Information Quality
A New Academic program at the University of Westminster

Tereska Karran, University of Westminster, karrant@wmin.ac.uk, Charles Poulter, University of Westminster, poultec@wmin.,ac.uk, Patrick Lees, University of Westminster, leesp@wmin.ac.uk

ABSTRACT

The University of Westminster has an ongoing program producing postgraduate courses for management and industry. We have extended this to Business Intelligence by introducing postgraduate courses at MSc level in Enterprise Information Systems, Data Mining, Decision Support and, most recently, Information Quality. These courses represent our research commitment in this area as well as our response to a perceived industry need in the area. The curriculum for the course reflects the fact that information quality is part of a wider enterprise need and that data quality managers will need to be part of a strong enterprise team which ensures that organizational needs are met at both user and data layers.

Key words: Information Quality, data warehousing, enterprise architectures, business intelligence, postgraduate curriculum

The adoption of new technologies allows enterprises to gain competitive advantage over their rivals. However, there is a complex relationship between the technology of the organisation and its social structures. If the technology changes, this affects the social structures [Espejo & Gill (1999)] and it is essential to factor this in to the adoption of the new technologies. As business information systems mature and develop, the theory underpinning the subject area becomes more clearly differentiated into distinct streams or pathways. This can be seen by the way that, in response to socio-technical changes in the employment sector, the school of Computer Science at the University of Westminster first split into several departments, including Computing and Information Systems and in 2006 was reorganised as the School of Informatics, with a concomitant shift of focus. This shift is in response to the fact that the subject matter of computing has become more and more about the delivery of information at all levels of the organisation rather than a pure technical subject.

One particular focus in both research and teaching at the School of Informatics has been the delivery of business intelligence. The delivery of this subject matter has resulted in the development and delivery of four distinct MSc programs, each of which is aimed at a different employment sector and focuses on a different aspect of the delivery of intelligent information within an enterprise. The quality of information within a business is at the core of business intelligence, and as such managing quality has been an endemic part of our teaching. However, as organisations become increasingly global and require more complex views of their information, it has become a subject specialism in its own right.

Our view is that business intelligence is the key to a successful enterprise. Several serious failures within organisation can be pinpointed as originating in the mismanagement of information [Sercu et al. (2006)], and it is particularly difficult to establish the causes because the architectural flows of information, and its quality, cannot be easily traced. It is clear that a successful enterprise needs a transparent data architecture together with a clear strategy for the management of information quality.

2. PEDAGOGIC AND RESEARCH BACKGROUND

The disciplinary sources for our view of information quality arise from several research and pedagogic origins.

Firstly, the Health and Social Care Modelling Group (HSMG) has a long research tradition in medical informatics, based on our proximity to the large teaching hospitals in central London and the historical needs of the National Health Service[1].

Secondly, we have a strong analytical research background based on statistics and operations research. Current research activities are in the areas of multivariate time series, correlation theory, data and text mining and computer simulation.

Thirdly, we have developed a generic intelligent architecture for use in enterprise applications [Karran et al (2003)] and which is being applied in a variety of research contexts [Karran & Wright (2007)].

Our expertise has been closely linked to the developments in industry and we have collaborated with the Defence Research Agency, Eurocontrol Air Traffic Control Authority, as part of WestmARC[2].

It became clear that there was a growing need for business intelligence experts in these fields. As a result, we have developed a program of business intelligence postgraduate teaching in order to cover this fast-growing field. The MSc in Information Quality is the most recent addition to the existing university offering. It is closely related to the MSc in Enterprise Information Systems, the MSc Data Mining and the

[1] http://www2.wmin.ac.uk/hscmg/people.htm
[2] KTPonline.org

MSc in Decision Sciences all of which are currently running. The additional modules taught in the new MSc further widen students' choices on these programs as well as providing a clear developmental pathway for graduates wishing to specialise in the information quality field. The course itself addresses the clear need for specialists in information quality while its specific modules can also provide a good introduction to the field for students undertaking related programs of study.

From a wider perspective, a postgraduate course in Information Quality has an important part to play in an international initiative towards developing expertise in managing and mining the vast stores of both structured and unstructured data needed by organisations within a global economy. Such data may be located across multiple servers both within and outside the organisation and information managers will need to resolve multiple data quality issues before information can be extracted from it effectively. The new course has been designed to meet the growing industry need for the provision of quality information for all levels of Business Intelligence Systems and enterprise applications. It is practitioner oriented and provides marketable skills relevant to the management of information and provision of accurate business intelligence for existing Information Systems. The course provides a balanced study, which aims at producing graduates are capable of

- Providing secure, accurate, on time and up-to-date information for all levels of the business, taking into account all the issues involved;
- Utilising their problem solving skills and their knowledge of various methodologies, techniques and tools to deliver quality information solutions;
- Creating models and architectures to restructure the Information Systems for the optimal delivery of Business Intelligence and Performance Management;
- Studying the context within which the restructuring information systems takes place, i.e. identifying the strategic, managerial and operational activities involved and delivering appropriate information for all monitoring and controlling activities;
- Critically evaluating alternative design and implementation strategies and the impact of the emerging technologies advances on delivering information;
- Managing independent in-depth analysis of a chosen topic involving the delivery of quality information involving information resources outside a teaching environment.

3. THE ACADEMIC PROGRAM

Our view of data quality is an offshoot of our understanding of the enterprise as a type of complex non-deterministic system. Complex system theory allows us to model an enterprise architecture as composed of multiple interacting and semi autonomous databases wrapped within a cell structure [Ramanathan (2005)]. It is possible to test the macro properties of the architecture without getting involved in the internal behaviour of components. One of the key macro properties of the architecture is the quality of information flow. It is our view that there are two elements to assessing this, namely the qualities of the information itself, and its fitness for purpose based on the profile of the user who intends to use it.

The core of the program is based on understanding the enterprise architecture consisting of layers, which can be loosely described as transactional, decision support, executive information and knowledge discovery. Figure 1 below shows a typical decision support layer. An important advantage of layering the information flows is that it becomes possible to add new component 'cells' and test their effects on the enterprise with the minimum of side effects [Bedau (1997)]. Once this has been established, it is possible to design and test both individual elements and the whole system in a variable environment [Bedau (1996)]. The management of information quality is therefore one of the key properties of the complex enterprise architecture and the data quality manager is one of a wider team managing business intelligence across the enterprise.

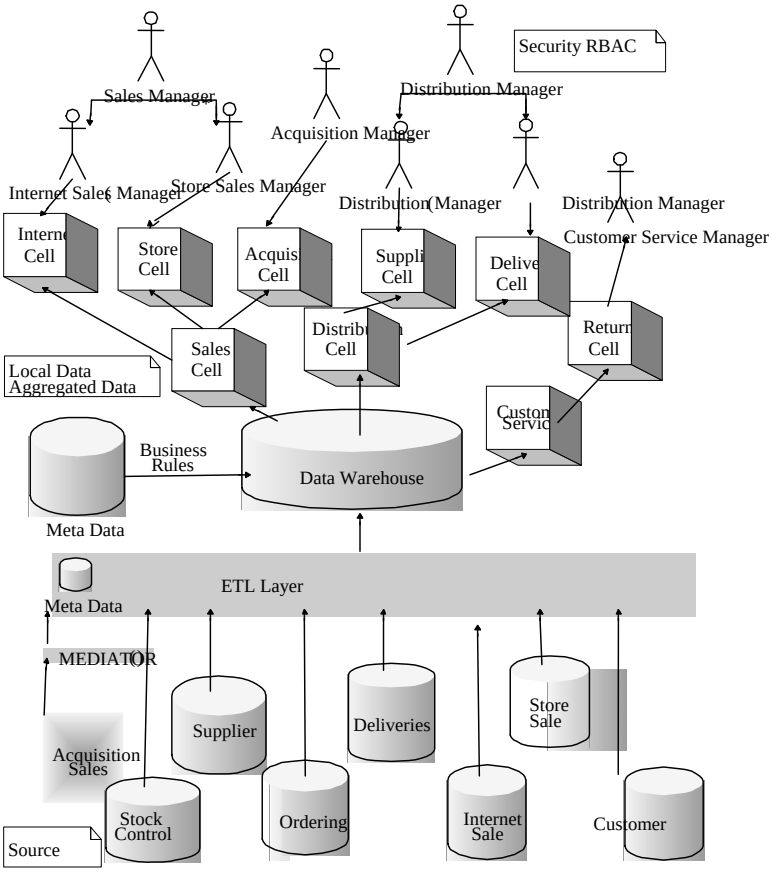

Figure 1. Data Warehouse Architecture showing a Decision support Layer composed of multiple cells

Each layer consists of multiple interacting component cells which manage the information needed by a specified user. The user is defined by the user profile which consists of security protocols and permissions as well as a history of accesses. Within each layer the component cells continuously adapt

their behaviour based on critical tolerances which are managed by cells at 'senior' layers.

The complete process involves several views of the original input data each with their own data quality requirements, each of which may be modified as a result of the intelligence gained by cells at the higher layers. The process has to be strictly managed in terms of data quality because although the source input data is likely to be highly detailed, it becomes highly generalised as it moves through the decision layers[3].

It should be clear that the higher layer intelligence cells may store base their decisions on generalised versions of core information, depending on what is needed for extrapolating rules about the environment, monitoring patterns of behaviour, and adapting to circumstances making data quality a fundamental property of good intelligence. Data quality is also vital to ensure a holistic and speedy adaptive response once new information has been discovered. This is managed by ensuring that there are efficient feedback loops linking the layers. The whole process is dynamic since decisions made must result in improvement. Otherwise they will be replaced by better adaptations based on analysis by higher layers. This data quality cycle holds true for any complex information system and not only business enterprises.

4. Managing Data Quality

The MSc program for Information Quality allows students to select a pathway through our business intelligence modules (see list in appendix 1). However, the students take a core of modules which provide them with a foundation in the two elements which we have identified as key to the management of information quality within the organisation. These are the **user**, who has a profile which is managed autonomously within the enterprise architecture, and the **data**, which has a quality profile independent of the user, based on formal analytic principles. This separation of function is shown in figure 1 above.

4.1. Managing Data Quality within the user Profile:

We distinguish between the intrinsic quality if the data and the data requirements of the user. For example, in figure 1 above, each user has a security profile detailing accesses and permission to cells. At each layer there is a data mining tool which analyses user accesses and transactions. It should help to ensure that users receive the right data in a timely manner and at the same time, monitor for potential unauthorised access.

Students will need to distinguish between the tools and processes needed for each layer and the types of data and user profiles which can be constructed for different purposes.

4.2. Managing Data Quality within the data warehouse and the data cell

Students on the MSc in Information Quality will be trained in areas of statistics and operations research appropriate to the subject area including:

a) Sampling methods and theory suitable for monitoring data accuracy and relevance on an ongoing basis, including stratified and quota sampling, ratio and regression estimates, sample size calculations, error estimation in combined estimators and special methods for rare characteristics [Cochran 2001].

b) Weighting methods for combining dashboard values into singled metrics, including principle components and the Analytic Hierarchy Process [Saaty 2001]

c) Forecasting and time series theory relevant to timeliness [Chatfield 2001]

d) Industrial control theory such as control charts and Taguchi methods relevant to the study of changes in metrics over time and to the general improvement of information quality.[Moen et al. 1991]

They will also cover enough SQL to write their own macros to clean data in various ways.

5. The MSc In Information Quality

The course is designed around sets of modules which are open to students according to their abilities and interests as shown in figure 2.

Core modules	***Option modules for Business Analysts (BA)***	***Option modules for all other students (DBA+)***
Postgraduate project Preparation & planning	Statistical data mining	Project management
Postgraduate Project: Final report and viva	Text mining	Text mining
Corporate systems and data management	Statistical inference	Database languages
Data warehousing and data mining	Data mining applications	Enterprise resource planning systems
Interoperability in data–centric applications	Project management	Enterprise applications
Information quality		
Information security		

Figure 2: Module Pathways through Information quality

The course will consist of a common core of five modules plus two option modules, and a major project. There are two sets of option modules. The sets are different for the business–analyst (BA) type of student and the database administrator type (DBA+) and other types.

The more technical modules in the BA set will only be open to students who can show that they have the necessary background or ability.

7. Conclusions

The Dearing National Committee of Inquiry into Higher Education (1997) had a profound effect on all aspects of

[3] We assume that information is logically structured so that the steps from the detail to the general are logical.

current UK Universities' practice. The recommendations included, most notably: a greater emphasis on teaching quality, an explicit rationale for a student's skills development, more effective partnership with industry, a clear statement of a course's intended learning outcomes with a shift from what topics are taught to what students can do practically and intellectually, a standard framework for all University qualifications and a progressive development of standard discipline benchmarks with which to reference any particular course's content and expected outcomes [Lees (2003)].

The University of Westminster, in common with other UK Universities, operates quality processes for the introduction, maintenance and renewal of its courses. The national context is the audited adherence to best practice, as defined by the UK Quality Assurance Agency in their *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education*, on behalf of the Higher Education Funding Council of England whose statutory duty is the allocation of public funding for the English Universities. Scottish, Irish and Welsh Universities operate in a similar but not identical context.

The MScs in Business Intelligence will continue to be revalidated and modified in accordance with changes and developments in research and modifications to best practice. The courses face automatic re-validation at the end of every five years, and every year we request minor modification to the programs if these are needed. We look forward to implementing the new program as we expand our provision to meet future change and progress.

APPENDIX 1 -MODULAR PROGRAM

Statistical inference

This is a modified version of the existing module to include study of Taguchi methods, Shewhart and CUSUM quality control charts and stratified and other sampling methods, as well as some of the existing material on probability and random variation, distributions and hypothesis testing. It will depend heavily on practical work with SAS including extensive use of PROC SQL.

Statistical data mining

This module covers the standard methods of statistical data mining plus some leading–edge material, with particular emphasis on techniques useful in management, healthcare and business generally. It uses software tools to provide practical experience of the issues involved.

Postgraduate project — preparation and planning

The module provides students with the necessary skills to begin research. It prepares students for further studies and gives them the foundation to propose, plan and organise a major piece of research activity. Students learn how to identify areas appropriate for research, to use research methodologies for qualitative as well quantitative analysis and hypothesis testing, and to acquire/improve academic report writing skills. The final output of the module is a project proposal that each of students will prepare, which could be used as the foundation for a postgraduate project.

Postgraduate project

The project module plays a unifying role within the course and aims to encourage and reward individual inventiveness, application and effort. It may take a variety of forms and provides students with the experience of planning and bringing to fruition a major piece of individual work. Although projects are not expected to be wholly original pieces of work, students are expected to show that they have exercised initiative and worked independently. The scope of the project is not only to complete a well-defined piece of work in a professional manner, but also to place the work in the context of the current state of the art of the subject area.

Corporate systems and data management

This introductory module covers theoretical and practical issues related to DBMS and data models. It discusses and evaluates the underlying technologies used in capturing and maintaining corporate data. Pursuing this, the evolution of database management systems, their components and the functionality thereof are discussed, along with some of the predominant data models. The module addresses issues related to conceptual data modelling, practical database design, and current trends in corporate systems database design.

The module introduces a database language for the definition and manipulation of data constructs in the context of a major commercial database system and it addresses issues related to procedural aspects and client/server programming. The exercises and materials used in the delivery of the module are based on Oracle and SQL.

Data warehousing and data mining

The module discusses and addresses recent technological developments in the database fields of data warehousing and data mining, both of which aim at extracting information from the overwhelmingly large amounts of data that today's IS are capable of collecting. The data warehousing part focuses on the multi-dimensional analysis of data, whereas the data mining part focuses on the types of models used for identifying compressed representations of data.

Interoperability in data–centric applications

This module provides an analysis of the problem of interoperability in data centric applications and gives an insight into different approaches that have been used to address the problem in the last decade. The module also emphasises the impact of internet technologies on the interoperability of current database systems. It discusses standards for data interchange and gives guidelines to solve interoperability of distributed heterogeneous database applications using XML–based web services.

Enterprise resource planning and systems

The module focuses on ERP systems and database integration, providing an overview of the aims and modular functionality of ERP systems, through lectures and hands–on experience. The module addresses problems associated with the design and implementation of ERP solutions, including vendor

selection, re–engineering and project management. It also examines the evolving relationship between ERP and e-business, customer relationship management and supply chain management.

Enterprise applications

The precise content of this module depends on current applications and issues in enterprise, and on the guest instructors/speakers and students' interests. Possible topics include: performance management, compliance issues in managing data, data protection and data security in warehouse systems, customer relationship management, supply chain management and e-business. Teaching includes seminars led by external speakers, by researchers and experts in the field. The unit works through a number of case studies and looks at some of the latest industry applications using project deconstruction in order to understand how ERP strategies have affected business. Students familiarise themselves with different solutions for SMEs and large enterprises and assess the success of the different strategies employed. Students will provide an ERP solution and business report for a case study.

Data mining applications

This module employs a seminar style of presentation to introduce current applications and issues in data mining, including web search engines and web mining. The contents are based on case studies related to current business and management issues in marketing (e.g. CRM), finance, healthcare, or biomedicine. Attendance at seminars given at the university and outside the university will be expected.

Text mining

The module provides essential text mining and information retrieval knowledge, with particular emphasis on techniques useful in scientific research, medicine and business.
Clustering and classification of documents are studied using a text mining tool and natural language processing is covered via Python/NLTK. There are guest lectures on copyright, freedom of information and legal issues and an introduction to the semantic web.

Project management

This module focuses on the importance of project management within organisations and examines why project management continues to be a growth discipline. The role of the project manager is examined together with techniques used for planning, scheduling, monitoring and controlling projects throughout the project life cycle. This module uses PRINCE2 as a framework for understanding the key issues, and provides students with practical experience in using project management software tools for project scheduling.

Information Quality

This module provides a deeper understanding of issues involved in delivering clean, integral and up–to–date data. Amongst the important topics studied will be: slowly changing dimensions, static and dynamic data, the creation of a meta–data repository, managing data history, consolidation and reconsolidation, compliance and data protection and freedom of information. The module will use the data quality tools where relevant.

Information Security

This deals with data integrity and role based access controls, managing a metadata repository, delivering data in batch form and in real time to cubes and different dashboards and scorecard type applications. It also covers security and how to deliver quality data in a timely way when managing accounts and financial data. This module uses a Case tool to create an enterprise architecture with users' profiles etc.

REFERENCES

Bedau M.A., (1997) Emergent Models of Supple Dynamics in Life and Mind. *Brain and Cognition* **34**: 5-27.

Bedau**,** M. A., (1996) The Extent to which Organisms Construct their Environments Adaptive Behaviour **4**: 476-482.

Cochran, W G (2001) *Sampling Techniques*, Wiley

Chatfield, C (2001) Time-series Forecasting, Chapman and Hall/CRC Press

Espejo, R., Gill, A. (1997) The Viable System Model as a Framework for Understanding Organisations. Phrontis .com.

Karran, T., Wright, S., (2007) Towards Database Autonomy: An Approach Using Complex Organic Distributed Architecture (CODA) *EEE'07 World Comp 07*

Karran, T., Madani K., Justo, G. R., (2003) CODA - Self Learning and Adaptive Systems in Dillinger M. (ed) Software Defined Radio, Wiley.

Lees P.F., Ptohos T. & Juric R. 2003 Experiences of revalidating the Undergraduate and Postgraduate Courses within the Information Systems Curricula at the University of Westminster, UK. *Journal of Computing and Information Technology – CIT 11,2003, 3, 1-10.*

Moen, R D; Nolan, T W & Provost, L P (1991) Improving Quality Through Planned Experimentation, McGraw-Hill

Rees, M. and Dineschandra, J. (2005) Monitoring Clinical Performance: The role of software architecture. *Journal of Healthcare Management Science.* Vol 8, p197 – 203.

Saaty, T (2001) Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World, *Analytic Hierarchy Process Series* Vol. 2

AUTHORS:

Dr Tereska Karran (karrant@wmin.ac.uk).Course leader for MSc Enterprise Information Systems, MSc Information Quality. Research interests include: Business intelligence, Business intelligent architectures, Public private partnership performance monitoring, Autonomous monitoring systems, Managing and monitoring data quality and Personalisation systems.

Charles Poulter (poultec@wmin.ac.uk). Course leader for MSc Decision Science. Research interests include Data and text mining, Data quality metrics, Multivariate statistics and Statistical computing.

Patrick Lees (Leesp@wmin.ac.uk) Head of Department of Information Systems and Computing. Research interests

include: ICT public sector procurement, computer forensics, systems modeling, natural language processing.

IQ Education Key Events

Date	Event Description
Early 1990's	Larry English founds Information Impact International, a company whose mission is to provide quality education and consulting in information management.
March 1995	The Data Quality Journal published by James Hurysz releases its first issue. This Journal would remain in existence until 2002.
October 25-27, 1996	First International Conference on Information Quality (ICIQ) held at MIT, Cambridge, Mass. ICIQ is now in its 12th year.
1997	Thomas Redman founds Navesink Consulting Group. Among its many IQ services, Navesink Consulting Group creates the Data Quality College, a series of seminars focuses on data quality issues.
September 1999	Omar Khalil, Diane Strong, Beverly Kahn and Leo Pipino author the paper entitled "Teaching Information Quality in Information Systems Undergraduate Education" in Informing Science, Vol. 2, No. 3. This article identifies the gap between the needs of organizations for high-quality information and the skills of university graduates from Information System program.
September 2001	Craig Fisher offers the first undergraduate course on Information Quality to IS seniors at Marist College, Poughkeepsie, NY. This course would become a regularly offered elective and form the basis for the book "Introduction to Information Quality".
October 29-31, 2001	Europe's launches its own annual Data Management and Information Quality Conferences in London, UK.
July 15-17, 2002	The MIT-IQ program offers its first three day workshop entitled MIT-IQ for Executives.
November 8-10, 2002	WooYoung Chung, Craig Fisher, and Richard Wang present their work on "Redefining the Scope and Focus of Information-Quality Work: A General Systems Perspective" at ICIQ . This paper classifies IQ skills into three categories: Technical, Adaptive, and Interpretive.
May 19-23, 2003	The MIT-IQ program offers its first week-long workshop entitled "IQ-1: Principles and Foundations"
January 2004	International Association for Information and Data Quality, a not-for profit, vendor-neutral professional society of people passionate about improving information and data quality was chartered.
June 7-8, 2004	CAiSE sponsors the Workshop on Data and Information Quality in Riga, Latvia. A second DIQ workshop is held by CAiSE in 2005.
June 18, 2004	SIGMOD sponsors the International Workshop on Information Quality in Information Systems (IQIS) in Paris, France. IQIS 2004 is followed by IQIS 2004 and IQIS 2005.
April 2005	Diane Strong, Craig Fisher, David Feinstein, and Herbert Longenecker publish their article "Teaching, Learning, and Curriculum Development to Support Managing Information as a Product" in the AMIS Monograph on Information Quality. This article explores various strategies for incorporating IQ into an IS or Business UG Education.

August 27, 2006	University of Arkansas at Little Rock launches the first Master of Science in Information Quality. 25 students enroll.
Fall 2006	At Northeastern University, Dr. Yang Lee teaches the Honor Seminar: HNRU302 Topics in Research and Inquiry: Focus on Analysis, Information Quality: Technology and Philosophy
January 2007	The International Journal of Information Quality published by InderScience releases its first issue.
January 2007	University of South Australia, School of Computer and Information Science announces plans to launch a Master of Science in Information Quality
March 2007	University of Westminster in the UK announces plans to launch a Master of Science in Information Quality
April 2007	The ACM Journal of Data and Information Quality begins accepting submissions.