ASMG
ADVANCED SYSTEMS MANAGEMENT GROUP

DATA INTEROPERABILITY

White Paper #2
22 January, 2003

Background

The current focus of many organizations is the integration of their Information Management (IM) environment. This activity is the precursor to the development of a knowledge management capability and the foundation of a competitive advantage in an ever-expanding global market. There is a growing realisation by many senior executives that a company's ability to develop and retain a competitive lead hinges on its ability to provide a fast, effective and trusted exchange of information between key corporate personnel, strategic partners and trusted suppliers.

- **FAST**: The idea of moving large volumes of information quickly lies at the heart of almost two decades of activity. Terabytes of information can be moved virtually anywhere at the blink of an eye. Current estimates indicate that only 3% of the overall capacity of the existing backbone is being used. From this standpoint, it would appear that moving information quickly is not a big challenge for organisations, unless they are situated in remote or isolated locations.
- **EFFECTIVE**: Within the IM context, effectiveness refers to moving the right information; to the right place (person or people); in the right format; at the right time. However, it would appear that this aspect of Information Management is posing a true challenge for the majority of organisations. Few companies have a clear program for managing information across their enterprise. As an organisation grows, knowledge becomes more diverse, information systems become more complex and islands conflicting information became progressively more widespread. Integrating these isolated areas of information is proving to be a complex and costly exercise, particularly for organisations that have large investments in Information Technology (IT) and legacy systems.
- **TRUSTED** The ability to provide trusted access to information is a double-edged sword. The leading edge involves the distribution of required, timely and accurate information to personnel, strategic partners and trusted suppliers, such that it forms the basis for making sound operational and business decisions. The business risks associated with not having access to accurate information are comparable to the risks associated with unauthorized access to information. The reverse edge is in keeping classified, proprietary, confidential and private information out of the hands of people or organisations that do not need or should not have access to the information. The recent focus of the IT industry has been on controlling the information in-flight (movement of information between source and target). There are numerous commercial off the shelf (COTS) technologies designed to secure the information flow between locations, such as Public/Private Key Encryption (PKI), and Virtual Private Networks (VPN. A major issue that has attracted far less attention is the delivery (inadvertent or malicious) of unauthorized information over a secure channel. As the amount of information being transferred between computers without human intervention increases, this will prove to be a challenge for CIOs and CTOs.

Data Interoperability

The scope of this paper will be limited to the discussion to data interoperability. Data interoperability describes the ability to effectively exchange timely and accurate data between operational systems (OPS), Operational Data Stores (ODS), Data Warehouses (DW) and Information Portals, while retaining the original translation or understanding of the data. The ability to effectively exchange data forms a fundamental building block in enterprise information and knowledge capability. To support this assertion, the following definitions are provided:

- **DATA**: Raw facts and figures which are processed into information.
- **INFORMATION**: Summarized or processed data.
- **KNOWLEDGE**: The combining of information and experience into a theoretical or practical understanding.
- **CORPORATE KNOWLEDGE**: The sum of an organisation's knowledge and experience.

The key to developing enterprise-wide data interoperability includes:

- Developing a framework for integrating the organisation's IM and IT environments;
- Developing and maintaining a clear understanding of business requirements;
- Adopting a technique that effectively documents the referential, semantic and transactional completeness of the data exchange;
- Implementing a process and infrastructure for managing meta-information about the IM environment and retaining corporate knowledge;
- Implementing a process and infrastructure for maintaining, defining and auditing security, confidentiality and privacy; and
- Implementing a process for managing change in the IM environment.

Framework for Integration

The Information Management environment encompasses numerous components which cross multiple application, information, security and technology domains, as illustrated in Figure 2. Often, each of these domains is managed, supported and maintained by different groups, organisations or companies (out-sourced). Few organisations have the ability to develop a composite view that identifies the relationships and interdependencies between business processes, applications, information or technologies. Enterprise architecture provides a framework for capturing and organising corporate knowledge about a company's IM environment.

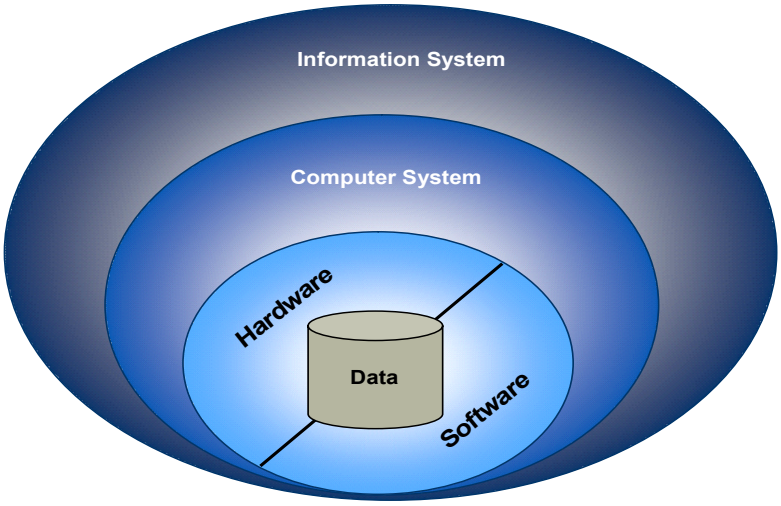

Figure 1: Knowledge Management System

After more than two decades of experience in systems engineering and system integration, the indisputable truth is that the effective capture, management, and dissemination of information is at the core of every successful integration project. Integration projects are executed by a multidisciplinary team comprised of business analysts, application architects, developers, database specialists, security specialists and test engineers, each of whom is responsible for specifying, designing or developing characteristics of system integration or data interface. Each of these team members requires access to current and accurate information that is developed during the integration or delivered by way of a separate on-going or delivered project. Information required by the team includes, but is not limited to:

- Source system data structures, target system data structures and applied technology (e.g. Database Management System, flat file, etc);
- Transfer form, format and frequency;
- Network considerations;
- Server considerations;
- API considerations;
- Corporate, sector or international standards;
- Security, confidentially and privacy considerations affecting the data/information transfers;
- Referential, semantic and transactional (business) integrity considerations;

- Business rules (e.g. integration of data from multiple sources, exception handling on the extraction and load methods);
- Test materials including requirements, test procedures, test cases, test data and test results;
- Applicable standards;
- Delivered models and documentation; and
- Delivered code.

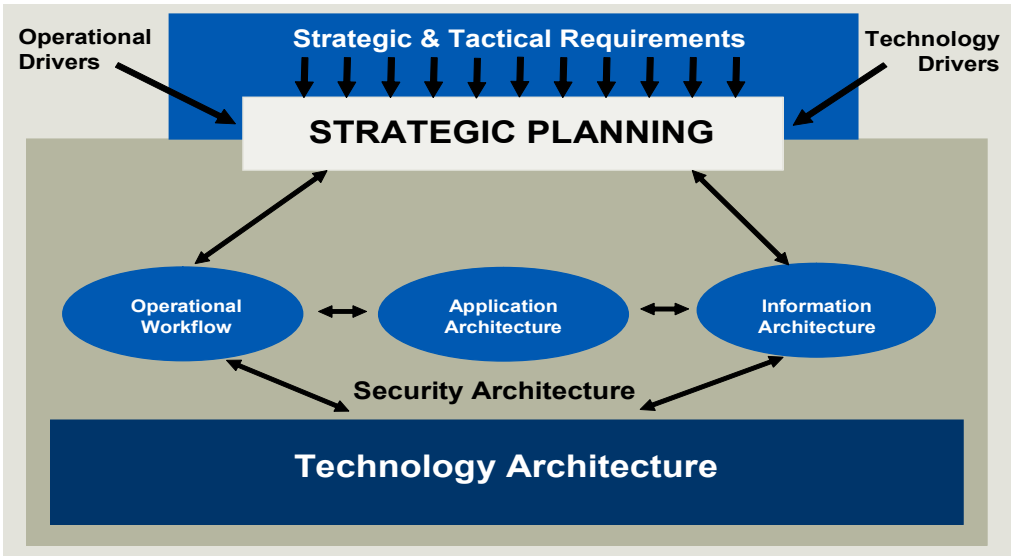

Figure 2: IM Architecture

The use of architecture to manage large-scale integration is not a novel concept. Adopted from other engineering disciplines, military and aerospace engineers incorporated architecture as a core component of the systems engineering process. In recent years, the use of enterprise architecture (viewing the entire information environment as a single integrated information or knowledge management system) has been enthusiastically promoted within the IM/IT community. ASMG does not promote a particular architectural framework for our clients, as we prefer to work with one that has been pre-selected by a client's organisation. In recent years, our clients have adopted several of these frameworks, with the most popular being those created by Zachman and Meta-Group.

As competitive pressures compel greater integration of IM components, it is becoming increasingly difficult and costly to deliver and maintain enhanced IM capability. A major factor is the increasing complexity and growing number of interdependencies. For IM groups to manage within an environment that consists of increased integration, greater complexity, increasing demands and

shrinking resources, it is more important than ever for organisations to maintain a persistent view of: how the IM capability (e.g. processes, functions, applications and technology) is deployed and how and where this data is being used and disseminated.

In addition to facilitating the integration process, maintaining an enterprise architecture (corporate knowledge) improves management's ability to mitigate several of the most pervasive IM risks. These include:

- Cost of staff turnover;
- Cost of continual business and technical change;
- Inability to validate deliverables against user and project requirements;
- Escalating operations and maintenance costs of the IM environment; and
- Increasing the risk of proprietary, confidential or private information inadvertently or maliciously being released out of the IM systems and portals.

Clear Understanding of Business Drivers

The requirements for data interoperability exchange between business units, strategic partners and supply-chain partners, are based on business requirements, not technical requirements. Far too many organisations jumped into electronic commerce (e.g. Business to Business [B2B], Business to Client [B2C] and Customer Relationship Management [CRM]) without clearly defining their objectives and requirements. These organisations were enticed by industry hype and technology-based solutions that promised a quick and painless transition to the e-world. In reality, this migration has proven to be far from painless and rarely has it delivered the expected riches. It has been repeatedly demonstrated that in order for an organisation to develop a project with favourable results, a clear plan based on sound business objectives must be implemented.

Many organisations have adopted a technology-based approach, viewing information distribution as a technical issue to be addressed by commercial off the shelf (COTS) technology. Subsequently, they have invested heavily in commercially available solutions such as the Enterprise Resource Planning (ERP) tool of the 90s. However, these COTS solutions have revealed several significant and often hidden risks that are usually identified after initial deployment, including:

- Inability to maintain corporate knowledge;
- Inability to audit the IM environment
- Inability to audit flow on information;
- Inability to control support and maintenance cost;
- Inability to replace aging applications and technology; and
- Inability to complete development, support and maintenance contracts resulting from technology and vendor lock-in.

In many instances, the COTS (e.g. middleware or Extraction Transformation and Load [ETL] tools) applications were not at fault. They delivered according to stated capability, but did not live up to expectations or requirements.

Numerous publications have identified the requirement for businesses to develop knowledge management capability and channel information into intellectual capital. However, there is little discussion on how business requirements are effectively captured and translated into interface definitions, or how requirements are managed, maintained and tracked throughout the development and maintenance life-cycles. This is perplexing since the inability of organizations to effectively manage requirements is repeatedly identified as the number one reason for IM/IT failing to deliver needed capability, on-time and on-budget.

ASMG focuses its process on the identification, definition and management of requirements for business and operational interfaces in a manner that: is technology independent and does not rely on specific technology implementations of factory objects, XML, SQL, XSQL, Data Replication, message broker, CORBA or other ETL/middleware tools. Our primary goal is to maximise the retention of corporate knowledge.

Requirements for Data Exchange

IM architecture is derived from business requirements, operational drivers including market/sector considerations, and technology drivers (e.g. available COTS and industry standards). By integrating several architecture views, a team of business and policy analysts and technology specialists can effectively describe interface requirements based on function, operation, performance and security). The key to this approach is to establish a framework through which these multidisciplinary teams can effectively communicate and integrate requirements.

By refining architecture requirements within the same framework, a development team can deliver the interface while maintaining a direct link between the requirements and the delivered code. Ideally, these refinements can be captured within an engineering tool that automatically generates the code based on the specification and design information.

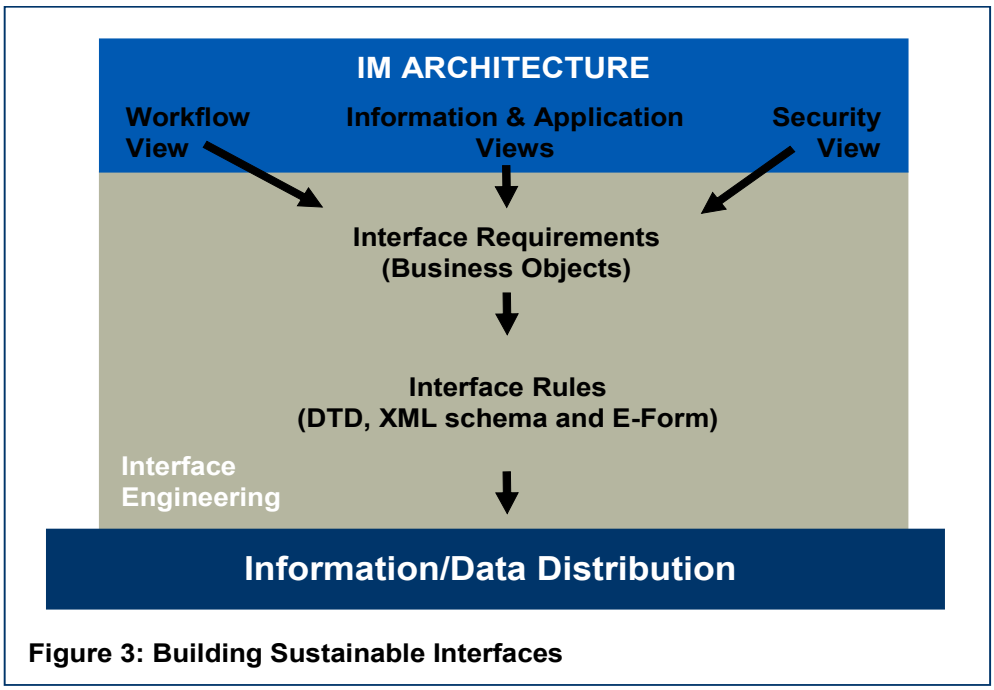

Figure 3: Building Sustainable Interfaces

The following architectural views are used to define business interfaces:

- **Workflow View**: Identifies the flow of information (interface) between business processes and organizations.
- **Application View**: Identifies the applications used to acquire, process and store business data and information. As part of the architecture model, application functions and components are linked to business functions and organisations. By providing these links, the architecture now maintains an understanding of which applications and components need to share information and what types of information they share, based on an identified business need. The application view also integrates an interface definition notation that permits analysts and designers to document the business and transformation rules for data moving between applications databases and information portals.
- **Information View:** Identifies the structure and format of the information being transferred. The data elements in the information view can be linked to the application views (application data model) and the workflow views (corporate data models and information flows) that identify the interfaces between databases, applications and organisations. The information view can also document interface standards (e.g. naming conventions or approved data tags such as XML tag value) and relationships or dependencies between data elements as part of the underlying data dictionary.

- **Security View:** Represents attributes and annotations on processes, interfaces and data elements which describe or specify security considerations related to the storage, processing and transmission of data.

Integrating the identified views over their conceptual, logical and physical levels of abstraction provides a direct link (traceability) between business requirements and deployed interfaces (code). More importantly, the architecture-based approach promotes the capture and maintenance of corporate knowledge and establishes an information asset that can be used to manage and audit the quality and security of an IM implementation.

Documenting Referential, Semantic and Transactional Complete Data Transfers

As identified, the application view should also provide the ability to define and document the extraction, transformation (integration) and load interfaces between corporate data sources including operational systems (OPS), operational data stores (ODS) and data warehouses. These components of information within an interface are the artifacts that must be managed. To completely describe a data interface we must address the persistent (physical) artifacts, the transient (logical) artifacts and the transactional (business) artifacts.

The objective for building artifacts is to move interface design from implementation and deployment phases back to design. This allows business analysts and IM designers to verify that the combination of data objects forms a complete business transaction that meets specified business requirements. The key benefit to this approach is the direct involvement of the business analyst in the process. Once complete, the interfaces are fully documented and traceable to business requirements.

The persistent model maps the physical database tables to persistent data objects. These objects and associated methods describe the structure of the data and the logic required to extract the information from a single table in a specific database.

The transient model maps combinations of persistent objects to logical artifacts. These models also permit the combining of logical artifacts in a progressive manner. This feature is required to address the complexity and integration of objects from multiple databases. Logical artifacts also document the relationships and interdependencies between persistent and transient objects.

The business artifact groups persistent and transient objects and forms the complete business transaction. During the construction of business artifacts, business analysts and IM designers verify that the combination of logical artifacts and data objects fulfils business requirements and that the target system is provided with sufficient information to discern its meaning and intent.

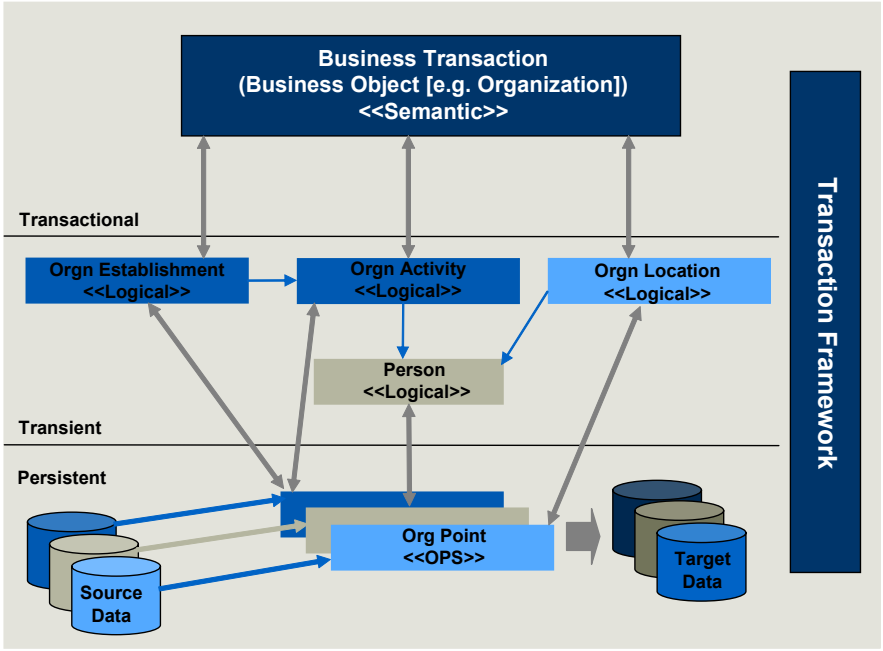

Figure 4: Documenting Interfaces

Managing Meta-Information

Underlying an effective strategy for developing, delivering and supporting data, information, and knowledge management systems is the requirement to manage information about that environment. This is truly Catch-22, since in order to develop and deploy a sustainable IM environment, one must first have an effective IM environment. Information sources for a project may include:

- Multinational standards organisations;
- Legislation;
- Business re-engineering or marketing activities;
- Business, IM and IT requirements;
- IM policy, standards and architecture;
- Other IM projects or organisations;
- IM engineering and development;
- Network management;

- Project management; and
- Vendors.

Until recently, the IM community has managed to limit itself to database and application development. In many instances, it was the same group of individuals that was responsible for both components, thus making the capture and communication of information relatively simple. As movement continues to shift towards a web-based distributed environment, these teams have evolved to include networking, telecommunication, warehousing, database, application development and security specialists. By maintaining existing practices, these groups are becoming progressively more isolated at a time when integration is crucial. The challenge to the IM community is to adopt mechanisms that facilitate the expression and communication of specification, design and implementation data such that other members of a multi-disciplinary team can review and reuse the data in whole or in part. These mechanisms also help to ensure that the knowledge developed during the course of an IM development is transferred to and maintained by the organisation.

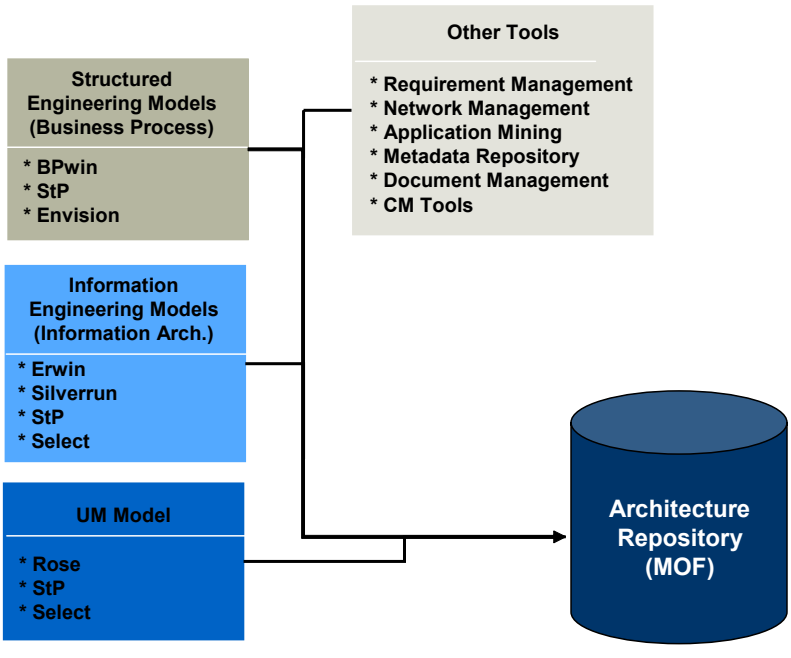

Figure 5: Architecture Repository

The IM community has often expressed concern at the military and aerospace IM processes because of their adherence to process and technical standards and its proportionally high cost to deliver and maintain technical capability. Now that IM systems are becoming ever more complex and approaching mission critical status for a large number of organisations and sectors, with the associated risks and

liabilities, it is time for the IM community to review the experience of the MIL-AERO sector. The goal of this review would be to identify and adopt strategies, practices, procedures and standards that mitigate the risks associated with complex system development and the integration of multidisciplinary teams.

The capture and maintenance of architecture information also underlies the development tools used to support the IM delivery process. Subsequent papers will describe how meta-information can be mined to provide decision support, change management and auditing tools to CIOs, CTOs and IM managers.

Maintaining IM Security, Confidentiality and Privacy

Information security has always been a major concern for government, military and police organisations. However, as a result of an evolution in new legislation (liability) pertaining to the confidentiality and privacy of personal information, it has become the concern of a much broader segment of the IM community, such as healthcare, banking and insurance industries. Using artifacts and a model-based approach provides business analysts with the ability to identify and document security considerations. Each of the identified persistent, logical and business artifacts can be annotated (tagged) with security restrictions and considerations.

Capturing security information during the architecture, specification and design phases establishes the data-set required to generate security documentation in accordance with international standards (e.g. Common Criteria), to provide traceability from requirements to implementation, and to develop a framework for auditing the movement of information within the defined IM environment. Future articles will provide details on how to deliver this capability.

Managing Change

Change is the only consistent feature of the IM/IT sector. Changes in client and user expectations, international and sector standards, and technology are continually imposing changes on the IM environment. The demand for change has also impacted the interfaces between systems and databases since interfaces are not only affected by changes in requirements, but by changes to applications and data structures as well. As IM environments are integrated, small changes to a simple data structure can have a major impact on a wide range of interfaces and effectiveness of OPSs, ODSs, Data Warehouses, Data Marts and information portals. IM groups must possess the ability to track the influence of changes and assess the potential impact of proposed changes prior to their approval. The modeling approach described in this paper captures the dataset required to assess the impact of change to the IM environment. Future articles will provide details on how to deliver this capability.

Summary

The approach to model business artifacts as part of the development life cycle can provide significant benefits to both technical and management personnel. The ASMG solution offers benefits to management by providing efficient access to the data and information that is required to achieve effective business practices and processes.