

# THE SIGNIFICANCE OF MARKUP LANGUAGES IN CONSTRUCTION

Edward Finch<sup>1</sup>

## ABSTRACT

The adoption of a standard markup language offers significant benefits for construction firms who today are having to manage considerable amounts of project information. The term *markup* originates from the early practise of publishers who routinely made notes in margins to identify elements of text. Electronic data interchange (EDI) in construction has tended to focus on the exchange of CAD based information, but little advance has been made in the adoption of textual formats. This paper discusses the shortcomings of conventional methods of text based exchange and points to some of the radical departures afforded by markup based systems. In particular; the opportunities for collaborative preparation of documents during the design process; the opportunities for maintaining and reusing information; and the possibility of increased connectivity between documents and sites.

Keywords: markup languages; document management; SGML; HTML; HTML+;VRML;

## SIGNIFICANCE OF DOCUMENT MANAGEMENT IN CONSTRUCTION

Text based documents play a major role in the construction process, including specifications, quality assurance documents, progress reports, procedures, and maintenance manuals. Some of these documents may relate to specific projects or may have a company-wide significance. The moment something goes to print it appears to become obsolete. Information has to be duplicated many times over in order to address different audiences or accommodate minor changes.

For many organisations involved in the construction process, the *information* about their products (buildings) is as important as the product itself. For management contractors the only tangible commodity they deal with *is* documents. Some of the direct costs associated with document management include printing, copying, distribution, filing and storage. For manuals, users will need access to complete copies and incremental updates. Perhaps the most costly of the above stages is the process of filing. Incorrectly filed updates will result in the use of out of date information and can often result in reworking or extra design effort to remedy mistakes. Another indirect cost is that associated with the availability of information: this in turn determines the likelihood of the information being used. Simply having access to the relevant document does not ensure that the information is easily obtained. Accessibility of information within a document is dependent on the size of the document, the richness of the index and the proximity of logically related information. Whilst having an entirely electronic information system

---

<sup>1</sup>E.F.Finch@rdg.ac.uk: <http://www.rdg.ac.uk/ConMan>

ensures that relevant documents can be transported electronically, the access to information within these documents may still prove difficult.

The annual cost of document management to many construction organisations represents a significant proportion of total outlay. In order to obtain an advantage in such a market, organisations which can make better use of the information are likely to have a considerable advantage. Much of this advantage pertains to future proofing, ensuring that changes in technology do not preclude future use of information.

## **STANDARD GENERAL MARKUP LANGUAGE**

The advent of the Standard Generalised Markup Language (SGML) represented a watershed in the evolution of text. It became an ISO Standard in 1986 (ISO 8879) and has since been adopted as an international standard in industries as diverse as car manufacture, defence, pharmaceuticals and telecommunications. The rate of adoption in the construction industry has been much less spectacular. Much of this can be explained by the industries failure to understand the importance of ensuring reusability of textual information.

All documents contain three different types of information:

- The *data* in a document which typically include text, tabulated data and graphics
- The *structure* in a document which refers to the relationship of the data elements. For example, in this document there is a title, an abstract, authors, body text and various headings to which body text is attached. In a delivery note, the structure is more precise, with a specific order and number of elements
- The *format* of a document which determines its appearance. For example, in this document, the title appears as centred bold text.

If a text document is used to convey information, its visual appearance does not ensure correct interpretation by the user. For example, we may infer from the appearance of text at the beginning of this document that it is a title: invariably titles appear at the beginning of a document and are usually emphasised in some way. However, it is often more difficult to unambiguously identify text, and some means of tagging is required to remove this ambiguity. This approach forms the basis of the Generalised Markup Language developed by Charles Goldfarb for IBM in the 1960's. His intention was to develop a language which would let text editing, formatting and information systems share documents. SGML was developed so that data, structure and format were treated as separable elements in any kind of text based document. SGML does not represent a single rigid authoring format, but a metaformat which can be used to specify a wide range of document formats.

## **BENEFITS OF USING SGML**

The Standard General Markup Language SGML is designed to describe documents in terms of the language and terminology of text. The design of SGML markup language is primarily optimised for identifying textual objects according to their

function or purpose; not according to their presentation characteristics which are described independently. It is designed to enable data description, data modelling and data interchange. The low uptake of SGML in construction is in part due to the level of awareness of this interchange standard which has been overshadowed by graphical exchange formats.

The benefits of using the standard markup language SGML in documents, as identified by Adler (1990), are:

- application independence
- device independence
- adding intelligence to the data
- aiding in the management of information resources

### **Device and Application Independence**

Application independence means that the exchange of documents, both internally and with external organisations, can occur with minimal reworking of information. This is beneficial to all companies involved because it means that they can make their own decisions about the tools and processes used in their companies. This also applies to the devices used for creating and viewing documents. SGML specification says nothing about the user interface. The eventual appearance of SGML documents is determined by the publisher (i.e. the person viewing the document). The vehicle for defining the way in which information appears (i.e. with respect to typographical appearance) is the *Document Style Semantic & Specification Language*. Individual companies are likely to have their own preferred 'style sheet'.

### **Adding Intelligence to Data**

Perhaps the most significant benefit of SGML is the use of unambiguous tags in text. This allows documents to be interpreted and searched, and individual items of information can be interconnected, abstracted and reused. In a strict sense, SGML is not a language but a meta-language. It defines the rules whereby application specific languages can be developed. The individual tag names are thus designed to suit the needs of the particular application. Document applications in construction might include tenders, contracts, quotes, specifications, catalogues and installation and maintenance instructions. It is the Document Type Definition (DTD) which defines the structure and rules used to markup text. The complexity of these DTDs depend on the particular application which can range from extensive and complex manuals which are rarely changed to installation instructions which would be simple in structure but variable. To illustrate the three components of an SGML markup system consider a standard form as an application as shown in Figure 1. The form would be associated with a specific *document type definition* (DTD) as shown in figure 2.

Maintenance Acceptance Procedure		
<b>A</b>	<u>Technical Information</u>	
1.	Manufacturer	Craig HVAC Systems
	Type of Equipment	Cooling Plant C-122
	Sub-Assembly	Cooling coil
2.	<u>Reason for Failure</u> ( 'wear and tear'; 'improper use', 'improper maintenance' "vandalism')	wear and tear
3.	<u>Corrective action</u> (immediate correction, correct in planned operation, run to failure)	immediate correction
4.	<u>Repair proposal</u>	see Drawing 8983-3
<b>B</b>	<b>Cost Information (±5%)</b>	
1.	<u>Cost of maintenance</u>	
a.	Equipment Cost	8988
b.	Labour Cost	7878

Figure 1 Example of Standard Form

```

<!--Document Type Definition for the form -->
<!-- "Maintenance Acceptance Procedure" -->

<-- ELEMENT modify -- (technical, costinfo)>
<-- ELEMENT technical --
(item1, reason, correction, repair)>
<!-- ELEMENT costinfo -- (costmod)>
<!-- ELEMENT costmod -- (equipment, labour)>
<!-- ELEMENT item1 -- (manufacturer, type, sub)>
<!-- ELEMENT (reason | correction | repair)
-- (#PCDATAT)>
<!-- ELEMENT (equipment | labour) -- (#PCDATA)>
<!-- ELEMENT (manufacturer | type | sub) -- (#PCDATA)>
<!-- ATTLIST modify eng-no CDATA #REQUIRED>

<!-- ENTITY #DEFAULT SYSTEM>

```

Figure 2 A Standard Form DTD

The final component is the generated tagged document, known as the *document instance*. Figure 3 shows the document instance for the form shown in figure 1.

```
<!-- An Instance of Maintenance Acceptance Procedure -->
<!DOCTYPE modify SYSTEM "modify.dtd">
<modify maint-no = "MAINT212">
  <technical>
    <item1>
      <manufact> Craig HVAC Systems </manufact>
      <type> Cooling Plant C-122 </type>
      <sub> Cooling coil </sub>
    </item1>
    <reason>
      Wear and tear
    </reason>
    <correction>
      Immediate correction.
    <repair>
      see Drawing 8983-3
    </repair>
  </technical>
  <costinfo>
    <costmod>
      <equipment> 8988 </equipment>
      <labour>7878 </labour>
    </costmod>
  </costinfo>
</modify>
```

**Figure 3** A Document Instance for Standard Form

## **BARRIERS TO SGML ADOPTION IN CONSTRUCTION**

There is a significant start-up cost involved in the adoption of electronic data interchange within a construction firm. Preliminary work is required to define requirements; design applications; procure hardware and software; and design training programmes. There may also be a substantial amount of legacy documents which may need to be converted into SGML format using optical character recognition followed by further editing. As noted by Maunder (1994) there may be a need to produce a cultural change and a new organisational structure in organisations. There is an armoury of solutions for data interchange, including E-mail, CAD exchange standards (e.g. IGES, DXF); knowledge base standards (e.g. STEPS) and graphical standards. For the user community there is no simple solution to what is a complex problem. Construction firms must commit effort in understanding their requirements and realise that there may not be a single approach to meet their needs.

## **ADDING THE HYPERTEXT DIMENSION**

Expert systems have often been presented as a solution to complex construction problems. However such systems are limited by; (1) the effort involved in

encoding expert knowledge; (2) the inflexibility of the resulting solution; and (3) the inability of such systems to reflect human reasoning. One alternative is to use hypertext. This is a method of internally cross-referencing written information so that a user can jump from topic to related topic. This provides an interlinked network of information comparable to the links which occurs in the human brain as a result of the learning process. Linkages can also be made to spreadsheet or database data, multimedia, or computing routines. Williams (1994) cites an example of how hypertext manuals were produced by General Electrical Systems for their technicians involved in the repair of imaging equipment. Previously, they had to carry manuals weighing a total of 90kg in their cars. The new method involved the use of portable computers and a CD-ROM player. The reported productivity saving was 9% as a result of using hypertext, with an increase in sales of \$25 million. A particular benefit of hypertext in construction is the freedom from using a mouse or keyboards. These methods of data entry and access are not convenient on construction sites. Pen based pointing systems in conjunction with hypertext provide a much more robust solution.

An often overlooked benefit of using hypertext is the presence of a hierarchical structure which arises from the links. This enables users to develop a hierarchical structure of the text they read. It also ensures that they are able to start from the 'big picture'.

### **Use of Hypertext in Co-authoring**

Co-authoring is a fundamental part of many programming, design and planning processes in construction. The authoring of documents such as a programming brief typically involve sequential stages of increasing levels of detail, with discrete approval stages. Creation of such documents using hypertext present significant opportunities, but also demand new ways of working.

A number of hypertext systems have been developed which enable computer-supported collaborative writing. Such systems have been used to:

- facilitate early stages of problem solving, supporting brainstorming and planning (Conklin and Begeman, 1988)
- support individual co-authors in the organisation of text, allowing the merging of text with other co-authors at a later stage (Delise & Schwartz, 1987)
- provide communications at specific stages of the co-authoring process to support collaboration at all stages of writing

The act of co-authoring involves a process of dissonance resolution or 'bringing together of minds'. People often find it difficult to relate new knowledge to their existing knowledge. For example, the designer of a HVAC system for a building may have difficulties in reconciling their knowledge with that of the architect in the preparation of a specification. Many of the developments in hypertext co-authoring applications attempt to address this issue.

The NEPTUNE co-authoring system developed by Delise and Schwartz (1987) provides a mechanism for checking multiple perspectives in collaborative writing. It uses the principle of nodes and links to ensure that the parts of the authoring work evolve in a consistent manner and efficient versioning transitions occur between co-authors. Perhaps the most advanced collaborative tool is SEPIA which

recognises four aspects of the authoring process; (1) planning space; (2) content space; (3) argumentation space, and (4) rhetorical space. The system is designed to monitor and organise interactions among co-authors.

Preparation of hypertext documents is known to present a greater cognitive load on authors compared to linear text. This arises from the need to make sense of connections and information chunks. Fisheye browsers have been developed (Chen *et al*, 1994) to provide a balance between displaying local details and the overall organisation of a large information space. These provide an outline window showing the hierarchical structure of the hypertext workspace together with a window for specific hypertext documents and one for word indexing.

Hypertext is a facility which can be readily incorporated into the Standardised Generalised Markup Language SGML and provides another means for enriching document intelligence.

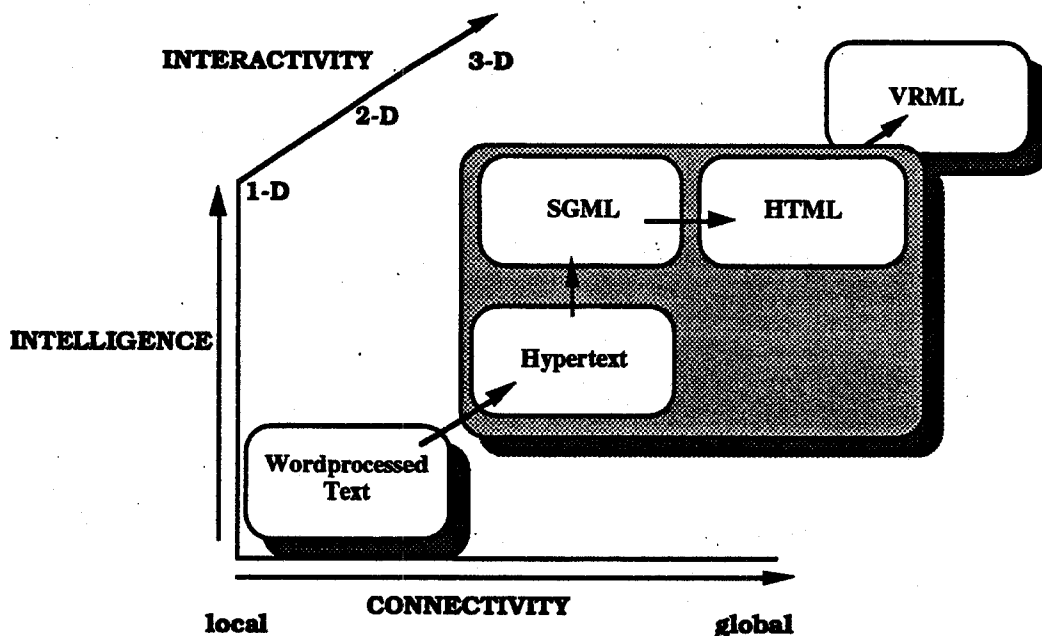
## **HYPERTEXT MARKUP LANGUAGE**

In the '70s the emergence of interlinking networks including NSFNET, NREN (US networks) and JANET (the British Joint Academic Network), gave rise to what is termed the *Internet*. This provided instantaneous access to global sources of information. However, the absence of a suitable user interface capable of using the many applications and platforms available on the Internet, compromised its exploitation for nearly two decades. It was the development of the *Hypertext Markup Language* (HTML) in conjunction with a browser for viewing information which opened up the latent capability of the Internet. HTML is a derivative of SGML containing a set of modular extensions for linking to remote sites and various multi-media formats. The combination of servers (computers providing the information), browsers (computers viewing the information) and the standard HTML protocol which determine how the information should be communicated, brought about a form of network communication known as the World Wide Web.

Its impact on the construction industry has been limited to date. In the UK much of this explained by the cost of line charges, particularly for firms wishing to become information providers. A number of technical resources are however becoming available in the UK including the British Standards, HMSO and Barbour Index. With the removal of restrictions to the Internet and an inevitable reduction in line charges, the cost effectiveness of links, even at the site level, will become more apparent. Most commercial organisations currently providing information on the World Wide Web are using it as a publishing or marketing medium. Documents are prepared as 'one-offs' for Web delivery using the HTML format. Few organisations are using it as an information system in the conventional sense, because few organisations routinely use HTML as a document format. However, for organisations opting to use the Standard Generalised Markup Language, presentation in HTML can be achieved without degrading the original document structure. The continued development of HTML is likely to lead to a universally accepted interface for accessing networked information. Ironically, it is the advent of HTML which has brought attention to the merits of its parent language, SGML.

The advent of Virtual Reality Markup Language VRML adds a further dimension to markup languages as shown in figure 4. It is currently being introduced on the back of HTML. It represents a complete departure from the use of text as a means

of communication. Every type of 'object' is represented as a 3 dimensional object in VRML. In construction, objects might include not only physical objects such as columns and beams: abstract objects such as 'contracts' could be represented visually, with all the elements of a contract and the variants of a contract appearing as separable items. Each object can be associated with an unlimited number of attributes. For example, a contract's status could be visually represented using a colour coding system. This would provide a way of improving the user's ability to interpret information and recognise patterns. Some aspects of construction clearly benefit from the introduction of a 3D interface. Buildability and clashing problems in services integration are good examples. Many proprietary systems already provide this form of 3D interaction as a development of CAD. Furthermore, many construction organisations are interested in exploiting such software technologies for internal advantage. Therefore, the benefits of a common virtual reality standard are not as obvious. However, as a non-proprietary standard, it may provide a way for designers from different organisations to resolve design issues involving 3 dimensional space.



**Figure 4** The evolution of markup languages

#### **FUTURE DEVELOPMENTS**

Electronic information exchange does not stop at the stage of electronic file transfer. Mechanisms must be incorporated into electronic text to allow access to the internal workings of a text document. This will prevent the production of automation islands. Documents will be capable of being re configured for different audiences and abstracted for future reference.

In construction, many of the formats used for electronic distribution of textual information embody the restrictions of the printed page: documents are broken into a sequence of fixed size pages to be viewed in a pre specified style. No methods are used for indexing or enriching the information present in text documents. As



observed by Raggett "global networks now make it easy to distribute information electronically and the time has come to shake off the limitations of the printed page"

The impact of HTML is likely to be less far reaching for individual construction organisations than SGML. However, it may stimulate a movement away from large monolithic construction organisations towards smaller satellite organisations feeding a hub organisation. These smaller organisations will be able to link up to the parent information systems without having to change platforms or applications. The current rationale for being part of a larger organisation, based on access to training, design libraries, and other sources of expertise, and a common method of information exchange, becomes less compelling. Use of other technologies such as videoconferencing, video site links and e-mail will further enable virtual working. This trend will benefit larger construction organisations confronted with fluctuating workloads by the introduction of flexible working arrangements.

## REFERENCES

- Adler, S. (1992). The Birth of a Standard. *Journal of the American Society for Information Science*, 43(8), 556-558.
- Chen, C., Rada, R. and Zeb, A. (1994) An extended fisheye view browser for collaborative writing. *Int. J. Human-Computer Studies*, 40, 859-878.
- Conklin, J. and Begeman, M. (1988). gBIS: a hypertext tool for exploratory policy discussion. *ACM Transactions on Office Information Systems*, 6(4), 303-331.
- Delise, N. and Schwartz, M. (1987). Contexts - a partitioning concept for hypertext. *ACM Transactions on Office Information Systems*, 5(2), 168-186.
- Li, S. and Juster, N. (1993). An Exchange Strategy for Production Information. *Interfaces in Industrial Systems for Production and Engineering*, 10, 175-186.
- Maunder, C. (1994). Documentation on tap. *IEEE Spectrum*, September, 52-56.
- Raggett, D. (1994). A review of the HTML+ document format. *Computer Networks and ISDN Systems*, 27, 135-145.
- Williams, T. (1994). Applying Portable Computing and Hypermedia to Construction. *Journal of Management in Engineering*, May/June, 41-45.