Information Dynamics, Perspectives, and Risks

DAVID GEORGE

Department of Computing, UCLAN, Preston, UK

Abstract

The complexity of 21st Century society is demonstrated in a dynamic information environment that is fuelled by the expanding digital world. This demands an effective information management approach that supports reliable information provision, worker collaboration, and satisfies business intelligence initiatives that drive marketing strategies. The Web and databases are now inextricably entwined and new solutions need to be identified that can provide more effective access to billions of data sources. Emerging Semantic Web technologies may be a solution on the horizon and may help to address the potential integration challenge and risks.

Key Words: business intelligence, database, information, integration, intranet, RDF, Semantic Web, OWL

Dynamics of an Information Society

Organisations develop as a result of the complexity of society and survive by satisfying the needs of other organisations and customers. They have to handle technological development, aggressive market competition and expanding markets (Johnson McManus and Snyder, 2003). These issues, allied with business reorganisation and mergers driven by evolving corporate strategies, all contribute to the concept of organisational change in the battle to stay ahead. Thus, the 21st Century workplace is a dynamic environment and many organisations demonstrate an insatiable need to reorganise and develop their information systems to understand markets, identify profitable customer segments, monitor performance and comply with government legislation (Rob and Coronel, 2002). Equally, financial constraints and profit maximisation, service or efficiency requirements or the desire for strategic marketplace differentiation, all drive systems development programs that bring the challenge of integrating legacy and new information systems.

The characteristic of effective organisation structures is how well they meet the challenge of harmonising three key components of task, individuals and groups. They achieve effectiveness through merged information extraction, performance monitoring, and enabling communication and understanding by the information consumer - individual, team or other organisation.

The Information Environment

Many companies have gradually evolved as global organisations with data distributed in many parts of the world. Organisations have also attempted to achieve large-scale vertical integration with suppliers and customers, by transacting e-commerce through the Web. However, despite new database application development, organisations are often characterised by legacy databases and associated applications being retained and supported (Stonebraker et al., 1993), and these can form a significant part of fragmented information structures.

- **Internet and Web**

The Internet and, more specifically, the World Wide Web (Web) has provided the platform for an "information space of interrelated resources" (W3C, 2004a) used to enable accessibility of digital resources. The vitality, and essential property, of the Web is its universality because it exploits the hypertext link, so that it is possible to link any document or data source to any other. This Web data access functionality can be demonstrated in various environments, from the public or "open" Internet to corporate intranets and extranets.

- **Intranet Origins**

Whilst public Internet sites tend to be open and not explicitly restricted to a particular class of users, intranets and extranets are more exclusive (Powell, 2002). An intranet is a shared information resource for employees, within a closed or discrete private network. Nevertheless, it employs standard Internet protocols, TCP/IP and HTTP, and Internet technologies (Bansler et al., 2000, Karlsbjerg and Damsgaard, 2001). Whereas traditional client/server systems manage multiple applications and often have interface issues, intranet protocols use a common language and communicate via a web-browser, thus providing a common interface. It is for these reasons that Web-based communications platforms have

represented the platform for emerging data and communication technologies.

▪ Information Resource Management

Organisations can improve operational effectiveness and productivity by enabling users to access data more easily. For example, cost savings can be realised simply by moving processes online and reducing paper and distribution costs (Ward, 2001). Sun Microsystems intranet, for instance, delivered a reduction on paper document costs of approx £2m p.a. (Horgan, 1997).

Recent Intranet/Extranet development, using Web technologies, has shown that successful management of emerging technologies has been vital in supporting new, and equally emerging management philosophies; where the focus is on "openness" and sharing of information (Bansler et al., 2000, Wagner et al., 2002). This has been facilitated by better organised and faster data access, using features such as information "portals", and the realisation that collaborative effort can harvest improved workforce productivity e.g. consider the empirical study of US West (Bhattacherjee, 1998). The most productive intranets focus on news provision, enterprise-wide directories with associated search facilities, custom portals and compliance systems that focus on ISO/Quality and legislative issues (Lamb and Davidson, 2000). These facilities generate widespread and frequent/daily usage because end-users use them as virtual libraries. However, their success depends on the integration of data sources.

Changes in organisation culture, allied with emergent technology, have harnessed the concepts of worker collaboration and information sharing to promote operational best-practise and the empowerment of employees to provide faster decisions and improved customer service (Wagner et al., 2002, Bansler et al., 2000, Bar et al., 2000). Web-browsers can access data held on different systems and stored in varied formats, thus providing a single, common graphical interface. Therefore, an organisation has the capability to instantly link geographically isolated operations with common, integrated, and up-to-date information.

The following comment provides an interesting perspective on the information challenge and relates to IBM's "Dynamic Workplace" Intranet (Eliot and Barlow, 2002, Smeaton, 2002), which has been attributed with revolutionising the way in which employees can communicate and access information:

"There were far too many sources of information to search through..... Key to our success ... was the goal of rendering the complexity of the organization irrelevant for employees"

To reduce complexity, IBM's challenge was to merge more than 8,000 "local" intranets, linking more than 11 million Web pages to support 300,000 employees.

▪ Caught in the Web - the Price of Success

The seemingly inexorable penetration of the Internet and Web into daily life, facilitated by basic underlying structures and protocols, has unearthed significant shortcomings because most web content is stored in natural language format. As a result, the current Web works well for creating and presenting different types of web content but affords very limited support for meaningfully processing the data. This is because it is very much dependent on the human users during search, extraction, and interpretation stages.

The task of accessing information sources ranging from unstructured and semi-structured data through to autonomous, federated and clustered database systems, has presented users with potential information overload and brought with it the resultant problem of how to identify meaningful and relevant data. An example can be demonstrated, at the most basic level: where different organisations post information on the Web, resulting in documents or databases containing related information appearing at different Web sites. However, whilst they may contain related information, i.e. "semantically" related from a contextual view, they are inevitably likely to be in multiple formats and employing different terminology or data schema, resulting in potential integration issues. Equally, consider a potential homebuyer seeking to identify a certain range and type of property, in an area with good employment prospects, low crime and highly rated schools and hospitals? In this case the data integration problem assumes different dimensions, because a search would require access to autonomous databases holding

property, demographics, crime, health services and education data, to provide an integrated meaningful answer.

These issues demonstrate the real world complexity that information systems must potentially seek to address and are consistent with the 1998 "Asilomar Report on Database Research" (Bernstein et al., 1998), which then highlighted the need for the database community to radically address the way that technology captured, stored, analysed and presented the vast and increasing amount of online data. It was considered that the database community should widen its research to encompass all Web content and online databases, with a ten-year "Information Utility" goal: to make it easy for everyone to manage most human information online.

The Information Challenge

Thus the dramatic growth in the Internet and Web has brought with it the need for effective and flexible mechanisms to provide integrated and meaningful views over multiple information sources and data types; in effect, in the homebuyer's use case, a "mediation" of potentially complex, multiple, real worlds that will support information and knowledge acquisition. Similarly, organisational metamorphosis driven by organic growth of businesses, organisational restructuring and corporate amalgamations have inevitably demanded strategic level information systems planning; to allow otherwise geographically isolated, national and global operations, access to shared, distributed and heterogeneous data stores containing heterogeneous data types and definitions.

However, simply managing and sharing data represents only the raw material or first layer in knowledge management. To derive knowledge from integrated data, it requires a logical flow of information from data that provides semantics or meaning i.e. data is both concept related and contextual with other data sources. The challenge of organisations is thus to provide an information systems framework that ultimately supports a "knowledge model" that depicts creation, retention, transfer and utilisation based on unstructured, semi-structured and structured "knowledge artefacts" (Newman and Conrad, 2000).

Next-generation technologies are now being developed to address these challenges, such as Web Services and Semantic Web. In Web Services, the traditional concept of the Web, being designed for human interpretation and solely a repository for text and images, is now being utilised as an integrated "provider of services" (McIlraith et al., 2001, Brodie, 2002). A typical example, such as holiday and flight-bookings, would use tools to build "virtual" advanced systems accessing multiple distributed systems supplied by different organisations.

The Semantic Web (Berners-Lee et al., 2001, Hendler et al., 2002) represents the next generation of the Web "to create a universal medium for the exchange of data". It incorporates several key Semantic Web technologies that have been adopted as World Wide Web Consortium (W3C) recommendations, the Resource Description Framework (RDF) core language (W3C, 2004c), and the RDF Schema and OWL Web Ontology languages (W3C, 2004b). These will serve as a platform to support a standardised query language for RDF that will permit widely distributed RDF/XML data collections to support integration and function as a universal data exchange mechanism. This will support enterprise application integration and provide the platform to interconnect personal information and global commercial, scientific and cultural data in a "seamless" manner. Therefore data can be shared and processed by automated tools as well as by humans, to provide an automated knowledge resource that accurately reflects real world "meaning" or semantics.

The Price of Failure

It is perhaps appropriate to briefly consider both motivators for, and the consequences of, not addressing the data integration issue.

On the 11th September 2001, an airliner ploughed into the North Tower of the World Trade Centre in Lower Manhattan. After having identified the "pervasive problems" of managing and sharing information across a large and unwieldy government, one of the conclusions of the 9/11 Commission (Kean et al., 2004) was that:

"The U.S. government has access to a vast amount of information. But it has a weak system for processing and using what it has"

Clearly, a tragic backcloth highlighted the issue of data integration as never-before.

In commercial environments, such as the Financial Services industry, organisations have identified the need to develop a complete, product-wide understanding of their customer base, to enable cross-selling and maximise customer retention (Liew, 2003). Organisations have determined that differentiation from competitors can be delivered by a company's ability to construct a single view of each customer, based on its "Business Intelligence" (BI) (Hill, 2004).

BI relies on data warehousing but a major challenge encountered by companies is the proliferation of "disconnected" customer product systems, simply because individual departments e.g. mortgage, savings and insurance, have historically operated autonomously – often with each pursuing its own IT infrastructure (Lamb and Davidson, 2000). Thus BI seeks to harness a spectrum of information to derive knowledge. Clearly, the absence of reliable, relevant and related data, to drive customer targeted and effective marketing, could spell disaster in an increasingly competitive commercial market where corporate growth demands attraction of new customers and retention of existing customer base.

In 2003 SAS commissioned a European survey among 500 marketing directors from companies across France, Germany, UK, Italy and the Netherlands (SAS, 2003). The survey revealed that 66% of organisations were of the opinion that data integration and quality issues had an impact on company profitability and further, it was considered that data integration was the greatest barrier to achieving return on investment in data warehousing and business intelligence software. In addition, 67% felt that customer satisfaction and loyalty with their organisations suffered as a result and only 25% of respondents had a "great deal" of confidence in data quality in their businesses.

The Costs of Integration

It can be seen that data integration represents one of the most significant challenges facing the information technology community. In the SAS survey, 52% of European organisations considered that poor integration of diverse systems was a major contributor to inaccurate data and caused significant problems, although no cost predictions were highlighted. However, according to Michael Brodie,

Chief Scientist at Verizon IT, one of the world's largest communications organisations following the merger of Bell Atlantic and GTE, the cost of data integration and improving data quality has been estimated at £500m p.a. (Brodie, 2003).

In a similar study, The Yankee Group conducted a survey of 400 enterprises to uncover the "hidden costs" of integration (Yankee-Group, 2004). Their findings indicated that labour represented approximately 50% of total costs including software and hardware.

Conclusions

Data integration remains the first stage in the process to enable organisations to deliver meaningful, comprehensive, and relevant information and knowledge provision. Information systems must be able to address the real world data complexities, now demonstrated in the Web environment, and database integration has the opportunity to address new horizons, in the form of emerging Web technologies.

The Semantic Web may represent the next generation of the Web to provide the universal medium for the exchange and integration of data. The challenge and costs of integration demand an examination of alternatives.

References

BANSLER, J. P., DAMSGAARD, J., SCHEEPERS, R., HAVN, E. & THOMMESEN, J. (2000) Corporate Intranet Implementation: Managing Emergent Technologies and Organizational Practices. *Journal of the Association for Information Systems,* **1**(10), pp. 1-41.

BAR, F., KANE, N. & SIMARD, C. (2000) Digital Networks and Organizational Change: The Evolutionary Deployment of Corporate Information Infrastructure. In: *Proceedings of International Sunbelt Social Network Conference. Vancouver, British Columbia, April 13-16 2000*.

BERNERS-LEE, T., HENDLER, J. & LASSILA, O. (2001) The Semantic Web. *Scientific American.*

BERNSTEIN, P., BRODIE, M., CERI, S., DEWITT, FRANKLIN, M., GARCIA-MOLINA,

H., GRAY, J., HELD, J., HELLERSTEIN, J., JAGADISH, H. V., LESK, M., MAIER, D., NAUGHTON, J., PIRAHESH, H., STONEBRAKER, M. & ULLMAN, J. (1998) The Asilomar Report on Database Research. *ACM SIGMOD Record,* **27**(4), pp. 74-80.

BHATTACHERJEE, A. (1998) Management of Emerging Technologies: Experiences and Lessons Learned at US WEST. *Information & Management,* **33**(5), pp. 263-272.

BRODIE, M. L. (2002) The Grand Challenge of Information Technology. In: *Proceedings of an Invited Talk, OntoWeb Meeting. Innsbruck, Austria, December 16-18 2002.* Verizon Information Technology.

BRODIE, M. L. (2003) *Enterprise Level Integration - It's Not About Technology* [online]. Verizon Information Technology. Available from: http://www.plmdc.engin.umich.edu/BrodieIntegratio n.pdf. [Accessed 29 December 2004].

ELIOT, S. & BARLOW, J. (2002) KM Can Enable the Enterprise. In: *Proceedings of KMWorld & Intranets Conference. Santa Clara Convention Center, Santa Clara, California, 31 October 2002.* IBM Lotus Software Group.

HENDLER, J., BERNERS-LEE, T. & MILLER, E. (2002) Integrating Applications on the Semantic Web. *Journal of the Institute of Electrical Engineers of Japan,* **122**(10), pp. 676-680.

HILL, J. (2004) *The Future of Data Integration Technologies* [online]. Stamford, CT, META Group Inc. Available from: http://www.dmreview.com/whitepaper/WID100658 1.pdf. [Accessed 8 January 2005].

HORGAN, T. (1997) *Developing Your Intranet Strategy and Plan* [online]. CIO Magazine. Available from: http://www.cio.com/WebMaster/strategy/tsld046.ht m. [Accessed 4 January 2005].

JOHNSON MCMANUS, D. & SNYDER, C. A. (2003) Synergy between data warehousing and knowledge management: three industries reviewed. *Int. J. Information Technology and Management,* **2**(1/2), pp. 85-99.

KARLSBJERG, J. & DAMSGAARD, J. (2001) Make or buy - A Taxonomy of Intranet

Implementation Strategies. In: *Proceedings of 9th European Conference on Information Systems. Bled, Slovenia, June 27-29 2001.* Dept. of Computer Science, Aalborg Univ., pp. 579-592.

KEAN, T. H., HAMILTON, L. H., BEN-VENISTE, R., KERREY, B., FIELDING, F. F., LEHMAN, F. M., GORELICK, J. S., ROEMER, T. J., GORTON, S. & THOMPSON, J. R. (2004) *The 9/11 Commission Report* [online]. Washington, DC, National Commission on Terrorist Attacks Upon the United States. Available from: http://a257.g.akamaitech.net/7/257/2422/22jul20041 147/www.gpoaccess.gov/911/pdf/execsummary.pdf. [Accessed 4 January 2005].

LAMB, R. & DAVIDSON, E. (2000) The New Computing Archipelago: Intranet Islands of Practice. In: *Proceedings of IFIP Working Group 8.2 Conference - The Social and Organizational Perspective on Research and Practice in Information Technology. Aalborg, Denmark, 10-12 June 2000.* Kluwer Academic Publishers, pp. 255-274.

LIEW, M. (2003) *Failure is Not an Option - Computerworld Singapore* [online]. ITWorld.com. Available from: http://www.itworld.com/App/103/031224failure/. [Accessed 6 January 2005].

MCILRAITH, S., SON, T. C. & ZENG, H. (2001) Semantic Web Services. *IEEE Intelligent Systems,* **16**(2), pp. 46-53.

NEWMAN, B. D. & CONRAD, K. W. (2000) A Framework for Characterizing Knowledge Management Methods, Practices and Technologies. In: *Proceedings of Third International Conference on Practical Aspects of Knowledge Management (PAKM2000). Basel, Switzerland, 30-31 October 2000.* pp. 16.1-16.11.

POWELL, T. A. (2002) *Web Design: The Complete Reference, 2nd Edition.* McGraw-Hill/Osbourne.

ROB, P. & CORONEL, C. (2002) *Database Systems Design, Implementation and Management.* Fifth Edition ed.: Course Technology.

SAS (2003) *Reducing the Cost of Data Integration, Increasing ROI* [online]. Carey, NC, SAS Institute. Available from: http://www.sas.com/news/newsletter/business/2003_ 07_15.html. [Accessed 7 January 2005].

SMEATON, J. (2002) *IBM's Own Intranet: Saving Big Blue Millions* [online]. Intranet Journal. Available from: http://www.intranetjournal.com/articles/200209/pij_09_25_02a.html. [Accessed 8 January 2005].

STONEBRAKER, M., AGRAWAL, R., DAYAL, U., NEUHOLD, E. J. & REUTER, A. (1993) DBMS Research at a Crossroads: The Vienna Update. In: *Proceedings of 19th International Conference on Very Large Databases (VLDB). Dublin, Ireland, 24-27 August 1993*. pp. 688-692.

W3C (2004a) *Architecture of the World Wide Web, Volume One* [online]. World Wide Web Consortium. Available from: http://www.w3.org/TR/2004/REC-webarch-20041215/. [Accessed 4 January 2004].

W3C (2004b) *OWL Web Ontology Language Guide* [online]. World Wide Web Consortium. Available from: http://www.w3.org/TR/owl-guide/. [Accessed 29 October 2004].

W3C (2004c) *RDF Primer* [online]. World Wide Web Consortium. Available from: http://www.w3.org/TR/rdf-primer/. [Accessed 15 November 2004].

WAGNER, W. P., CHUNG, Q. B. & BARATZ, T. (2002) Implementing corporate intranets: lessons learned from two high-tech firms. *Industrial Management & Data Systems,* **102**(3), pp. 140-145.

WARD, T. (2001) *Measuring the Dollar Value of Intranets* [online]. Intranet Journal. Available from: http://www.intranetjournal.com/articles/200104/ii_04_25_01a.html. [Accessed 4 January 2005].

YANKEE-GROUP (2004) *Uncovering the Hidden Costs in Data Integration* [online]. Boston, MA, The Yankee Group. Available from: http://wp.bitpipe.com/resource/org_1013484794_935/Data_Integration_Executive_Report_Yankee_Group_Bitpipe.pdf. [Accessed 7 January 2005].