It is unknown for BI initiatives themselves to experience the problems of system silos if they are developed without a strict common set of business definitions – you can easily end up with different versions of the truth being created if duplicate business data entities exist in the warehouse environment.

SOA and a semantic model in a BI architecture

BI has been around for decades but it is only now, armed with a common semantic model, that its potential has really started to be realized.

The service-oriented world also offers benefits for BI. Figure 2 shows standard BI components deployed within a service architecture.

- (1) Here we see the metadata representing business transactions being passed to a transformation step. Some of these systems will save up transactions and pass them to the warehouse on a daily basis (lots of BI requires only daily or weekly reporting; the data needs only to be as current as the requirements for reporting), but the service-orientated architecture also allows individual transactions to be passed to the data warehouse as the business event happens. This

Figure 3: Information on the move.

Figure 4: Bash-up using a map.

allows an organization to keep up-to-date with performance, such as being able to see which programs have been streamed most in the past few hours.

- (2) The transformation step takes disparate data structures and definitions and converts them (using business understanding) to a standard format. This standard format is based on the common semantic model, which we discussed earlier. Within a service-orientated architecture, the transformation layer will be able to perform the same transformations on individual transactions as it will to bulk files, and software tools, which extract, transform and load data into warehouse environments (commonly referred to as ETL tools), can be set up to use the same logic for both transactional and batch sets of records. The transforms will all be based on the common semantic model.
- (3) Once received, the messages are loaded to a common format (again dictated by the same semantic model) and often extracted into a set of KPIs and focused data marts (or views of data). Once more, common business definitions of data are used to define these business views. These business views can be queried by a data exploitation tool, directly connected to the warehouse.
- (4) The SOA layer can also be used to pass the results of standard business insights across to any connected application (or user) within the architecture.

In some industries, the benefits of having up-to-the-minute information available are clear (such as managing queues in a call center operation). In media, the benefits of up-to-the-minute information are less clear cut, for example, the question posed by our production manager is probably relevant to the picture a day ago (you would imagine that costs are not entered into a system at the exact time they are incurred). However, the ability to share information across the organization is still a big benefit for media.

The power of exploitation

If we refer back to our definition of BI, this architecture helps us to deliver the right information to the right people (through an understanding of the Business Process Architecture) and at the right time (in time to make the decision and of an appropriate currency)—but what about the right format?

Getting the format right for the audience is all about defining different paths to the same underlying information. Technology has taken great leaps in BI in the past few years and it now facilitates a vast array of paths.

Business information can be delivered to web pages (via portals such as Microsoft SharePoint or via Java on web pages), direct to the end users' desktop and even to mobile devices such as PDAs (as seen in Figure 3).

Going back to our production manager, this means that he could be alerted when the

production cost gets close to the budget rather than having to constantly check by running reports.

BI content can also be merged with other web content (something which has been coined a 'Bash-up'). A good example of this is using maps, which can be a very powerful visual tool for regional analysis. Figure 4 shows how audience figures can be linked to a Google map.

This is also particularly useful in news production, to track where resources have been deployed (for example, when a story breaks it is vital that the closest available satellite truck is sent to the scene).

BI tools also provide powerful extensions to traditional search engines. As well as being able to search through documentation, web content and blogs within the organization, the tools are able to search through information stored in or produced from the data warehouse. Again, this BI search is facilitated by the use of common business terms.

SUMMARY

In the first part of this paper, we saw the origin of data architecture in the growing realization of the need to optimize enterprise investment in

information systems. The principles surrounding data architecture were outlined along with the business drivers, which are focusing our attentions on our data assets. We then described some of the main components of data architecture, the architectural deliverables required and the use of the architectural products within the business.

Finally, we looked at how, by developing a clear, shared view of data, we can take advantage of the information to be derived from our systems – bringing business benefit by informing business decisions.

As Benjamin Disraeli once said, 'As a general rule, the most successful man in life is the man who has the best information'.

REFERENCES

- 1 Zachman, J. A. (1987) A framework for information systems architecture. *IBM Systems Journal* 26(3): 276–292.
- 2 Jordan, J. (2005) Evolving your semantics: Feedback between data projects and the corporate standard. *Journal of Digital Asset Management* 1(6): 386–398.
- 3 TowerGroup. (2001) *Reference Data: The Key to Quality STP and T+1*. Needham, MA, USA: TowerGroup.