CHAPTER 1

Information Integration: A Review of Emerging E-Business Technologies

Liping Liu and Bindiganavale S. Vijayaraman
College of Business Administration
The University of Akron
Akron, OH 44325-4801, USA
Email: liu@acm.org; vsb@uakron.edu

1 Introduction

There is no consensus on what constitutes e-business or e-commerce[1]. As the term implies, e-business is the conduct of business on the Internet. IBM, who was the first to use the term in October 1997, defines *e-business* as the transformation, through the use of Internet technologies, of business activities involving information collection, organization, and exchange. In essence, e-business aims at improving business processes by replacing traditionally paper-based processes with Internet-based applications. As a special case, *e-commerce* refers to the buying and selling products and services over the Internet[2]. In practice, e-business and e-commerce are often used interchangeably, and other related terms such as e-tailing, e-procurement, and e-marketing are also used in special occasions.

Conducting business electronically is not new; for three decades corporations have been using EDI (electronic data interchange) over private networks to exchange business documents such as Invoices, Purchase Orders, Bills of Lading, and Acknowledgments. What makes e-business a phenomenal trend today is, however, the advent of modern e-business technologies, which have been developed to take advantages of the Internet's low-cost, widespread accessibility, and its potential for integrating information across the boundaries of devices, organizations, and physical locations.

HANDBOOK OF INFORMATION TECHNOLOGY IN ORGANIZATIONS AND ELECTRONIC MARKETS

In the last ten years, e-business technologies have undergone four generations of transitions from static web pages, to interactive media, and to dynamic commerce[3,4]. Early adopters of e-business primarily used static web sites for broadcasting business information. The second generation web sites began to support information requests and customer profiles, moving beyond the one directional information flow of the initial web sites. The third generation e-business technologies transformed relatively primitive web sites into the world of e-commerce, transferring the Internet into a new sales channel[5,6]. By enabling continuous interaction, they presented businesses with a new way to exchange information and to reach customers that is independent of the physical presence of the traditional "brick and mortar" storefronts. For example, they allowed customers to browse products, fill shopping carts, and make purchases from anywhere and at any time. They also allowed distributors to check for inventory availability and automate reorder generation, and vendors to take advantage of on-line invoicing and faster electronic payment cycles.

The focus of e-business today is on information integration with the aim at transforming business processes across organizations and throughout the entire value chain, including customers, suppliers, distributors, and even competitors, and extending the Internet to the physical world. The challenge is to automate last-mile processes at the edge of the Internet, including capturing data into computer systems and moving data from one system to another. These processes are still largely manual, becoming the bottleneck of corporate workflows and the inhibitor to newer e-business initiatives.

The inefficiency of one manual process may easily offset the gain in productivity of many e-transformed processes. This is evidenced by the failure of EDI initiatives. Corporations once believed that EDI was the solution to their business problems but found their expectations largely unfulfilled; EDI implementations only automated the flow of information between trading partners while leaving manual and/or inefficient processing activities remained for both senders and receivers. For example, many companies simply sent their incoming EDI transactions to the printer while traditional manual processes remained unchanged.

The e-business era has not solved this last-mile problem and companies are still facing the same challenge as before. The majority of businesses today are still in an extremely disintegrated state. Most businesses still rely heavily on information delivered on paper, whether it is machine print, handprint, handwriting, tick boxes or bar codes[7]. Almost all but the smallest organizations have multiple systems that do not communicate with each other[8]. About 75% of companies are still in the process of integrating their front office with back-office systems[9] and 71–91% of executives could not access information from different systems in a single view[10]. As each Internet ready e-business application hits the market, companies are faced with new, costly integration projects to connect these applications to existing business applications and legacy systems. A recent survey in US found that 74% of executives believed that information integration was their top priority[11], and companies spent up to 25 percent of their technology budgets trying to integrate (and reintegrate) their business systems. A similar survey in the UK reveals that 69% of organizations saw a lack of integration impacting productivity, while 57% felt it would decrease revenues if not addressed. Furthermore, 43% of the respondents are currently working on projects driven by the need for better integration and more than 50% of companies had integration requirements in half or more of new projects[12].

Responding to the challenge, newer e-business technologies, including radio frequency identification (RFID) and e-services are emerging to take up the core of the fourth generation e-business revolution. RFID is a technology that captures product data and track inventory throughout its supply chain without human interaction, enabling the Internet to reach the physical world[13]. E-services refer to a group of technologies, including web services and grid services that create an open standard and emerging trend for application and information integration[14]. In this chapter, we review these two emerging technologies from the perspective of supporting cross-boundary information integration. We introduce the basic concepts of these technologies and show how they may be used to meet the challenge to the fourth generation e-business era. We provide critical assessments of their current status and identify the challenges to their adoption.

2 Information Integration

E-Business today is challenged to create the real-time enterprise that is responsive and proactive in the face of current business events[15], and develop new ways to respond the new types of market demand — online orders that provide immediate payment but also bring expectations that the product will be delivered just in time[4]. This entails a high-level of information sharing and functional integration across the entire supply chain. For example, it requires that suppliers are able to monitor the inventory level of their products in stores and shops worldwide, dispatch new shipments when necessary, track shipping and routing status in the distribution channel, and receive payments upon delivery.

There are two particular challenges to the fourth generation e-business era. The first challenge is the disconnection between logical and physical worlds, leading to the asynchrony of data and underlying physical objects. A database is simply a model of the real world and information maps its entities, relationships, and dynamics. Unfortunately, creating and updating such a map has never been automatic and manual processes are the state-of-art. In the logical world, information may travel at the speed of light but it jams or stops at the point of the disconnection, waiting for human interactions. In material handling and manufacturing applications, for example, these processes include shipping, receiving, bookkeeping, product sorting and counting, order fulfillments, material assembly, and monitoring the status of work-in-process, machine utilization, worker attendance, and other measures of factory operations and performance. Manual processes are also prevalent in other business settings such as retail sales and inventory control, warehousing and distribution center operations, mail and parcel handling, patient identification in hospitals, check processing in banks, and security systems.

In a typical organization, more than 90% of information is still on paper[16]. With the rise of e-business, the need for automatic processing of this information is greater than ever. Research in this area is ongoing under the umbrella of automatic identification and data capture (AIDC), which refers to data collection by means other than manual notation or keyboard input. AIDC technologies can be broken down into roughly six categories: Biometric, Electromagnetic, Magnetic, Optical, Smart Cards, and Touch. They include fingerprint identification and verification, hand geometry, iris scan, signature identification and verification, voice recognition, facial thermography, magnetic stripe, magnetic ink character recognition, bar code data capture, optical character recognition, machine vision, smart cards, touch screens, button memory, and pen-based computing. All these technologies require no keyboarding but human involvement or presence to certain extent in the data capture process. Thus, they are not suitable for collecting a large amount of data for material handling and manufacturing applications.

The second challenge is the lack of technical interoperability between systems. A typical personal example can illustrate the case. For instance, today I decide to attend a conference and so go to their web site to register. At this point, I would like my calendar to have an entry, and my Outlook to download the contact information. I would like my travel agent to access the same information to make flight and hotel reservations. I would like my GPS to download the addresses and compute the best routes to get around. I would also like my Quicken to register associated expenses. I would like to do all this with one click. Unfortunately, currently I cannot. What I in fact have to do is to manually enter the time into my calendar, laboriously cut and paste contact details into my address book, retype the same information into the travel agent's reservation system, and transfer invoices into my Quicken.

The revelation about student visa approvals for 9/11 hijackers is another example. Like many government agencies, INS has dozens of separate systems that do not communicate with each other. INS admits that the problem in this case was that the contractor that notifies schools of student visa approvals gets its information from INS on paper, which must then be manually entered. INS approved the visas well before September 11th, but it took years for the paperwork to catch up.

Years of piecemeal deployment have created a phenomenon common in all organizations where there are numerous non-compatible systems[8]. Each was developed or purchased to provide a specific function and many have been developed on different platforms using different languages, creating numerous barriers between the systems.

Interoperability is often lacking at two levels: applications and information sources, and manifested in symptoms like application inaccessibility, workflow discontinuity, and information heterogeneity. For example, it is often the case that the output of one application has to be reprocessed to be input into the other or combined with the output of the other. In the literature, data heterogeneity has long been recognized and it includes *structural heterogeneity* — different systems store data in different structures, or *semantic heterogeneity* — different systems do not understand the meaning of each other's data[17]. Manual processing is often the only viable approach to handling data heterogeneity. Companies often hire dedicated staff to do nothing more than translate information from one format to another, or enter data generated by one system into another.

Breaking down application barriers has been the central theme of research and development under the umbrella of EAI (enterprise application integration). EAI solutions typically include the following five components[18]: Message-oriented middleware, application servers, adapters, adapter development frameworks, and workflow and process management tools. Middleware serves as the glue to connect two otherwise separate applications. Increasingly it is used in the development of applications which communicate across platforms as they traverse the Internet. Different approaches have been proposed as a platform for middleware solutions, including CORBA (the Common Object Request Broker Architecture), DCOM (Distributed Component Object Model), EJB (Enterprise JavaBeans), RMI (Remote Method Invocation), and DSOM (Distributed System Object Model). Consequently, middleware is largely proprietary in nature and the languages used are generally not compatible. This makes it difficult to connect middleware islands inside a single organization. Crossing the boundaries between the islands is the key challenge to businesses as they try to incorporate e-business applications[8].

The EAI approach has had some success and may be the best solution for specific, mission critical and high-performance systems. However, EAI has some serious limitations and drawbacks when it is applied to enterprise-wide integration and inter-enterprise integration. EAI solutions depend on many point-to-point links that connect specific elements and functions of each system. Every one of these links has to be created and maintained and every integration project is beyond connecting one organization to a few partners. Instead, each organization has multiple information nodes sending and receiving transactions to an array of internal and external data sources and applications. For example, a customer may demand integration with its ERP system so that it can search for new products, view production schedule, track order status, and update sales ledger information. This would require at least four separate interfaces and probably four different platforms. Imagine 100 nodes join in a network, and each uses a unique piece of software on a unique computing platform. Using middleware technologies, each node would have to incur significant expense writing custom bridges to "hardwire" to the other 99 nodes. Once implemented, each of these links is inflexible and needs to be updated to respond to any changes to the integrated systems or the processes. If one node changes its internal system, all other nodes would have to respond. If a new node needs to join the network, all have to incur an enormous cost of entry to maintain the status of integration. In fact, complexity and bespoke nature demands, for each integration project, hundreds of hours of EAI consultant's time in addition to the cost of the middleware solution itself. Gartner estimates that 40% of a project's costs and time are spent on integration issues. It further predicts a $500,000 entry point and 12 months to see any value even from discrete internal integration projects[8].

Research on information integration has focused in three areas: schema integration, semantic mediation, and ontology merging[19]. The general problem is to define a global schema over a set of local schemas[20]. Schema integration[21] as well as its implementations, e.g. CLIO[22], focuses on resolving relation and attribute conflicts between to-be-integrated schemas. Semantic mediation introduces a mediator as transparency layer that hides the distribution of information providers from applications, and focuses on integrating new data sources into the mediator[19]. Ontology merging was initiated in artificial intelligence[23] and is now studied under the umbrella of *semantic web*, which is an extension of the current web in which information is given well-defined meaning[24], and a set of portable standards to represent ontologies, enabling Internet search engines to retrieve information in a much more intelligent way. Research in this area focuses on representing and finding term-matching relationships for resolving semantic heterogeneity[25], and developing languages to represent ontologies and processing mechanisms[26].

Research has generated a few commercial products like IBM's DataJoiner[27] and Attunity's Connect (attunity.com) and tens of research prototypes, including TSIMMIS[28], information manifolds[29], SIMS[30], MIX[31], DIKE[32], and ODB-Tools[33]. Unfortunately, all these technologies have focused on integrating a previously known fixed set of data sources and developing an integrated schema for these sources. They lack the flexibility to handle dynamically changing sets of data providers and they do not aim at semantically integrating the data of different providers[34]. Adding a new data source typically requires human processing, and often leads to loss of information that does not fit in the global schema. These technologies may be used locally to transform the data of a provider into a common domain schema compliant representation.

3 Radio Frequency Identification

Radio Frequency Identification (RFID) overcomes the limitation of traditional AIDC solutions, and is a technology that captures data or track inventory with minimum human intervention. Unlike other AIDC solutions, RFID is the non-contact, non-line-of-sight nature of the technology for data retrieval (www.AimGlobal.org). Data that is stored on RFID tags can be read through a variety of substances such as snow, fog, ice, paint, crusted grime, and other visually and environmentally challenging conditions, where barcodes or other optically read technologies would be useless. This has many significant implications to e-business. For example, using a barcode system, if a pallet arrives at a dock door, it will not be registered until an operator walks over and scans it. The number of operators available to perform the scanning limits the data rate that the barcode system can generate. In RFID, however, the acceptance of the pallet can happen automatically by using a reader to constantly monitor its space for incoming shipments. In this way, RFID systems have more of a sensor-network or monitoring-system flavor than do bar-code systems[35].

A basic RFID system consists of three main components[36]- an RFID tag, a scanner, and a computer. Antennas emit radio waves that send or receive data to and from the RFID tags. A scanner (receiver) is then used to capture the data and send the information to a computer system for interpretation, storage, and action.

There are two main types of RFID tags being used today; active and passive with passive being the more commonly used and the least expensive. Each type has its own advantages and disadvantages. *Active* tags contain built-in batteries and an internal transmitter, and are capable of transmitting up to 64 KB data. They are continuously powered and able to send and receive data at all times. Active tags are usually much larger in size and more expensive than the passive tags. Active tags generally have a much shorter lifespan than passive tags. The main reason to justify the use of an active tag is its range of communication; an active tag's read range can extend well beyond 300 feet.

Passive tags do not contain an internal power source or transmitter and must be activated through an RF signal from a reader in order to transmit information. They can be extremely small and have a long operating life. The cost of a passive tag can be as low as only a few cents per tag, making them much more affordable for companies wanting to track low-cost product. Generally, passive tags have a much shorter read range than active tags do. Most passive tags must be within a few inches to a few feet of a reader in order to transmit data.

RFID tags can store more information than any other devices currently in use for identifying inventory items. All RFID tags contain microchips that have either "read-only" or "read-write" memory. A chip with "read-write" memory allows for the original data on the tag to be read, modified, or overwritten. On the other hand, a chip with "read-only" memory comes stored with pre-existing data from the manufacturer of the tag that can be read but not modified. Many passive tags use "read only" memory in order to keep their overall cost down.

In a sense, RFID technology is not new. One of the first recorded uses of RFID technology dates back to World War II, when RFID transponders were used as a means to distinguish friendly aircrafts from enemies. The real innovation in RFID, however, is probably not in the technology itself, but in its application to real world situations[37].

Today, the primary application of RFID is to locate, identify, and track objects with the goal to reduce costs, wastes, production cycle time, and inventory[38]. RFID tags allows for much-improved tracking and inventorying of products, especially where multiple products and product-lines are involved. They can help provide operational efficiencies and improve stock level transparency in short shelf-life product distribution in supply chains[39]. They can also potentially be tied into security systems, helping to prevent theft of product.

The benefits of RFID include improved data accuracy, enhanced asset visibility, reduced information latency, reduced out-of-stocks, reduced inventory, reduced shrinkage, improved forecasting, enhanced product pedigree, reduced costs of product recalls, improved accuracy for return processing, reduced product counterfeiting, theft deterrence, etc. A study by Accenture[40] indicates that the retailers implementing RFID technology will reap "significant improvements" in both supply chain operations and inventory accuracy, including increasing revenue up to 1 percent, cutting working capital by 2 to 8 percent, and reducing fixed assets by 1 to 5 percent. The most significant is their claim that retailers can achieve up to a 65 percent reduction in in-store labor expenses when it comes to received goods and a 25 percent drop in stocking and cycle counting, and would totally eliminate the need for physical counting. RFID would also reduce product loss by nearly 1 percent in sales.

It is not surprising that, with such potential benefits, large retailers such as Wal-Mart, Target, and Tesco are making it mandatory that their suppliers use RFID tags. This is quite a big demand, but most of the suppliers are left with no other choice than to upgrade their current supply chain systems in order to integrate RFID technology. It is estimated by Yankee Group that over the next three years, manufacturers will spend $2 billion on RFID tags and another $1 to $3 billion on related infrastructure to implement RFID tags[41]. Indeed, it is this initial "start up" expense that challenges many manufacturing companies. They will be forced to make a decision if the up-front cost is worth the benefit or if they simply cannot afford to make the move. Some may choose to wait until deployment cost comes down further as tags and equipment cost becomes more affordable. A study by AT Kearney[42] concluded that, "retailers can expect extensive inventory and labor cost savings from the adoption of RFID technology, but some consumer product manufacturers will face higher costs and delayed benefits from adopting the technology." In other words, RFID technology is expected to result in "long term" cost savings by retailers and eventually manufacturers but, due to initial startup costs, it may take a while for those savings to be passed on to everyday consumers. According to Alvarez[43], by 2008-2009, enterprises will tag more than 70% of their assets and generate operating cost reductions of 1-3% through reduction of lost assets, improved tracking of asset maintenance, and protection of assets from theft, fraud, or injury.

In addition to retailers, other industries and government agencies are also looking forward to using RFID. For examples, Sullivon[44] noted an application by Georgetown University Hospital's Blood Bank and a study of RFID wristbands in comparison to the use of barcodes. MeadWestvaco Intelligent Systems is evaluating the inventory tracking solution of RF tags in order to assist the U.S. Food and Drug Administration with its multi-faceted approach, which has been launched to reduce and/or eliminate counterfeit drugs. Garskof[45] noted a few applications: 1) A Las Vegas casino is expected to begin implementing RFID-enabled gambling chips to stop counterfeiters and spot high rollers; 2) MasterCard will debut a contact-free RFID payment chip called PayPass that will be available in credit cards and Nokia cell phones; and 3) Helena Regional Airport (Montana) is installing new RFID-enabled ID readers, complete with onboard fingerprint sensors, to keep tabs on its employees.

RFID has the potential to automate supply chain processes and allow companies to determine the exact location of their products in real time. Highjump (3M) was one of the first vendors to develop RFID supply chain software. Other big name vendors including Sun, SAP, Oracle, and IBM are rewriting their existing enterprise applications to integrate RFID data[46]. In 2004, Microsoft announced its first RFID supply chain management pilot project. Microsoft plans to market its RFID software to mid-sized companies, which total 39 million worldwide.

Even though everyone in the supply chain sees some benefit from RFID, the manufacturers will end up paying for the cost of tags. It is estimated that it could cost a combined $2 billion dollars for the top 125 manufacturers to implement RFID tags[42]. Indeed, the main criticism of RFID technology is that it is too expensive and it is unlikely that the investment will pay off implementing at the item level tagging[39]. The majority of manufacturing companies do not expect a positive ROI although retailers were very optimistic[47]. Despite some early trials of RFID tags at item level by Gillette and Tesco, the ROI for item level tagging remains to be a concern for many companies because the costs of tags and associated infrastructure are still greater than anticipated benefits[48].

Although cost is a concern, lack of worldwide tag standards posts a bigger challenge to RFID adoption. The standards for RFID technology are still evolving. EPC Global is currently working with its member organizations to establish standard for tags. The acceptance of these standards depends on big retailers such as Wal-Mart, Target, Tesco, and other companies requiring suppliers and manufacturers to use tags that use a common standard.

Another challenge is the management of voluminous data the RFID system will create and the complexity of the system integration[49]. Implementation of RFID will generate a massive amount of data that needs to be stored, processed, and used in real time. RFID systems will therefore need to be integrated with existing data warehouses and other e-business systems. This will only complicate the difficult task of information integration. Existing systems may not be able to cope with the level of data or the complexity RFID technology may bring. The continuous flow of product data as an item moves through a supply chain demands that e-business systems be scalable to support the data flow, capable of distributed data gathering, and able to route and integrate information to internal applications and external business partners.

The debate that revolves around privacy and security currently dominates the press. Unfortunately, empirical data and clear analysis are scarce. The perception among consumers is that RFID will violate their privacy. RF tags can take suppliers and distributors beyond where the product was manufactured, its shipping history, and the date and time it went through the store's checkout. Electronic readers could take the information from the tag and feed it into a computer database, which can then perform any operations, for example, to determine when last time you purchased toilet paper was or which books you bought. In addressing this issue, a bill was presented to the California State Senate in 2004 concerning the use of RFID. It outlines three requirements for a business to use an RFID system that can track products or people: Tell customers it's using RFID, get consent to track or collect any information, and detach or destroy RFID tags on products before the customer leaves the store[50].

Some authors have a different take on the privacy issue. For example, Cline[51] believes that the privacy scare is simply overblown. "No company or government agency will be secretly scanning your house to find out what products you have purchased, because there is no feasible way to do so." He also admits that, if RFID chip-makers do not soon allay these fears, the escalating public emotion about this issue may effectively ban the most valuable implementations of this remarkable technology. Bacheldor[50] confirms this assessment. IBM recently conducted a study on the Cost of Privacy[51]. Based on 44 U.S.-based multinational companies, they found that setting up a privacy program office is the costliest part of privacy efforts at major corporations. They suggest that privacy is not something to be taken lightly and the privacy program will not come without expense or effort but it is an aspect of the future that must be evaluated and considered very carefully.

4 E-Service Technologies

E-service technologies are developed in response to the interoperability problem. Among them, web services and the Grid are the latest trend and the hottest sector in the information technology industry. A few years ago, *application services* were proposed to provide packaged applications over the Internet as an alternative to hosting them in house[52]. Sharing a similar vision, *web services* aim at providing modular software components to achieve collaborative computing via Internet protocols[53]. The *Grid* seeks even larger-scale sharing of computational and informational resources across platforms, devices, and institutional boundaries[54]. Although the grid was positioned to be independent of the Internet, it is now rapidly converging with web services to form a single set of standards[55] and has conveniently served as a marketplace for exchanging web services[56], as an environment to organize web services[57], or as a framework for architecting next generation information systems[58].

4.1 *Web services*

The idea of web services is not new. For many years, application services provision (ASP) has been a business model of supplying and consuming computational services over computer networks. In this model, a provider assumes responsibility of buying, hosting, and maintaining a software package on its own facilities, publish its user interfaces over computer networks, and provides its clients with a shared access to the published user interfaces. The client organizations, on the other hand, subscribe and receive the application services through the Internet or a dedicated network connection as an alternative to hosting the same application in house.

The ASP model essentially allows organizations to hand over the responsibility of systems deployment or its execution to an outside vendor while still satisfying self-information needs. It reduces the complexity associated with the traditional make-or-buy model while allowing an effective control of the deployment costs and risks[59]. It can amortize expenditures over the entire client base, enabling it to improve quality of services, security, and risk reduction measures that individual client may find cost-prohibitive.

Despite the benefits, the growth of the ASP industry has been slow. Some researchers have attributed the problem to limited user satisfaction[60]. However, some community has tended to blame users for not being ready for change[61], and has attributed the problem to limited end-user acceptance. Liu and Ma[62] empirically examined this problem based on the technology acceptance model[63].

The ASP model enables multiple organization to access a central data repository and therefore, provides a relief to the interoperability problem. For example, in the medical industry, physicians who subscribe to the same ASP may share patient data. However, the ASP model does not meet the need for information integration especially for data exchange between business partners. Instead, like ERP or EAI solutions, it simply replaces the old silos with large data islands. In the last few years, most ASPs have struggled financially and never gained widespread acceptance. No vendor is big enough to provide services to all partners in the supply chain. In the medical industry, for example, no software vendor today is able to cover every aspect of HIPAA compliance[64].

Embracing the idea of the ASP, Web services improve upon the EAI approach to systems integration. The interoperability problem with middleware has forced companies including Microsoft and IBM to seek a new open standard, i.e., web services, for distributed computing across applications, platforms, and devices. Unlike middleware solutions, web services are loosely coupled software components that exchange XML-based data[14]. Web services have the following distinct characteristics[65]. First, each web service represents a business function or business service, whose interface is exposed to the Internet and accessible by another program remotely[66]. It encapsulates a task; when an application passes a message (e.g., data or instructions) to it, the service processes the input and, if required, generates an output message back to the application. Second, all messages are written in XML-based text, which follows SOAP (the Simple Object Access Protocol) coding and formatting specifications, instead of cryptic binary strings. This enables web services to communicate with other applications that may be developed in different programming languages and reside on different platforms[67]. Third, web services are self-describing; each is accompanied by a description, written in the WSDL (Web Services Description Language), regarding what it does and how it can be used. Fourth, web services are discoverable; service consumers can search for and locate desired web services through UDDI (Universal Description, Discovery and Integration) registries.

A simple marginal costs perspective would suggest that each organization can expose some or all functionalities of its internal legacy (open or proprietary) systems as web services. These services can be as simple as scheduling appointments, validating credit cards, browsing product catalogs, or submitting invoices. They can also be as complex as the functions carried out by an entire supply chain, customer relation management system, or ERP applications. These services hide their internal complexities such as data types and business logics from their users, but expose their programming interfaces using the WSDL and their locations using UDDI protocols. Since every service complies with one set of web services standards, there is no need for writing custom bridges in order to accommodate different computing platforms. Instead, organizations can exchange data by directly invoking data exchange services.

Web services might provide businesses with an idea platform for information integration. Essentially, web services allow code to speak to code without human intervention. This will enable organizations to create an integration platform capable of crossing internal boundaries between intra-enterprise applications and connecting middleware islands with middleware-to-middleware integration. The business press calls such an integration platform SOA (service-oriented architecture)[68] or software bus[8]. Technically, SOA refers to an application architecture within which all functions are defined as independent services with well-defined invokable interfaces that can be called in defined sequences to form business processes. SOA aims at simplifying the integration of disparate systems. The traditional EAI approach quickly leads to hundreds of custom bridges for every new application or process change, whereas SOA requires just one integration point for each. New systems and platforms can be plugged in to a software bus to be immediately integrated with all the other systems on the same bus. SOA can be also used to integrate middleware islands and thus leverage existing EAI investments[8]. For example, an organization that may have selected a CORBA-based integration in one area and DCOM for another can integrate the two without discarding either.

Organizations can extend the SOA beyond the boundaries of the enterprise to link to suppliers and customers, and offer collaboration with outside systems via the same architecture. They can leverage SOA to establish a single connection to each vendor or customer organization rather than building point to point connections. If one organization were to change how a certain function is processed internally, as long as the programming interface does not change, the rest of the world can remain still. Thus, the cost of entry and exit will be greatly reduced.

To realize the benefits of SOA, of course, five infrastructure management issues must be resolved[68]: 1) security management that handles validation or authorization of the request, encryption, and decryption; 2) deployment management that allow services to be moved around the network for performance, availability, or other reasons; 3) logging management for auditing, metering, and tracking; 4) dynamic rerouting for fail over or load balancing; and 5) maintenance management.

Web services have the potential for instant connectivity with a universe of services over the Internet. However, some authors have reservation, believing it is a mirage e.g.,[8]. They caution that the security issues will make it very difficult for any organization to really trust its core processes to unknown third parties that they found on the Internet. According to a recent survey by ITToolBox.com, 54% of participants indicated that they are currently using web services for integration while 42% cited security concerns as the biggest challenge to widespread acceptance of web services. Another challenge is with complexity. At present, web services are capable of device-to-device communication but not much so of workflow processing. Although they have the potential to combine individual services into more complex, orchestrated services that will realize sophisticated business process and workflow automation, such composition and orchestration is still on the drawing board [69–71].

4.2 *Grid services*

Besides security and complexity issues, the logistics of distributing web services is also a concern[72]. Regardless how easy it is to search for and locate a web service through a UDDI registry, after all, a consumer may have to contact the provider and negotiate a service contract. If there are thousands of web services to be subscribed, it is practically impossible for anyone to contact these many providers individually and renew contracts periodically. At the same time, service providers are facing another dilemma. Since a web service is usually a small component, it may not be cost effective to spend money to market the service in a distinctive way. However, a lack of marketing effort will reduce the consumer awareness about the service, which in turn, will reduce the number of subscribers. Consequently, most providers cannot even afford the expense of maintaining their services and may have to exit, leading to a shrunk service market that diminishes the viability of web services as a business model.

Then how do we resolve this logistic issue? One approach is to have a few large providers responsible for assembling and delivering suites of web services or packaged applications to consumers. The approach might work, but it may be insufficient because these providers, like other consumers, still need to find and negotiate contracts for required services. What seems more desirable is to have a global market for efficient exchange of web services between service consumers and providers. This solution leads to the notion of a service grid, which is a distributed computing infrastructure consisting of large and diverse sets of distributed resources and services. The concept revolves around the idea of service creation and delivery through coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations[73]. Peer-to-peer computing (P2P) was an early implementation of a service grid[74]; it aggregates the unused computing power of individual personal computers into a computer power grid to create a virtual supercomputer. With the advent of web services, now grid and web services are rapidly converging to form a single set of standards and technologies as manifested in the Open Grid Services Architecture[55], which is so promising that many people believe it is one of ten key emerging technologies that will change the world[75].

Enabling collaborative computing, grid services have the potential to serve as an effective market mechanism for distributing web services. A service grid may act as an intermediary between service providers and service consumers and break a typical many-to-many business (between providers and consumers) into simple one-to-many relationships. It buys web services from the providers and then sells the services to consumers. Then consuming web services becomes as easy as watching TV programs from a cable network or obtaining electricity from a power grid; requesting and delivering web services becomes as easy as plugging an appliance into the grid. In the meantime, those small service providers do not have to incur prohibitive expenses to advertise and run its businesses. They can focus on their core business — developing and upgrading web services, and then plug the services into the grid to sell.

Taking advantage of grids, new approaches have been proposed for information integration[34]. These include semantic mediation systems tailored to peer-to-peer architectures and data grids for the management of massive amounts of distributed data. P2P systems are highly dynamic, allowing for ad-hoc addition and removal of data sources. Instead of a fixed, global schema, P2P architectures couple the peers directly and map their local schemas via domain relations and coordination formulas[76] or peer mappings[77]. An application directly sends a query to a single peer using its local schema, which may access other peers to answer the query. Using UDDI, discovery services may be implemented to achieve location transparency and can be searched by mapping logical domains and predicates onto actual data sources. In a medical application, Liu and his colleagues proposed the notion of a data grid for managing patient records[56,72], which spreads over loosely coupled data nodes and works together with a separate service grid that utilizes the data resources. Research has mostly focused on replica management and consistency issues[34]. In the case of medical records, however, applications do not need to send complex queries over various data sources but use the grid to find one data source that has the file they need. There is no need for data integration or merging. The main issue here is probably the lack of a privacy control protocol that specifies how medical records at each node might be accessed and secured[56].

5 Conclusion

E-business technologies have undergone four generations of transitions. The Internet today offers unlimited potentials for transforming business processes, from making online orders to managing inter- and intra-organizational interactions and building customer and vendor relations. E-transformation has accelerated most of business processes and information flows. However, the bottleneck is at both ends. In typical organizations, more than 90% of information is still on paper and the majority of information systems are still in an extremely disintegrated state.

The challenge to the fourth-generation e-business era is to bring e-transformation throughout the entire value chain. It entails transforming last-mile processes, including capturing data into computer systems and moving data from one system to another. These processes are still largely paper-based and become the inhibitor to new e-business initiatives. In particular, there are two obstacles to overcome. The first is the disconnection between logical and physical worlds, leading to the asynchrony of data and underlying physical objects. The second is the lack of technical interoperability between distributed systems.

Research on information integration has produced tens of products, technologies, and prototypes to meet the challenges. AIDC technologies aim at collecting data without human interaction. A lot of progress has been made in the areas of optical recognition and machine vision. However, most technologies are still not suitable for material handling in e-business applications. RFID is one of few emerging AIDC technologies on the e-business scene, which has the potential to extend the Internet to the physical world but it may take time to realize its promise. RFID, the technology for managing supply chain, is continuously evolving. Many benefits such as supply chain visibility will materialize only when RFID technologies work with complementary technologies like Bluetooth and GPS[13]. RFID standards and products such as tags, scanners, and software are evolving rapidly. Companies that are trying to implement RFID technology will have to make short term strategic plans based on the current technology and standards but will have to develop long term strategic plans based on appropriate information resources and projections. Companies should determine how to integrate RFID technology with their existing e-business systems that connect to their partners. Real time tracking of all the items in the supply chain and making effective use of the data generated by RFID technology could be a big challenge to many companies.

System interoperability issue arises in the form of application inaccessibility, workflow discontinuity, or information heterogeneity. Human interaction is often the only viable approach to resolve the issue. Using proprietary middleware, EAI offers some success in making point-to-point connections between fewer systems. However, it becomes too expansive and complex when it is applied to enterprise-wide integration and inter-enterprise integration. Emerging web service technology offers a great promise in advancing the mission of EAI. Instead of incompatible middleware, it offers open standards and protocols that enable XML-based communication across platforms, devices, and languages. Instead of point-to-point connections, it allows service oriented architectures that reduce complex many-to-many connections in typical EAI projects; within an organization, integration is as simple as plugging a system into a software bus, and between organizations, integration is as simple as connecting two software buses. Instead of integration with a specific set of systems, emerging grid services promise an open architecture for collaborative computing over the Internet; integration becomes as simple as plugging an appliance into the power grid. Someday, a business may find itself in a position not dealing with one specific vendor for performing a certain process, say validating credit cards or shipping orders. Rather it will deal with a grid that offers a line of similar services that the business can choose, negotiate with, and use on the fly.

What follows the fourth generation e-business? We have seen where the train is going from its base station: the Internet eliminates the constraints of time and space and allows people to conduct businesses anywhere and anytime. After information integration, distance collaboration will be the next theme of e-business initiatives. Through e-transformation, jobs that are usually performed on corporate campuses can now be performed over the Internet. Thus, fifth generation e-business technology will redefine the landscape of employment and enable virtual corporations. It will open up more opportunities for job outsourcing and improve resource optimization and labor division for the mankind.

References

1. Riggins, F.J. and Rhee, H.-S.S., Toward a unified view of electronic commerce. *Communications of the ACM*, 41, 1998, pp. 88-95.
2. Kosiur, D., *Understanding Electronic Commerce*. Microsoft Press, RedMond, WA, 1997.
3. Mackey, C., The Evolution of E-Business. *Darwin*, 2003 May 1.
4. Moncla, B., *The Convergence of E-Business and Business Intelligence: The I-Business Hurricane*. 2005 [cited 2005 April 8]; URL: http://datawarehouse.ittoolbox.com/browse.asp?c=DWPeerPublishing&r=%2Fpub%2FAS070601.pdf.

5. Amor, D., *The E-business Revolution*. Prentice Hall, Upper Saddle River, NJ, 1999.
6. Kalakota, R. and Whinston, A.B., *Frontiers of Electronic Commerce*. Addison-Wesley, Reading, MA, 1996.
7. Brierton, R., *Automatic data capture - why handprint is a challenge*. 2001 [cited 2005 April 15]; URL: http://www.findarticles.com/p/articles/mi_qa3947/is_200108/ai_n8963346.
8. Baker, S., *The three steps to Web service integration*. 2002 [cited 2004 April 15]; URL: http://searchwebservices.techtarget.com/originalContent/0,289142,sid26_gci841585,00.html.
9. Maselli, J., Lack of Integration Hurts CRM Efforts: Study. *InformationWeek*, 2001 December 19.
10. Anderson, H., *2004 Systems Integration Survey*. 2004 [cited 2005 June 7]; URL: http://www.healthdatamanagement.com/html/SISurvey061004.cfm.
11. Anderson, H., Tackling the Challenge of Systems Integration. *Health Data Management*, 2005 April 8.
12. M2Communications. *Data Integration Still Important to UK It - Report*. 2005 August 17 [cited 2005 September 1]; URL: http://projectmanagement.ittoolbox.com/news/dispnews.asp?i=132666.
13. Walker, J., *et al.*, *What You Need to Know About RFID In 2004*. 2003 [cited 2005 April 15]; URL: http://www.forrester.com/ER/Research/Brief/0,1317,33298,00.html.
14. Stal, M., Web services: Beyond component-based computing. *Communications of the ACM*, 45(10), 2002, pp. 71-76.
15. Abbas, A., Productivity paradox and information technology, in *Grid Computing: A Practical Guide to Technology and Applications*, A. Abbas (Ed.), Charles River Media, Inc., Hingham, MA, 2004, pp. 19-29.
16. Hemphill, B., *File Act or Toss*. 2005 [cited 2005 June 7]; URL: http://www.appearfirst.com/business/business_37.html.
17. Kim, W. and Seo, J., Classifying schematic and data heterogeneity in multidatabase systems. *IEEE Computer*, 24(12), 1991, pp. 12-18.
18. Ruh, W.A., Maginnis, F.X., and Brown, W.J., *Enterprise Application Integration: A Wiley Tech Brief*. Wiley, New York, 2000.
19. Gupta, A., Ascher, B.L., and Martone, M.E., Registering Scientific Information Sources for Semantic Mediation. In the 21st International Conference on Conceptual Modeling. 2002. Tampa, FL.
20. Ramesh, V. and Ram, S., Integrity Constraint Integration in Heterogeneous Databases: An Enhanced Methodology for Schema Integration. *Information Systems Research*, 22(8), 1997, pp. 423-446.
21. Batini, C., Lenzerini, M., and Navathe, S.B., A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys*, 18(4), 1986, pp. 324-364.
22. Miller, R.J., *et al.*, The Clio Project: Managing Heterogeneity. *ACM SIGMOD Record*, 30(1), 2001, pp. 78-83.

23. Uschold, M. and Gruniger, M., Ontologies: Principles, Methods, and Applications. *Knowledge Engineering Review*, 11(2), 1996, pp. 93-155.
24. Berners-Lee, T., Hendler, J., and Lassila, O., The Semantic Web. *Scientific American*, 284(5), 2001, pp. 34-43.
25. Wache, H., *et al.*, Ontology-based Integration of Information — A Survey of Existing Approaches. In IJCAI-01 Workshop: Ontologies and Information Sharing. 2001. Seattle, WA.
26. Hendler, J. and McGuinness, D., The DARPA Agent Markup Language. *IEEE Intelligent Systems*, 15(6), 2000, pp. 67-73.
27. IBM. *DataJoiner Implementation and Usage Guide*. 1995 [cited 2005 June 7]; URL: http://www.redbooks.ibm.com/redbooks/pdfs/sg242566.pdf.
28. Garcia-Molina, H., *et al.*, The TSIMMIS Approach to Mediation: Data Models and Languages. *Next Generation Information Technologies and Systems*. 1995.
29. Levy, A.Y., Rajaraman, A., and Ordille, J.J. Querying Heterogeneous Information Sources Using Source Descriptions. In 22nd Conference on Very Large Databases. 1996. Mumbai (Bombay), India.
30. Knoblock, C.A., *et al.*, Modeling Web Sources for Information Integration. In 15th National Conference on Artificial Intelligence. 1998.
31. Baru, C., *et al.*, XML-Based Information Mediation with MIX. In International Conference on Management of Data (SIGMOD). 1999.
32. Palopoli, L., Terracina, G., and Ursino, D. The System DIKE: Towards the Semi-Automatic Synthesis of Cooperative Information Systems and Data Warehouses. In ADBIS-DASFAA Symposium. 2000.
33. Beneventano, D. and Bergamaschi, S. Extensional Knowledge for semantic query optimization in a mediator based system. In International Workshop on Foundations of Models for Information Integration (FMII-2001). 2001.
34. Schwarz, T., *et al.*, Efficient Domain-Specific Information Integration in Nexus. In Workshop on Information Integration on the Web. 2004. Toronto.
35. Sarma, S., Integrating RFID. *Queue*, 2(7), 2004, pp. 50-57.
36. Finkenzeller, K., *RFID Handbook : Fundamentals and Applications in Contactless Smart Cards and Identification*. John Wiley & Sons, New York, 2003.
37. Brewer, A., Sloan, N., and Landers, L.L., Intelligent Tracking in Manufacturing. *Journal of Intelligent Manufacturing*, 10, 1999, pp. 245-250.
38. Want, R., The Magic of RFID. *Queue*, 2(7), 2004, pp. 40-48.
39. Kärkkäinen, M., Increasing efficiency in the supply chain for short shelf life goods using RFID tagging. *International Journal of Retail & Distribution Management*, 31(10), 2003, pp. 529-536.
40. Internetweek.com. *Accenture Finds Significant Benefits With RFID*. 2003 June 12 [cited 2005 September 1]; URL: http://www.internetweek.com/showArticle.jhtml?articleID=10300865.
41. Barlas, D., *$4.2 Billion RFID Opportunity*. 2004 [cited 2005 September 1]; URL: http://www.line56.com/articles/default.asp?articleID=5361&TopicID=3.

42. ATKearney., *Meeting the Retail RFID Mandate*. 2003 [cited 2005 June 15]; URL: http://www.atkearney.com/shared_res/pdf/Retail_RFID_S.pdf.
43. Alvarez, G., RFID Helps Enterprises Increase Return on Assets Through Tracking. *InformationWeek*, 2004 December 21.
44. Sullivan, L., RFID Tested Against Barcodes. *InformationWeek*, 2004 March 15, p. 16.
45. Garskof, J., The Many Faces of RFID Tech. *Popular Science*, 2004 April.
46. Schwartz, E., *RFID Ripples Through Software Industry*. 2003 September 26 [cited 2005 April 20]; URL: http://www.infoworld.com/article/03/09/26/38NNrfid_1.html.
47. Osyk, B.O. and Vijayaraman, B.S., RFID in the Warehousing Industry: Where Do WERC Members Stand? *WERC Watch*, 2005 May 19.
48. Luckett, D., The Supply Chain. *BT Technology Journal*, 22(3), 2004, pp. 50-55.
49. Helders, B. and Vethman, A.J., Beyond 2005: How RFID will change the GSC. *Chain Store Age*, 2003 December, pp. 39-48.
50. Bacheldor, B., Bill Aims at RFID Privacy. *InformationWeek*, 2004 March 1, p. 16.
51. Cline, J., The RFID Privacy Scare Is Overblown. *Computerworld*, 2004 March 15.
52. Fantasia, A., Decoding ASPs. *Harvard Business Review*, 78(10), 2000, pp. 33-42.
53. Ferris, C. and Farrell, J., What are web services? *Communications of the ACM*, 46(6), 2003, pp. 31-34.
54. Foster, I., Computational Grids, in *The Grid: Blueprint for a Future Computing Infrastructure*, I. Foster and C. Kesselman (Eds.), Morgan Kaufmann Publishers, San Francisco, CA, 2004.
55. Gannon, D., *et al.*, *A Revised Analysis of the Open Grid Services Infrastructure*. Indiana University, Bloomington, IN, 2002.
56. Liu, L. and Ramaprasad, A., An integrated e-service framework for electronic medical records. In 7th International IEEE Conference on E-Commerce Technology. 2005. München, Germany.
57. Zhuge, H. and Liu, J., Flexible retrieval of web services. *Journal of Systems and Software*, 70(1–2), 2004, pp. 107-116.
58. Hagel, J. and Brown, J.S., Your next IT strategy. *Harvard Business Review*, 79(9), 2001, pp. 105-113.
59. Dewire, D.T., Application Service Providers. *Information Systems Management*, 17(4), 2000, pp. 14-19.
60. Susarla, A., Barua, A., and Whinston, A.B., Understanding the Service Component of Application Service Provision: An Empirical Analysis of Satisfaction with ASP Services. *MIS Quarterly*, 27(1), 2003, pp. 91-123.
61. Voelker, K.G., *Primer on Electronic Medical Records*. 2002 [cited 2002 January 3]; URL: http://www.elmr-electronic-medical-records-emr.com/electronic_medical_record_Primer.htm.
62. Liu, L. and Ma, Q., The Impact of Service Level on the Acceptance of Application Service Oriented Medical Records. *Information & Management*, 42(8), 2005, pp. 1121-1135.

63. Davis, F.D., Bagozzi, R.P., and Warshaw, P.R., User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8), 1989, pp. 982-1003.
64. MacVittie, L., Survivor's Guide to 2003: Business Applications. *Network Computing*, 13(26), 2002, pp. 78-85.
65. Flessner, P., XML Web services: More than protocols and acronyms. *Software Development Times*, 2001 August 15, p. 31.
66. Fremantle, P., Weerawarana, S., and Khalaf, R., Enterprise services. *Communications of the ACM*, 45(10), 2002, pp. 77-82.
67. Deitel, H.M., *et al.*, *Web Services: A Technical Introduction*. Prentice Hall, UpperSaddle River, NJ, 2002.
68. Channabasavaiah, K., Holley, K., and Edward M. Tuggle, J., *Migrating to a service-oriented architecture*. 2003 [cited 2005 April 20]; URL: http://www-106.ibm.com/developerworks/webservices/library/ws-migratesoa/.
69. Kraunelis, L., Schmelzer, R., and Bloomberg, J., *Using Web Services for Integration*. 2002 [cited 2005 April 20]; URL: http://www.xml.org/xml/wsi.pdf.
70. Srivastava, B. and Koehler, J., Web Service Composition — Current Solutions and Open Problems. In Workshop on Planning for Web Services. 2003. Trento, Italy.
71. Pistore, M., Supporting the Composition of Distributed Business Processes: Research Challenges. In WWW Service Composition with Semantic Web Services. 2005. Compiegne University of Technology, France.
72. Liu, L. and Ma, Q., Emerging e-business technologies for electronic medical records. *International Journal of Healthcare Technology and Management*, 6(1/2), 2004, pp. 1-22.
73. Foster, I., *et al.*, *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*. The Globus Project: www.globus.org. 2002.
74. Oram, A., *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*. O'Reilly & Associates, Inc., Sebastopol, CA, 2001.
75. Foster, I. and Kesselman, C., 10 Emerging Technologies That Will Change the World: Grid Computing. *Technology Review*, February, 2003, pp. 33-49.
76. Bernstein, P.A., *et al.*, Data management for Peer-toPeer Computing: A Vision. In 5th International Workshop on the Web and Databases. 2002. Madison, Wisconsin.
77. Halevy, A.Y., *et al.*, Schema Mediation in Peer Data management Systems. In 19th International Conference on Data Engineering. 2003. Bangalore, India.