XML, FLEXIBILITY AND SYSTEMS INTEGRATION

Mark John O'Brien[1], **Nasser Al-Biqami**
1. School of Computer Science and IT, University of Nottingham

ABSTRACT: There are two primary ways in which data can be integrated. One involves the establishment of a centralised data store that meets all the needs of a construction project; the other recognises the geographical and functional fragmentation of the industry and views data integration as a conceptual process. From a purely technical point of view the first is perhaps the easiest, but it fails to meet the organisational and economic demands of the construction industry. Thus the second approach is more likely to be adopted by the participants of that industry. The problem then becomes one of mapping the metadata structures of one participant onto those of another. Various efforts at the development of standards have attempted to address this issue. However, standards can be both complex and inadequate. The complexity is a demand of the industry while the inadequacy stems from the impossibility of coping with every eventuality - a severe problem given the essential uniqueness of each building product. This is not to say that standards are not required, merely that their limitations are fully realised from the outset and that expectations are not raised to the point where disappointment sets in and they fall into disrepute. EDI is a perfectly good standard but has failed to make a great impact on the construction industry. The volume of application-to-application communications remains small.

This paper argues that while standards such as EDI can form the backbone of data communications - and therefore provide a vehicle for data integration in the construction industry – they are insufficient to cope with the desired flexibility demanded by the industry. The paper then develops this idea by suggesting that something more is required, something flexible. Extensible markup language (XML) is a tool which can help provide the necessary flexibility.

XML is a language which provides a common syntax for expressing the structure of data. While it can be seen as an extension of the commonly used Hyper Text Markup Language (HTML) this fails to recognise that XML has uses beyond the creation of Web pages. In its broadest sense XML allows systems developers to define the structure of a document. Currently its main uses are for data interchange between humans and machines, but the ability to facilitate machine-machine interactions is the most exciting concept for construction industry systems.

Now EDI is a perfectly good tool for such interactions but in the event of any new requirements the standards need to be extended. This is such a long process that by the time it is completed it is of no use to the original users. XML however provides a dynamic mechanism which can be adapted as required to meet the needs of the users. This is its great strength for the construction industry - an industry that is "document-rich". In effect by using XML to specify meta-data structures one overcomes the differences between the data structures of different trading partners. No longer will we require all parties to conform to the tramlines of a strictly enforced standard, but rather those parties will be able to exchange data merely by changing the XML description of their documents. Thus in conclusion this

paper shows that the use of XML within the construction industry will facilitate data, and hence systems, integration.

KEYWORDS: XML, EDI, flexibility, standards

1. INTRODUCTION

Systems integration remains problematical within the construction industry. As a problem it has been the subject of numerous analyses and to a large extent the arguments are both well rehearsed and well understood. In part the problem remains intractable because it is made up of a combination of narrow technical limitations and wider aspects of the application domain. If the technical limitations could be taken in isolation then it would take little effort to solve the problem of systems fragmentation. Equally, if the industrial fragmentation were to be tackled independently of the technical issues then it would present a relatively minor problem: industrial consolidation is a common phenomenon in the modern business world. The real difficulty arises because of the co-existence of these two aspects. It is probably true to say that they will only be solved together.

This paper considers a recent technical development which could present an opportunity to move forward. XML is a rather simple technology which, despite the rather dubious extravagant claims of some of its proponents, nevertheless holds out a good deal of promise. At base it provides a method to describe tree structured documents but within this apparently restricted ability it has sufficient flexibility to tackle a wide range of applications. Indeed, the most interesting of these applications were not even in the minds of those who devised the language: the use of the language to describe database structures.

The paper first of all briefly runs through the key aspects of systems integration. In particular, it addresses the concept of data integration, since data integration is a necessary prerequisite for systems integration. Data integration requires the use of agreed standards between the integrating parties to align their various data stores. Traditionally, this has been seen to be an industry-wide issue, where standards are set for all participants within the industry. This paper, however, points out that standards can be created between just two trading partners, and moreover those standards may only exist for a single economic exchange. The reason this has not be prevalent in the past has been the difficulty of creating and implementing such unique one-off standards. The effort involved has been just too great. XML, as a language and technology, offers the opportunity to reduce these difficulties to such a level that the creation of "on-the-fly" standards becomes feasible.

Thus the paper concludes with two observations. The first is that the use of XML to create industry-wide standards is a rather sterile activity. Yes, it can be done but represents a waste. The second, and key observation, is that XML can provide a mechanism for the quick and easy creation of metadata structures to facilitate the exchange of data between trading partners. It is this flexibility that holds out most promise for the construction industry.

2. DATA AND SYSTEMS INTEGRATION

As a term *systems integration* can be taken to mean whatever a particular speaker wants it to mean. This is due to the indeterminate use of the word *system* and the ill-defined word *integration*. At one extreme it can be taken to be the complete harmonisation of structures and functions throughout a wide area of application, whereas at the other it can be seen as the mere shuffling of data from one system to another in some semi-automatic manner. Yet

however it is taken an absolutely necessary prerequisite for *systems* integration is *data* integration. And at least in the non-object oriented world data integration involves the syntactic and semantic mapping of data from one store onto some other store.

O'Brien (1997) outlined the two primary ways in which data can be integrated. One involves the establishment of a centralised data store that meets all the needs of a construction project; the other recognises the geographical and functional fragmentation of the industry and views data integration as a conceptual process. From a purely technical point of view the first is perhaps the easiest, but it fails to meet the organisational and economic demands of the construction industry. Thus the second approach is more likely to be adopted by the participants of that industry. The problem then becomes one of mapping the meta-data structures of one participant onto those of another.

This pragmatic approach is the one that is now generally taken to be the correct paradigm for systems integration in the construction industry. The main reason for this is that it does not attempt to alter the fragmented industrial structure of the industry. The myriad of small organisations, which are distributed by both function and geography can maintain their own data stores. Moreover, all the potential problems pertaining to data ownership and control within the first approach are avoided. In short, provenance implies ownership.

Yet the multiplicity of data stores suggests that they have separate existences for very good reasons. The primary reason is that they are all essentially unique. If they were not then industrial consolidation would have happened a long time ago since each of the small organisations would have been providing exactly the same services and products. It was Porter (1985) who pointed out that *differentiation* is one of the main competitive drivers. But not only do these data stores contain different data, but more importantly, they will have different structures, that is to say they will have different metadata. The attributes will be different in both syntax and semantics, constraints will be organisationally determined and the structural linkages between different parts of data stores will vary from one store to another. Differences in syntax usually present a minor problem - data values can be cast from one abstract data type to another. Semantic differences are far more serious and usually defy any formal translation from one system to another. Indeed, it is those data items which demonstrate close, yet different, semantics which present the most problems.

3. PROTOCOLS AND STANDARDS

Nevertheless, the use of translators or intermediary software allows data to be transferred from one system to another in an automatic or, at least, semi-automatic manner. Yet translators rely upon the existence of a mapping of metadata structures from one system to another. These mappings can be set up through a process of negotiation. That negotiation can be bilateral in which case two trading partners can establish their own linking rules. Until recently however, the more normal route has been to establish wider standards. Such standards attempt to create frameworks or templates which describe generic trading exchanges: sales orders, invoices and so forth remain fairly constant and lend themselves to such an approach. That is to say, standards work best when the messages are relatively fixed in format. While the standards are costly to establish, and even more costly to conform to, such costs are outweighed by the sheer volume of messages exchanged in these fixed formats. Invoices, delivery notes, sales orders and so forth are excellent examples. They are ubiquitous.

Standardised messages are the mass produced artefacts of the connected world: cheap and common. But here we come up against the developments in the world to produce customised artefacts. (A detailed description of this development would unbalance the argument at this point, for more details see Pine, 1993). Any business text will tell you that the operations and means of production must match the nature of the artefact being produced. In particular the IT strategy of any corporation must be coherent with the overall business strategy. Thus in the world of mass customised, unique, non-standardised artefacts the use of database systems which are repetitive and standardised is no longer appropriate. Moreover, while the common standardised messages will remain, they will not be crucial to business success. That will rely upon unique messages with individual trading partners.

But here we observe the poor level of integration within the construction industry. If it were such a simple matter of setting up industry-wide standards then they would already be in existence. The truth is that whereas the standard trading exchanges do indeed take place they represent a non-critical element of the trading process. More important are the unusual, unique, and differentiated trading exchanges which are a natural result of the construction processes and the industrial structure which provides the matrix for those processes. It is precisely at this point that standardised, industry-wide negotiated protocols have no role to play.

An excellent example of this industry-wide approach has been the development of Electronic Data Interchange (EDI). EDI is a very simple yet highly useful technology. It extends the idea of emails, which are essentially person-to-person communications, into the domain of application-to-application communications. Now whereas humans can react to messages in whatever form they appear in, applications can only react to valid inputs. Thus in EDI, email messages are structured. The originating application creates the message using a predefined structure and then the receiving application also has an understanding of the structure and can thus unwrap, parse and react to the message. The entire data exchange relies upon the existence of the predefined structure – it is this structure which is established by standards bodies. And, as ever, the issue is not one of the particular data in question but rather its format, type, structure etc: in short standards are concerned with *metadata.* (A detailed analysis of EDI within the construction industry can be found in O'Brien and Al-Soufi, 1993.)

4. THE ADVENT OF XML

This brings us to XML and the opportunities it presents to the construction industry. XML is being developed by the World Wide Web Consortium (W3C) and details on their work can be found at www.w3.org. Yet before proceeding a brief description of XML might be useful. Many will be familiar with HTML - hyper-text markup language. This is the language which is used to describe most web pages. More generally it is a language used to describe documents. It is a markup language in the sense that it describes how data is displayed rather than what the content of that data actually is. It does this through the use of *tags.* Now the tags of HTML represent the language itself and are fixed.

XML is also used to describe documents. Equally it also has tags. But there are two important differences to HTML. First the tags are not predefined, they need to be set up by person creating the document. And secondly the tags are primarily concerned with document structure rather than how it is displayed (although this can be part of the tags). XML views documents as tree structures. Thus a simple example might be that a book consists of chapters

and each chapter consists of sections. In such a document it would be up to the user to create tags `<book>`, `<chapter>` and `<section>` and to establish the one-to-many relationships that exist between them.

Now the originators of XML envisaged precisely this use for their language. They have been taken aback, however, by the adoption of XML in an entirely different area: databases. XML is so simple and so clean that it has immense flexibility. This flexibility means that rather than use the tags to describe large scale data items such as a chapter, or a section of text, they can also be used to describe fine granule data such as date of birth, gender, stock number, quantity or any other of the data primitives so commonly found within databases. It turns out that XML is perfect for specifying metadata. Indeed, some proponents of this approach are now suggesting that post-relational databases will not be object-oriented in nature but rather based on XML. Such a view is certainly speculative.

5. XML AND STANDARDS

Now XML has begun to be used to create standards. There is a certain irony here because XML itself is not a standard. Rather it is a recommendation of the World Wide Web Consortium (W3C). Of course, within a few years whether or not it becomes an official standard, it will probably attain the status of a *de facto* standard. Now the very ease with which XML can be used to create metadata descriptions has prompted some people to create, or in some cases re-create, high-level messaging standards. It appears to be perfect for the job. It's quick, cheap, and has a high degree of flexibility. Within the construction industry such an effort is the aecXML (see their web page at www.aecxml.org) in the United States who are proposing standards for documents common within the industry. Equally CITE in the United Kingdom is considering such a move. (The geographical fragmentation of the construction industry is particularly marked at the national level: there is very little cross-border trade – even with the European Union.) Yet these efforts are really just a way to create the standardised messages commonly found within EDI standards. If this were truly the technique of the future then EDI messages would have become the lingua franca of the construction industry. The sad fact is that EDI has not achieved any great degree of penetration within the industry and it is unlikely that any similar efforts will succeed where it has failed.

Thus these XML standardising methods misunderstand the real power of XML. XML is really at its best when it is dealing with *non-standard* documents. Its speed and ease is perfect when trading partners must deal quickly with each other using documents and data which is unique to that particular trading exchange. The construction industry is full of such exchanges and the description of the move towards mass customisation suggests that this will become more common rather than less. Moreover, not only will it become common but it will become essential to competitive advantage. Without the capability to trade quickly with partners in one-off exchanges will mean a rapid loss of market share.

Thus the issue becomes one of utilising the highly flexible nature of XML to facilitate one-off messages where the structure is unique to the message. Certainly, one can handcraft an XML exchange but its real power will be appreciated when tools allow the generation of XML documents. (These are beginning to appear now, but remain expensive – see, for example, the system WebMethods at www.webmethods.com.)

All of this is not to deny the necessity of standards: they will still be required for the repetitive parts of business. But the very nature of repetition means that it can be repeated by *anyone.* Those repetitive exchanges cannot form the basis of any business advantage because they are open to all competitors in a business sector. It is the things which make corporations *different* from each other which hold the key – and if trading partners have different databases, data structures, data systems, then we need something to "glue" them together: XML is that glue.

Standards essentially derive their power from establishing *fixed* metadata structures, but XML provides the capability of creating *new* and *flexible* metadata structures. This is the difference in approach which means that it is essentially a waste of a flexible technology to create inflexibility; unfortunately the sheer convenience of XML as a tool suggests that large-scale, well-publicised efforts will continue to promote this approach which is so inconsistent with the capabilities of the tool itself.

REFERENCES

O'Brien, M.J. (1997). Integration at the limit: construction systems, *International Journal of Construction Information Technology,* **5**, pp 89-98.

O'Brien, M.J. and Al-Soufi, A., (1993). Electronic data interchange and the structure of the UK construction industry, *Construction Management and Economics,* **11**, pp 443-453.

Pine, B.J., (1993). *Mass customisation,* Harvard Business School Press, Boston, Massachusetts, U.S.A.