INTEGRATING CONSTRUCTION INFORMATION: AN OLD CHALLENGE MADE NEW

R. Amor
University of Auckland, Auckland, New Zealand

ABSTRACT: Integrating construction information for a project is one of the major issues faced in IT supported projects. The change in representations and transport formats from paper to electronic provides a chance to improve upon current methods. Many tools, techniques, and processes are posited to manage this problem, though how comprehensive a solution they provide is not always clear. This paper enumerates the issues around different facets of the information integration problem to help provide an understanding of the scope of issues to be tackled.

KEYWORDS: Integration, Information management, Integrated systems

1. INTRODUCTION

Integration is a fundamental process underlying almost all aspects of construction IT. It is only through integration, of varying degrees, that it is possible to support many of the processes involved in a modern, IT-based, construction project. A well structured integration framework supports and enhances all forms of communication and collaboration on a project; it provides the basis from which up-to-date information across all partners in a virtual team can be accessed and visualised; and it provides the basis upon which more complex analysis, and intelligent systems, can be developed.

Integrating construction information has always been a part of normal project work, with paper-based information management requiring a large overhead in a normal practice. DMS (Document Management System) vendors argue that approximately 40% of an engineer's or architect's time is spent in the management of information related to a project. Certainly, in a well managed office there is a large overhead in classifying information, filing information, managing different versions of drawings and files, and ensuring the appropriate dissemination of information (document) sets to the participants on a project.

With computers becoming pervasive within organisations, and the adoption of LAN (Local Area Network), WAN (Wide Area Network), and Internet/Extranet systems providing simplified and rapid communication mechanisms, the challenge of integrating construction information is taking on a new dimension. The change in representation and communication media provides a chance to not just replicate previous practice with speed improvements, but to enhance and restructure the whole integration process. There are many drivers for this restructuring including business reorganisation, new methods of business, new tools and functionality, and most depressingly for poorly suited IT systems and applications (i.e., to suit the way the computer system works).

The scope of the integration of information covered in this paper will be very wide, encompassing all types of information that may be associated with a project from all sources. While no definition of the scope will be attempted, the reader should envisage information sources and types including the following: personal repositories, department and organisation repositories (e.g., libraries), project repositories, manufacturers' information, legislative information, industry news, project documents (e.g., CAD files, contracts, requests for information, etc), simulation (design tool) analyses, structured project data (shared and personal repositories), process definition, commercial transactions (e.g., EDI), etc. The only constraint that is applied in this paper is that all of this information is electronic in some form (e.g., electronic document, scanned document, reference to a physical document, CAD file from which printouts are derived, etc). From this very broad scope the many aspects of the integration problem are considered.

2. ASPECTS OF THE INTEGRATION PROBLEM

This main section of the paper provides an eclectic grouping of the range of issues which need to be addressed when considering integration of construction information in the modern environment and into the future. For some of the listed issues there are already solutions (some work and others do not), for other issues there are proposed solutions which still have to be proved in practice. Other issues are still in the research domain, and a small number seem not to be currently addressed. There is enormous cross-over between the various issues described below and no effort is made in this paper to show the interdependencies between the issues. This paper provides an initial enumeration of the range of issues that need to be addressed in the new integrated information age.

2.1 Information Sources

Provision of data

There is a need to identify who is going to provide all of the information which is required in a project aside from project partners. There are many existing information sources (e.g., product manufacturers, standards institutes, research organisations, etc), where the benefit to these organisations in providing electronic versions of their information must be clarified. In many cases electronic versions of current paper information sources are not suitable for new approaches (e.g., data models), and this requires great effort by some actor in the industry to populate a partial data model (e.g., from a product specification sheet to an explicit data structure). Issues include who: gains from this process; creates the tools to enable this process; trains those who need to understand the data models; will pay (extra) for it; and what are the business drivers to make this happen. It is also likely that portals will be created that provide one-stop shops for many major information types in the industry (e.g., CONNET, see Turk and Amor 2000) requiring a further level of information management than in current projects.

Design tool outputs

The output of an analysis or simulation is often very valuable (e.g., time taken to obtain, or cost of getting an expert to perform the analysis). It needs to be clear where this information is stored in an integrated system, and how is it tied to the version of a building from which the analysis took place (obviously as the model changes the analysis may be rendered obsolete). The analysis may supply information either as a file (e.g., a rendering) or into a data model (e.g., beam size calculations).

Business data

Increasingly project-based information repositories are able to be linked with an organisation’s business systems. Integrated systems need to be able to cope with links through to the wide variety of business tools utilised by organisations (c.f., design tool integration). However, this information resource usually has greater constraints upon security and access than other project-based information, especially for external organisations.

Referencing information sources

For audit purposes it is necessary to reference the source of all information utilised within a project (e.g., was the most up-to-date standard used?). Given the wide range of sources of information (paper, electronic catalogues, Internet, etc) not all of which are able to be copied to a project repository (copyright issues), there must be a mechanism or structure available through which the source, version, etc of all referenced information can be recorded.

Data consistency

Multiple information sources are likely to provide inconsistent heterogeneous views of the same information type or item. The management of information sources needs to recognise the structuring and semantics of information being incorporated into the project information space (see Section 2.3 for issues in this area).

New media types

Any system developed for integrated construction information needs to provide easy extensibility for new media types which are appearing (e.g., white boards, streaming video/audio, etc).

2.2 Information Representation

Sufficient models

Before an integrated information system can be used for a particular project, or phase of a project, there must exist sufficient data models to represent the domains within which the project is to operate. For systems dealing purely with document-based information this may be possible today. Where shared data models are to be utilised then further analysis needs to be performed to determine whether sufficient data models exist for the project requirements.

Data model comprehensiveness

Where a standard data model is to be utilised for a project it is necessary to be sure that this model is comprehensive enough for the job. Issues to consider are whether the requirements of all design tools to be utilised on the project can be met by the data model being considered, whether the data model captures the outputs of the design tools as required for the project. How this comprehensiveness can be demonstrated for potential users is a difficult issue, as a data model can never model the full spectrum of products that can be designed, yet describing the constraints of a particular data model within a design space is not simple.

Data model validation

Data models are not developed in a manner which allow confidence in the validity of the completed system (Bjork 1999). This includes low levels of testing during development, simplistic examples used for testing and initial demonstration, insufficient documentation to comprehend the scope of a model and hence its ability, development by a small number of individuals representing vested interests (e.g., one type of organisation, or a single country), etc. While the reasons for this approach are understandable, it provides a situation where independent appraisal and verification is nigh on impossible. In comparison to software development methods this is comparable to the 'big-bang' method where one hopes everything works right at the end.

Multiple representations

Unique and comprehensive data models for a particular product or domain are not achievable. There are many very different views of a product, held for different processes and needs. This tends to lead to data models of very different structure for each view. These data models are often unresolvable without making drastic simplifications or concessions. This leads to multiple representations of similar constructs for different views, domains, and processes (see Section 2.3 for issues in this area). This extends to the notion of a single complete data model for a domain, a goal which is now accepted to be unachievable, leading to multiple independent (or loosely related) data models for particular domains and processes.

Data models with behaviour

The data models that exist today invariably have no behaviour incorporated within them (i.e., they are pure data models). Though they are defined within modelling formalisms which allow some level of behaviour to be defined, the difficulty in specifying and agreeing upon behaviour has restricted what has been achieved. The definition of data semantics is a difficult enough process, the definition of behaviour raises many issues in terms of implementation, transportability, checking, etc. The fact that a formalism could define some of this behaviour does not solve the major issues of how and where it would be implemented.

Dynamic data models

The ability to dynamically define and extend data models has been incorporated into existing data models, for example as property sets in the IFCs (IAI 2000) or in EDM (Eastman et al 1995). These are interesting ideas whose utility are not well understood. Issues include definition and agreement of dynamic data semantics, discovery of semantics by new users and applications, control of specification, validity, name space clashes, etc. The use and management of property sets for a large multi-partner and information provider project has not been shown.

Data modelling formalisms

The formalisms currently utilised for data model specification (e.g., EXPRESS and EXPRESS-G) provide a limited set of descriptive capabilities. Already, extensions to EXPRESS are being defined and a major effort is being undertaken in the USA to identify successor formalisms to the object-oriented paradigm (which EXPRESS follows) to resolve weaknesses identified over the years. This would indicate that models developed in EXPRESS may not be well founded or complete.

2.3 Information Mapping and Enhancement

Integrating data views

In a full integrated system for a project there will co-exist multiple data models (e.g., APs in STEP) populated with data for their domains and processes. There will exist some overlap between these data models, and as common data could be transferred between models across processes mappings will be required between these different views. A range of mapping languages have been developed over the last few years (Verhoef et al 1995) with EXPRESS-X (ISO 2000) currently being the main language for the A/E/C domains. The utility of these languages for real projects is also not proven (as for data models in Section 2.2), nor is it certain what the expressive power of these languages is. Work to date has identified a range of structures where automated mappings are not possible (e.g., aggregate to component views) and it is clear that human intervention will be necessary in many mappings between large data models. How this human intervention is invited, managed, and presented has not been researched widely. Where mappings are unresolvable there are issues of what can be done to record constraints imposed by these unresolvable mappings. Some work (Amor 1997) has looked at the ability to record constraints against data models based upon mappings, which are managed with rights much as users have rights to various portions of an information repository. Data must be mapped in both directions between repositories, which requires duplicated (but inverse) mappings in procedural mapping languages (e.g., EXPRESS-X) leading to a greater chance for incorrect specifications.

Integration of applications

Integration of applications poses many problems in an integrated system due mostly to the constrained or highly specialised model of a building (or construct) which often exist within these tools. Constraints are for example the limits on the number of spaces or elements which can be modelled in an application where often human expertise is required to manage the mapping. Highly specialised tools often require greater detail for products than would typically be found in a data model (e.g., detailed material properties from testing procedures).

Integration of incomplete models

It is unlikely that any particular data model will be complete and totally consistent until very late in the project. When integrating construction information there is a need to handle incomplete and inconsistent models over long periods of time. This has a profound effect on the ability of many mapping languages to complete their mappings, and to recover from the lack of complete or consistent information that will be presented at any one time. Research projects such as ToCEE (1998) have looked at mechanisms through which incomplete and inconsistent data models can be managed across a project's life span.

Application interfaces

How well an application can operate within an integrated environment depends upon the interfaces it provides for receiving information. Obviously, where an application has an in-built interface for a particular data standard there is likely to be better integration, though there are issues of how well it complies with the data standard, and how much of the data, and how well the data, is mapped into internal representations. This will depend to a large extent on the closeness between the application's data representation and that in the transfer format. Where an interface does not exist and translators need to be developed for another format there is an extra level of mapping which is likely to introduce further inconsistencies in the final model. What the level of confidence a user can have with loaded data is not easily measured or expressed.

Classification of information

For less structured information (e.g., documents), and even to some extent structured information, there is an issue of how to manage and classify the information content it represents. Paper-based libraries utilise various classification systems to locate and store information. Many integrated information systems (e.g., DMS) utilise full-text search across the full content of a document. Where there are a large number of documents (e.g., most construction projects) this becomes less and less useful with an inordinate number of matching documents being found for a search. The continued use of classifications, either supplied by users, or automatically derived in some fashion (e.g., Extractor 2000, Kea 2000), needs to be integral to integrated information management.

Intelligence in information retrieval

Until integrated information systems are readily available it is unlikely that general purpose intelligent systems will be able to be developed within the A/E/C domain. Very specialised tools will be able to operate in restricted domains (as they do today), but intelligent tools (e.g., agents) able to find best prices, find best products, locate relevant design information, offer holistic design alternatives, etc all require a large base of information which is well organised and controlled, as well as a significant user community.

2.4 Information Communication

Communication protocols

A wide range of communication protocols are available for access to integrated project spaces. The capabilities of the protocols used for particular implementations of a system need to be made explicit to understand the support offered for particular processes. For example, session-based work is not supported by Internet protocols (i.e., TCP/IP). Integrated project spaces need to be easily extensible for new protocols (e.g., white-boarding, streaming video, etc).

Data transport mechanisms

Several data transport mechanisms are utilised by different systems (e.g., SPF, SDAI, CORBA, XML). The abilities of the differing mechanisms are disparate and the power of an integrated system will be affected strongly by the mechanism utilised. For example, SPF (STEP Physical File – Part 21) transports the complete data repository, while SDAI and CORBA provide partial access either from central stores or distributed heterogeneous stores. The usage scenario for new technologies (e.g., XML, RDF) are widely hyped but have not been demonstrated in contexts which show their utility for construction projects and repositories.

Distributed information

A single global information repository can be difficult to negotiate for multiple partners whilst ensuring acceptable security and control of documents and data by originating partners. The more general (and politically acceptable) solution is a virtual repository comprising multiple distributed sources of information. The constraints on protocols supporting this type of store need to be explicitly defined for a system especially in terms of access rights, session control, and transaction management across participants (see Section 2.5 for greater detail).

2.5 Collaboration and Co-operation

Concurrency and collaboration

Working methods vary widely depending upon business practice, project type, and even the project information system functionality. Concurrency may or may not be allowed within different processes of a project and needs to be defined and managed across a project, impacting on the use of an integrated project store. Collaboration between project members needs to be facilitated by an integrated system, with issues of private workspace, partially shared workspace, and open workspace being required for different processes.

Process management

The process developed for a particular project impacts heavily upon the usage of an integrated information system in terms of staff involvement, access rights, information provided and information required. Tight integration with the global project process and macro-level organisational processes needs to be accommodated within an information store.

Project information space

A project information space (Augenbroe 1999) needs to exist for every project, either as a physical or virtual construct. The project information space needs to accommodate disparate and heterogeneous sets of information. In the majority of cases the project space will be a virtual construct with information sets being widely distributed across partner organisations and other relevant information providers. Methods to manage this disparate collection of information need to be adapted to individual project's needs. Information spaces can be provided for varying information types in different manners, for example, document sharing in a controlled manner across multiple organisations (e.g., Columbus) through to distributed data stores (e.g., CORBA). Common protocols need to be agreed, or systems developed, which can accommodate a wide variety of existing systems' protocols.

Legal obligations

The project information space exists within the bounds of legal and regulatory frameworks for the project being undertaken. Regard needs to be taken of various regulations (e.g., data protection act) as well as the system being structured to facilitate use in a court of law (e.g., legal admissibility of electronic documents). Support must also be offered for organisational requirements (e.g., sign-off procedures) and contractual obligations as specified for the project (e.g., authorisation hierarchy).

Co-ordinating systems

Partners coming together for a specific project are likely to have their own preferences and in-house tools for information management. Often, especially with larger organisations, the use of these tools is mandatory for all projects undertaken. The level of interoperability that can be provided between different information management systems ranges considerably. Research in the CONDOR (Rezgui et al 1997) project established a low level of common functionality across a range of DMS. Interoperability between information management systems and the common protocols to support this needs addressing.

Security systems

The security of a shared project information system is vital for a real project. The possibility of unauthorised access to the system has to be shown to be minimal. The security of information being passed to and from the repository needs to be sure. The system needs to be running from secure sites (physically) with backup devices and procedures available to handle problems (e.g., denial of service attack) which could potentially delay the whole projects.

Access rights

It must be possible to specify the rights of all involved in a project for individual pieces of information. This must take into account process stages as well as individual pieces of information (e.g., no one can modify a contract document after it is signed). Specifying a user's access rights for a structured data repository is no easy task. Current matrix systems with rights for each user, for each attribute of an object, at each process stage are not feasible on a real project (Ott 1998). Cooper and Rezgui (2000) provide a higher level model for this issue. Systems need to be established which control who can specify rights for which process stages, and also provide an audit of decisions which were made. Given the complexity of what is specified it will also be necessary to provide a method to verify that correct access is provided for particular project participants.

Verifying information receipt

An integrated information system will have to be able to show that project participants requested and received particular documents and pieces of information in case of dispute (e.g., I never saw that document). In relation to the project process the system should also ensure that users get the relevant and necessary information at the right time

Conflict resolution

During the course of a project there will be points of disagreement between participants involved in varying aspects of the project. An information repository (especially a structured data store) allows users to change and modify information as they wish within a process.

However, when no agreement can be found it will be necessary to provide conflict resolution procedures. Meetings, phone calls, email, etc may provide the most appropriate method, but more structured exchanges can be provided through the project information space (e.g., ToCEE 1998).

Consistency management

Inconsistencies in the information store need to be managed across a process stage with the ability to work concurrently on incomplete states, merge work sets (which may still be inconsistent), and, finally, determine whether a completed model (portion of a project) is actually consistent (e.g., all constraint satisfied). This process is made more difficult by the long transaction times which are normal within construction projects (hours to days in duration). Research has been undertaken in this area (ToCEE 1998, Amor 1997) and there are commercial offerings which provide a level of transaction and consistency management (e.g., Bentley, see Aish 2000).

Version management

Within projects there are key points at which new versions of the project model are released. A project information space needs to be able to record versions against all its information items in order to serve the most up-to-date information to its users as well as for auditing purposes. Within a project there are not just the major public version releases, but also private versions utilised by participants working on a particular aspect of the project. These private versions need to be managed for the individual (or small group) before being reincorporated with the public model version system. Similar to relational database systems, versions often need to be managed so that a project can be rolled back to a previous version (e.g., where problems with a particular approach are identified) and for the project to progress again from that earlier version.

Archiving

While usually considered within the functionality of a DMS the ability to archive a project version across all information types (e.g., data repository) raises issues of consistency and retrieval. The time periods for which project information needs to be stored guarantee many changes in the version of software systems being used, and may well see some systems not in existence after five to ten years. Archive formats for a wide range of information types need to be established and managed in a co-ordinated way across the whole project information space.

2.6 User Issues

User interface

The human computer interface (HCI) is the most difficult component to deliver in a manner which matches the user's requirements at any one time, let alone matches their model of the world. Varying modes of interaction can be incorporated into an integrated system, but working from user requirements for information use, and working through processes to be undertaken in a project will provide a range of differing requirements. The level of adaptability of a system to user requirements, process requirements, self customisation, etc should be addressed and made explicit.

Information search and retrieval

Different methods of access and provision of information will be appropriate at different stages of a project and for different information types. File hierarchy navigation will work for some project stages (Columbus 2000), word search across the repository (e.g., DMS functions) at others, classification-based search will be used to identify specific information types, and structured data access (e.g., SDAI or CORBA) is also needed. Integrated systems will need to offer a wide variety of access and storage modes to cover all information types and user needs. Whether the provision of information is static or dynamic will change depending upon project stages and work items (i.e., push and pull of information). Intelligence can also be added to this process to provide more tailored and appropriate information to users. How these requirements are specified and managed by users needs to be explicit.

Visualisation

The HCI need not be purely textual, with graphical and virtual reality interfaces providing alternate methods of navigating and manipulating information stores. The utility of these new techniques is not widely shown on real projects, nor the cost benefit. The learning curve and ease of use of these new techniques in viewing and interpreting information needs to be addressed to show where benefits lie.

2.7 Business issues

Cost benefit analysis

Organisations are presented with a wide array of new tools and techniques which can be utilised in an integrated information system. The cost benefit of any of these tools or techniques versus the particular processes followed by the organisation is very hard to quantify. The impact new systems will have on an organisation are also not well quantified as to the process change that will need to be undertaken to adapt to the technology. To ensure take-up of new tools and technologies these aspects need to be accessible to organisations.

Interface with business functions

Integrated project systems store information of use to intra-business processes or drawn from these processes. Methods of accessing and communicating with internal business processes will be a common requirement. This includes commerce (and E-Commerce) transactions undertaken for a project, management of internal processes (e.g., authority and sign-off), staff allocations to projects over time, etc. The integrated information system can not be totally independent so levels and methods of interaction need to be defined.

Sharing of information

Organisations are wary of the potential loss or sharing of certain categories of information across project partners. Where a shared information repository sits outside the organisation there are further concerns over the ownership, control, and access to information belonging to, or supplied by, the organisation. Issues of security, access rights, etc as described in Section 2.5 need to be addressed along with procedures and guides for use of project spaces by staff within an organisation. These guides and levels of interaction allowed with an integrated system are likely to vary across organisations.

Provision of data

The provision of information, especially as it becomes highly structured, brings issues of cost and benefit for the providing organisations. There is a high cost to the semantic enrichment of current information bases where the benefit to the organisation is often not clear, and usually not in the short term. How this is financed, where the costs are recouped, and over what time frame needs to be highlighted to ensure buy-in from providing organisations. Many models are mooted, but the acceptance and practicality of these models is not visible in practice.

REFERENCES

Aish, R. (2000). Computing fundamentals and process re-engineering for collaborative design, Proceedings of the UK National Conference on Objects and Integration, Watford, UK, 13-14 March, pp. 184-195.

Amor, R. (1997). A Generalised Framework for the Design and Construction of Integrated Design Systems, PhD thesis, University of Auckland, Auckland, New Zealand.

Amor, R. and Faraj, I. (2000). Misconceptions of an IPDB, Proceedings of the UK National Conference on Objects and Integration, Watford, UK, 13-14 March, pp. 124-135.

Anumba, C.J. and Amor, R. (1999). A Survey and Analysis of Integrated Project Databases, Proceedings of Concurrent Engineering in Construction'99, Espoo, Finland, 25-27 August, pp. 217-228.

Augenbroe, G. (1999). Guest editorial, International Journal of Computer Integrated Design And Construction, 1(1), pp. 1-2.

Bjork, B-C. (1999). Information Technology in construction: domain definition and research issues, International Journal of Computer Integrated Design And Construction, 1(1), pp. 3-16.

Columbus (2000). Columbus DMS, http://columbus.arup.com/.

Cooper, G. and Rezgui, Y. (2000). Objects and Integration in the "Wired Society", Presented at the UK National Conference on Objects and Integration, Watford, UK, 13-14 March.

Eastman, C., Jeng, T-S., Assal, H., Cho, M. and Chase, S. (1995). EDM-2 Reference Manual, Centre for Design and Computation, UCLA, Los Angeles, USA, 50pp.

Extractor (2000). Extractor text summarization software, http://extractor.iit.nrc.ca/.

IAI (2000). International Alliance for Interoperability, http://iaiweb.lbl.gov/.

ISO (2000). Express-X Language Reference Manual, ISO-10303, SOLIS, WG11, N103.

Kea (2000). Kea automatic keyphrase extraction, http://www.nzdl.org/Kea/.

Ott, E. (1998). The building site without paper - legal aspects in an IT environment., Proceedings of the CIB W78 Information Technology in Construction conference, 3-5 June, Royal Institute of Technology, Stockholm, Sweden, pp. 335-346.

Rezgui, Y., Cooper, G., Bjork, B-C. and Bourdeau, M. (1997). From Construction Product Information to Consistent Project Documentation: The CONDOR approach, Proceedings of the CIB W78 conference, Cairns, Queensland, Australia, 9-11 July, pp. 337-346.

ToCEE (1998). Towards a Concurrent Engineering Environment, http://wwwcib.bau.tu-dresden.de/tocee/.

Turk, Z. and Amor, R. (2000). Architectural foundations of a construction information network, International Journal of Construction Information Technology, 7(2), pp. 85-97.

Verhoef, M. Liebich, T. and Amor, R. (1995). A Multi-Paradigm Mapping Method Survey, Proceedings of the CIB W78 Computers and Information in Construction, Stanford University, Stanford, USA, 21-23 August, pp. 233-247.