A SERVICE-DESIGN OF IT INFRASTRUCTURE

Pratim Datta
Assistant Professor of Information Systems
A408, Department of M&IS
College of Business
Kent State University, Kent, OH 44242
Email: pdatta@kent.edu

Marvin Troutt
Professor of Operations Management
Department of M&IS
College of Business
Kent State University, Kent, OH 44242
Email: mtroutt@kent.edu

David Booth
Professor of Statistics
Department of M&IS
College of Business,
Kent State University, Kent, OH 44242
Email: dbooth@kent.edu

Murali Shanker
Professor of Operations Management
Department of M&IS
College of Business
Kent State University, Kent, OH 44242
Email: mshanker@kent.edu

Keywords: Modular Design, IT Infrastructure, Service design, Open Systems Theory, Integration

ABSTRACT

The purpose of this paper is to forward a conceptualization of information technology (IT) infrastructure as a portfolio of IT infrastructure services that can be dynamically reconfigured and rebalanced to adapt to evolving opportunities. The paper forwards a dynamic service-driven infrastructure framework that accommodates emerging notions of reconfiguration, outsourcing, and scalability to explain how IT infrastructure reorganize and adapt themselves via differentiation and integration.

INTRODUCTION

An organization's information technology (IT) infrastructure is constantly evolving, becoming "complex ensembles of heterogeneous artifacts, which are increasingly connected with and dependent upon one another" [10:1]. Services drive the IT infrastructure. Every process reengineering initiative, at its core, is a service redesign that drives the reconfiguring of IT infrastructure. For example, if a firm decides to shift from a UNIX mail server to an Exchange mail server, the decision is likely to be driven by a variety of service demands. Once a service case has been made, a firm needs to redesign its infrastructure to match its service portfolio. Different services require different levels of IT infrastructure integration. As a new service is introduced or removed from the organizational portfolio, IT infrastructure technologies, personnel, and processes have to be reorganized to match the service offered by the firm.

Despite advances made by past research in explaining IT infrastructure, existing research has been captive to traditional silo-based functional IT artifacts while the overarching dynamics of the IT infrastructure has generally escaped scrutiny [3, 4]. Yet, IT infrastructure is evolving. For example, while some mainframes are being used for standalone computing, others are being reassigned as content servers; while some legacy systems are being maintained for their functional efficiencies, other legacy systems are being rehauled through device-level and operating and application levels updates. Simply put, companies organize IT infrastructures in novel ways: catering to diverse demands to remain competitive but archetypes addressing such novelty and diversity are rare. If IT infrastructure is perplexing, it is because we do not have a good model to apply to its analysis [6, 9]. Most existing models treat corporate IT infrastructure as insular, static, overly simplistic, supply-centric, and lacking adaptation or fit. To address these limitations, we propose a new service-design view of IT infrastructure capable of capturing the evolving nature of IT infrastructure. However, it is important to note that our definition of IT infrastructure services refer to the service portfolio internally governed by the firm that justifies the presence of an IT infrastructure. The paper does not consider outsourced services as a part of the firm's IT infrastructure portfolio because a firm's IT infrastructure shows minimal variance when services are hosted or managed.

A SERVICE DESIGN OF IT INFRASTRUCTURE

IT infrastructure is defined as *a service-driven heterogeneous portfolio of modular and configurable services consisting of technologies, applications, people and IT-enabled routines (processes).* This definition emphasizes three key elements: (a) IT infrastructure is service-driven; (b) IT infrastructure is a portfolio of modular and configurable services; and (c) IT infrastructure encapsulates technologies and people services. Our definition reframes IT infrastructure as a key enabler instead of a base, rigid substrate of technologies. As Hanseth and Lyytinen [10] argue, IT infrastructure is capability oriented. IT infrastructure must be capable to offer requisite services for objective benefits. Departing from an old supply-push adage that services follow IT infrastructure, we find it more relevant to consider a demand-pull effect where services drive IT infrastructure. However, as IT grows to be more pervasive and ubiquitous, service-delivery becomes key- forcing reconfigurations of IT infrastructure to maintain pace with changing business models. According to Applegate et al. [1: 473], "as service delivery models proliferate and improve, the variety of IT service configurations will increase."

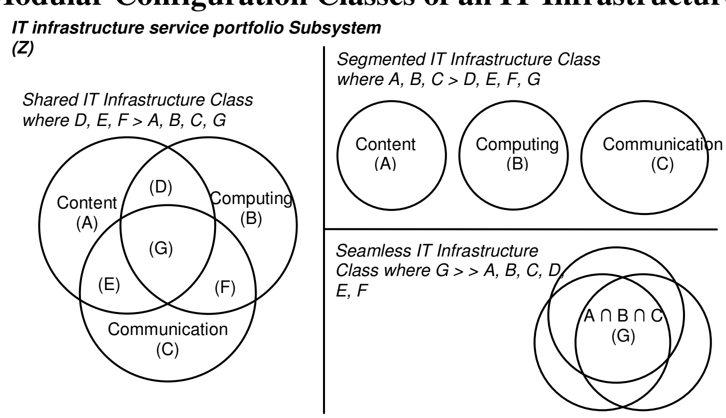
The starting point for the model is the following: IT infrastructure is predominated by three major categories, best understood as modular services subsystems: *content infrastructure* (e.g., databases, hard-drives), *computing infrastructure* (e.g., processors, monitors, programming tools), and/or *communication infrastructure* (e.g., routers, network operating systems). As modules of the IT infrastructure system, content, computing, and communication subsystems build on a combination of services that are "structurally independent of each other, but work together…a framework…that allows for both independence of structure and integration of function" [2: 63].

Each IT infrastructure subsystem consists of specific technology and human services. Technology services are resources such as devices, applications, databases, and networks that are available as packaged components or as in-house customized developments. The technology subsystem is the installed base of IT infrastructure comprising of operating- and application-level services. Operating-level services are mainly hardware and device-specific technologies while application-level services are program technologies that sit over the operating platform. Commonly, operating-level technology services require proprietary adaptors to communicate directly with the hardware while application-level technology services often rely on openly available adaptors (e.g., APIs) for indirect communications. Consequently, application-level technologies are relatively more flexible than operating-level technologies, given the proliferation of adaptors.

Human services complement technology services. Human services, by virtue of their "knowledge, skill sets, and experience." serve as "the mortar that binds all technical IT components into robust and functional services" [15: 333]. Human services are the knowledge base of a corporate IT infrastructure. In an era where technology services, by themselves, can no longer offer competitive advantages, it is only through the alignment and interaction of human services and technology services that the firms can capitalize on their corporate IT infrastructure. Moving away from a deterministic view of technology as a necessary and sufficient representation of IT infrastructure, the strategic choice view advocates the importance of the recursive cycle of interaction and influence between human and technology services, [e.g. 13, 6].

Figure 1 depicts the configurations of subsystem modules. The collectively exhaustive classes used to depict the IT infrastructure subsystem (Z) is shown in the Figure 1 where $A, B, C, D, E, F, G \subset Z$.

Figure 1: Modular Configuration Classes of an IT Infrastructure Subsystem



The seven IT infrastructure subsystems occupy one of the three classes by virtue of their underlying technology and human services, as shown in Table 1.

Table 1: Technology and Human Services for IT Infrastructure Subsystems

Subsystems	IT Infrastructure Subsystem Services		
	Technology Services		Human Services
	Operating Level	Application Level	(trained in)
Content	Disk drives, Storage Devices	Databases, Spreadsheets, Word Processors	Database Administration, Data Modeling
Computing	Hardware, Device upgrades	OS, Systems development	System Analysis, Software Engineering
Communications	Routers, Directory services	Network security, Network monitoring	System Administration, Network Design
Content/ Computing	Backup, Storage Systems	Content administration, Data mining	Business Intelligence, OLTP
Content/ Communications	SAN, NAS	Search tools, Distributed Storage	Web Development, Information Security
Computing/ Communications	Network OS, Thin Clients	Collaborative computing, Network Applications	Grid Computing, Application Security
Content/ Computing/ Communications	Enterprise servers	ERP suites, Collaborative tools	Enterprise Integration, Enterprise Security

Content Subsystem (information-based services) (A): The content subsystem includes all data and information-related services governed by the corporate IT infrastructure. The content subsystem includes technology and human services used for the acquisition, allocation and development of data and content resources needed to organize data for the purposes of cross-referencing and retrieval through the creation of information or data repositories [11]. *Technology and Human Services:* Operating-level services include magnetic-media storage (disk drives, external/removable storage devices), optical-media storage (CD, DVD); application-level services focus on data creation and manipulation (databases, spreadsheets, text/graphic editors, statistical software). Human services are resources such as database administrators, designers, and modelers used to develop, support and maintain content technologies.

Computing Subsystem (processors and system-based services) (B): The computing subsystem includes processor-based resources focused on input-output, control, and processing, consisting of operating systems environments, system applications software, and technical standards for the hardware for operation and multi-vendor compatibility [11]. *Technology and Human Services:* Operating-level services include hardware such as processor-based systems (Sun, Unix, PC, Apple), mobile-devices (PDAs, pagers), input/output devices (keyboards, monitors, printers); application-level services include stand-alone developmental software (compilers, debuggers, programming tools), system administration software (backup/recovery, emulators, system monitoring software, user management applications). Human services for the computing subsystem include programmers, systems analysts, software engineers, testers, and systems maintenance personnel.

Communication Subsystem (network-based services) (C): The communication subsystem deals with network-based resources used to support communications and provide organizational connectivity using voice and data networks, protocols, and standards [11]. *Technology and Human Services:* Operating-level services include physical hardware technologies (telephones, fax machines, routers), directory services (ADSI, X.500/LDAP), connectivity technologies (ATM. Gigabit Ethernet), network architecture (LAN, client/server, peer-to-peer); application-level services include applications pertaining to network administration (network solutions, traffic management), network protocols (VoIP, DHCP, HTTP) and network troubleshooting. Human services include personnel such as network administrators, network designers, telecommunication analysts, network service representatives, among others.

***Content & Computing Subsystem (information & system-based services) ($D = A \cap B$)*:** The convergence or integration of content and computing services gains significance especially in light of services such as data mining and business intelligence. Integration of computing and content services relates to large scale processing of databases and application data, based on human and technology services. *Technology and Human Services:* Operating-level services primarily include computing (system) hardware resources that provide access to stored content such as separate backup and storage devices while application-level services include applications pertaining to content administration, heterogeneous storage integration (data migration and synchronization) and content processing (data warehousing, data mining, data query processing). Human services are people who can develop, support and manage integrated content and computing services such as specialists in application data integration, OLTP, and data mining.

Computing & Communications Subsystem (system & network-based services) ($E = B \cap C$): The integration of system and network services is evidenced by the growth of distributed large scale processing services where processing resources are being connected via popular network protocols. This infrastructure subsystem refers to technologies that address and help integrate computing (system processing) and communications (networks), typically high end computing clusters, by connecting processors and workstations over networks based on load distribution to optimize processes and resources (e.g., Sun UltraSPARC III based computing clusters). *Technology and Human Services:* Operating-level services include technologies pertaining to secure systems-access, web applications, thin clients and terminals, network OS, distributed processing; application-level services include distributed application performance monitoring, collaborative computing, heterogeneous system connectivity (CORBA, COM+/DCOM, middleware interoperability). In this subsystem, human services include personnel trained in the operation, development, and maintenance of grid computing, application integration and security, clustering, middleware development among others.

Content & Communication Subsystem (information & network-based services) ($F = A \cap C$): As information sources have become distributed over networked environments, the need for information integration has grown steadily [14]. The Internet, particularly web-based developments have propelled the growth of integrating distributed content. Linking content repositories across the globe is becoming more and more popular, evidenced by the growth of the Internet and the World Wide Web along with enterprise search tools and networked content. What used to be a discrete, self-contained application on a server is being replaced by large relational databases at the back end and a flexible interface at the front end, connected by middleware [7]. *Technology and Human Services:* Operating-level services include technologies for the preparation, deployment, and management of content over large networks (e.g., file and content servers, Network attached Storage, Storage Area Networks). Application-level services include programs related to networked content security and assurance, search engines and interfaces and standards. Human services include personnel involved in web-development, data security, server platform engineers, among others.

Content, Computing, and Communication Services (information, system, and network-based services) ($G = A \cap B \cap C$): The growth of enterprise systems and needs for enterprise-wide services has driven the integration of content, computing and communication subsystems, merging information, system and network-based services. Enterprise application integration (EAI) is an example of that combination of processes, software, standards, and hardware resulting in the seamless integration of two or more enterprise systems allowing them to operate as one, such as building CRM systems, business-to-business integration, or leveraging legacy systems. *Technology and Human Services:* Operating-level services include enterprise servers and enterprise storage systems for processing, hosting and serving information from and to distributed sources and recipients. Application-level services include groupware, CRM, SCM, and ERP suites. Human services support the development, installation, and maintenance related to content, computing and communication subsystem integration and include personnel trained in enterprise application integration, configuration management, ERP consultants, among others.

SERVICE-DRIVEN IT INFRASTRUCTURE DESIGN

A service-driven IT infrastructure design is demand-centric. Rather than deploying routine services from specific infrastructure subsystems, a service-driven IT infrastructure is designed to accommodate service requests. Central to this thesis is that generic service deployment is no longer value-added. Correspondingly, firms need to maintain a portfolio of IT infrastructure resources that can be mapped to their portfolio of value-driven services. It is to be noted that a service-driven IT infrastructure design does not make generic services obsolete. Instead, service-driven IT infrastructure design allows firms to periodically reorganize its IT infrastructure design to emphasize on value-added IT services.

Underpinning the design and reorganization (reconfiguration) of content, computing, and communication subsystems is the reconfiguration of underlying technology and human services for reengineering processes to deliver service. For example, the need for Web-based services (content and communications) drives the employment of web-development personnel while the outsourcing of data centers would reduce the number of database personnel. A common example is how enterprise integration services lead to the reconfiguration and reengineering of common business processes and require a class of seamless services for positive service outcomes.

Different service orientations require different levels of IT infrastructure design decisions. A firm that emphasizes more on remote access services may reorganize its IT infrastructure to focus more on its communications resources while outsourcing its data center (content) and maintenance (computing) services (as managed services). A company emphasizing on collaborative computing services (e.g., grid computing) can focus more on integrating its computing and communication resources. Because a firm relies on delivering a portfolio of services, some outsourced (managed) and some in-house, their IT infrastructure needs are different, leading to different IT infrastructure designs. As new services are demanded by evolving strategies, firms have to reorganize, upgrade, and scale their IT infrastructure resources to accommodate these service changes.

A PORTFOLIO MODEL OF IT INFRASTRUCTURE SERVICES

A firm's portfolio of IT infrastructure services must be aligned with a set of service objectives that the firm wants to achieve. We agree with Ciborra et al.'s [6] recommendation that the IT infrastructure portfolio must be managed like an investment portfolio consisting of modular investment assets (various financial instruments). The portfolio must be reorganized and reconfigured to balance the weights of assets to suit particular service outcomes. Service-driven IT infrastructure portfolio decisions can therefore be formulated as a portfolio selection problem for maximizing service returns and minimizing service return gap, where the expected return is the return from all allocated subsystem assets in an organization's IT infrastructure portfolio. In optimizing the portfolio by reducing the gap from the difference between expected and actual service returns from a given IT infrastructure portfolio made up of different allocation of subsystem assets, organizations can reconfigure their IT infrastructure subsystem assets

Assume that a firm has an IT infrastructure portfolio budget M_0 that can be distributed/ configured to build an IT infrastructure portfolio by choosing the optimal services mix out of n (n=7, i.e., A, B,….G) possible types of IT infrastructure services, S_j, j=1, 2, …, n. Let R_j be the set of service benefits (measures in $) offered by the IT infrastructure services type S_j (a random variable). The choice of IT infrastructure services to build the IT infrastructure portfolio can be understood by its allocated weight, x_j, from M_0 towards IT infrastructure services S_j where $0 \leq x_j \leq 1$ and $\sum_{j=1} x_j = 1$.

M_0 = *IT infrastructure portfolio fund in dollars*
n = *Number of available IT infrastructure services categories where* $n = (n_1, n_2, \ldots . n_7)$
S_j= *Types (nominal variable) of IT infrastructure services,*
x_j= *Allocation, in %, of IT infrastructure services portfolio fund towards* S_j. *where* $\sum x_j = 1$
R_j= *Actual variances in services from IT infrastructure services type* S_j
r_j = *Expected service variance from IT infrastructure services type* S_j
q_j= *Difference between expected and actual service variances.*

Let $E(R)$ denote the mathematical expectation of a random variable R to define expected service returns (output) from a particular type of IT infrastructure services:

$$r_j = E(R_j), \quad q_j = E\,(E(R_j) - R_j)$$

where r_j and q_j are the expected rate of service returns and the expected entropic gap (deviation between expected and actual service returns) for asset S_j, respectively.

Therefore, the expected service returns from IT infrastructure portfolio allocations $\boldsymbol{x} = (x_1, \ldots, x_n)$ is shown by

$$R(x_i, \ldots, x_n) = E\,[\sum_{j=1} R_j x_j] = \sum_{j=1} E(R_j) x_j = \sum_{j=1} r_j x_j$$

Following Feinstein and Thapa [8] and Cai et al. [5], we construct a portfolio selection model where $x_j \geq 0$ and the acceptable/tolerable service variance can be restated as,

$$w(x) = max_{1 \leq j \leq n}\, E\,(E(R_j) - R_j) x_j = max_{1 \leq j \leq n}\, q_j x_j$$

assuming that, $q_j > 0$, j = 1,….,n (i.e., all service returns from subsystem assets exhibit some entropic gap) and where the degree of reconfiguration of IT infrastructure portfolio φ_{ITR} is subject only when $R_j - E(R_j) < 0$ and not otherwise (i.e. 0).

The aforementioned perspective adds to our current understanding of IT infrastructure in several ways. We believe that the proposed IT infrastructure model represents a major step in addressing the call for reconceptualizing IT artifacts from being "multiple, fragmented, partial, and provisional," to developing models capable of supporting theoretical work that reflects "the emergence and evolution of IT artifacts as complex and changing techno-social processes existing in time and over time" [13: 132].

First, an immediate contribution to practice is the availability of a new representation of IT infrastructure that provides an analytical tool for decision-makers and infrastructure designers. For firms in the process of developing or updating their IT infrastructure, the ability to classify their adopted/proposed IT portfolio using the systems model we have described should prove valuable in representing and evaluating different options. If firms can logically separate technologies by their functionality, human resources and services, resource allocation choices and the implications of those choices can become more apparent. A systemic view of IT infrastructure allows firms to unbundle their IT infrastructure into observable and classifiable components.

For researchers, use of the systems paradigm to understand IT infrastructure offers new horizons for investigation. First, it removes the traditional bundling of all technologies into one basket, equating databases with servers. Instead, it offers a logical classification based on functionality. Furthermore, the systems lens allows us to capture the dynamics of an IT infrastructure. Not only can we unbundle IT infrastructure technologies, but we can also represent the modular variations through configurations and reconfigurations over time.

REFERENCES

[1] Applegate, LM., Austin, A, and McFarlan, FW. *Corporate Information Strategy and Management: Text and Cases*. 7th ed. New York: Irwin/McGraw Hill, 2006.

[2] Baldwin, CY and Clark, KB, (2000) *Design Rules*, MIT Press.

[3] Broadbent, M (1995) *Honda's Civic Lesson*, Business Week, September 18, 26-28.

[4] Broadbent, M, Weil, lP and St.Clair, D (1999) The implications of information technology infrastructure for business process redesign. *MIS Quarterly* 23(2), 159-182.

[5] Cai, XQ, Teo, KL, Yang, XQ, and Zhou, XY, (2000) Portfolio optimization under a minimax rule. *Management Science* 46, 957-972.

[6] Ciborra, CU, Braa, K, Cordella, A, Dahlbom, B, Failla, A, Hanseth, O, Hepso, V, Ljungberg, J, Monteiro, E, and Simon, KA (eds.) (2001) *From Control to Drift. The Dynamics of Corporate Information Infrastructures*. Oxford University Press, New York.

[7] Coffin, C (2006), As Networks and Applications Converge, New Challenges Emerge, Analyst Corner, *CIO*, February 28th.

[8] Feinstein, CD and Thapa, MN (1993) A Reformulation of a Mean-Absolute Deviation Port-folio Optimization Model. *Management Science* 39, 1552-1553

[9] Fomin, VV (2003) Standard Making: A Critical Research Frontier for Information Systems, *MISQ Special Issue Workshop.*

[10] Hanseth, O and Lyytinen, K Theorizing about the design of information infrastructures: design kernel theories and principles, Working Paper

[11] Keen P (1991) *Every Manager's Guide to Information Technology*. Harvard Business School Press, Boston, MA.

[12] Orlikowski, WJ (1993) CASE Tools as organizational change: investigating incremental and radical changes in systems development. *MIS Quarterly*, 17(3), 309-340.

[13] Orlikowski, WJ and Iacono, CS (2001) Research commentary: desperately seeking the 'IT' in IT research--a call to theorizing the IT artifact. *Information Systems Research* 12(2), 121-134.

[14] Rudensteiner, E., Koller, A and Zhang, X (2000) Maintaining data warehouses over changing information sources. *Communications of the ACM* 43(6), 57-62.

[15] Weill, P and Broadbent, M (1998) *Leveraging the New Infrastructure – How Market Leaders Capitalize on Information Technology*. Harvard Business School Press, Boston, MA.