

STRUCTURAL AND CONCEPTUAL CHANGES IN THE INFORMATION LANDSCAPE : THE NEW CHALLENGES FOR INFORMATION PROFESSIONALS

Saji Gopinath
Associate Professor, IIM Kozhikode

ABSTRACT

The information space has witnessed phenomenal changes in the last few years. The fast growth of information repositories and trend towards protocol standardisation had completely revised even the conceptual frame work of the ubiquitous platform of internet. In the application domain, need for systems that mine out knowledge from the distributed data and information resources, revealing the hidden footprints in the seemingly unrelated resources is on the rise. The developmental philosophy of the information/knowledge repositories is moving from proprietary standards to the open standards, leading to several issues on the long term management of such systems. All these developments, happening at a rapid pace is posing newer challenges to the information professional involved in the design, development, deployment and management of the information repositories in the cyber space. This paper provides an overview of the developments happening in this area and focuses on the new issues emanating out of such changes. Rather than presenting a chronological review of the specific developments, the paper looks at the short and long term impact of philosophical and structural shifts happening in the information space and explore the methodologies for exploiting such changes. Specifically three aspects of structural and conceptual transformation are focussed in detail. On the conceptual front we focus on the transition of the internet from the “web” to “bow-tie architecture” and the subsequent death of “small world theory” in the design, deployment and access strategies of the information. On the structural front, the paper looks at the impact of “open standards” and “free ware” movement as against the dominance of proprietary domains. The strategic and tactical implications of this is analysed from the perspectives of the developer, user and the infomediaries. On the application front the paper focuses on the challenges of repositioning of information repositories as “knowledge estates”. As the users are increasingly demanding the extracted knowledge from the information resources, the information landscape is witnessing radical changes forcing information science professionals to reposition as “estate managers” of knowledge space. Even though these three changes are happening independently, led by professionals from different domains, they pose several conflicts and dilemmas to the integrators. The paper provides an exploratory approach to understand and analyse these dilemmas which are being faced increasingly by the IS professionals.

Introduction

Ever since the academic community and business organisations started using the interlinkage of network of computers-popularly known as internet- for speedy and seamless transactions, several research studies have been conducted to understand the conceptual nature of the internet (<http://www9.org>). As Internet is self managed, under the loose framework prescribed by organisations like IANA and InterNic, the multifold development of websites and information repositories were quite rapid in the recent past challenging the very nature of the Internet. The emergence of various access channels, servers with different processing capabilities and the positioning strategies adopted by the web developers, etc. had its impact on the structure of the Internet. The changing nature of the net is very significant for the information science

professionals, who are engaged in the physical consolidation of distributed information repositories

The age old debate between data and information is currently being replaced by the conflict between the demand for information and knowledge. With the emergence of multidisciplinary applications, need for integrating the information points, and to mine out the knowledge contained them is on the rise. The corporate world is witnessing a phenomenal growth in Application Integrators while the Information users are increasingly demanding systems for seamlessly integrate incompatible- syntactically or semantically- information/data resources from the IS professionals. The role of IS professionals are thus changing from one of gatekeepers to that of knowledge estate managers

While rapid development is happening on the storage and access technologies of information, serious debates are also focusing on the economic relevance of such developments. With majority of information science projects are State funded, the commercial viability of the IS and IT projects is coming under close scrutiny of financial pundits and corporate strategists. The financial impact of adoption of open standards is being questioned widely due to the “strategic inefficiencies” resulting from such a move. With IT falling into infrastructure domain, the proponents of Open software movement argues strongly for its adoption indicating the “operational efficiencies” achieved by such a move. While long term sustainability of any venture is heavily dependent on the financial viability, the need for a clear understanding of this dilemma is very crucial for the Information Science Professionals.

This paper looks at these three themes. The structural and philosophical changes happening on the internet is discussed in the next section. The need and impact of transition from information space to knowledge estates is discussed in section 3. The section 4 reveals the ongoing discussion on the strategic and operational impact of the adoption of freeware standards in development. Each section discusses the relevance of the respective theme on the working of modern IS professional. Section 5 provides the summary and concluding remarks stressing the need for more focussed research on this area.

1.0 Theme 1: Structural Changes in the Information Space

The early studies on the internetworking of computers-internet- revealed a connected graph structure (www.www9.org/w9cdrom/160/160.html). With all the websites believed to be interconnected, the *small world theory*¹ (www.searchengineposition.com) was in vogue till recently. Small world theory, which estimates an average of 17 clicks (steps) for transiting from any webpage to another webpage is very relevant to information professionals in formulating search strategies as well as to provide accessibility to the information resources. The small world theory had conceptually modelled internet as a *web* (see fig 1) with seamless transition possible across the web sites

2.1 Small World Theory

Small world theory had its own impact on the design of digital libraries. With modern digital libraries providing seamless access to information distributed across the web, the digital library² developers have to create search and access strategies for providing faster information to the consumer, who is unaware of the underlying structure. With small world theory promising access to any resource in the world within 17 steps, the IS professional need to adopt only generic search strategies which skim from site to site in search of relevant

¹ The small world theory is initially proposed in 1960 by a social psychologist Stanley Milgram to indicate that every individual in the planet could be linked to anyone else by a chain of only six other people (Small World Research Project; Columbia University 2002). This was later extended to Internet

² here digital libraries is used in a generic sense to include virtual libraries also

information. The focus will be more on the faster point-to-point transition rather than on the structural aspect of the web. Most of the digital libraries (www.digitallibraries.com/design.html) are built assuming Small world theory as the underlying postulate.

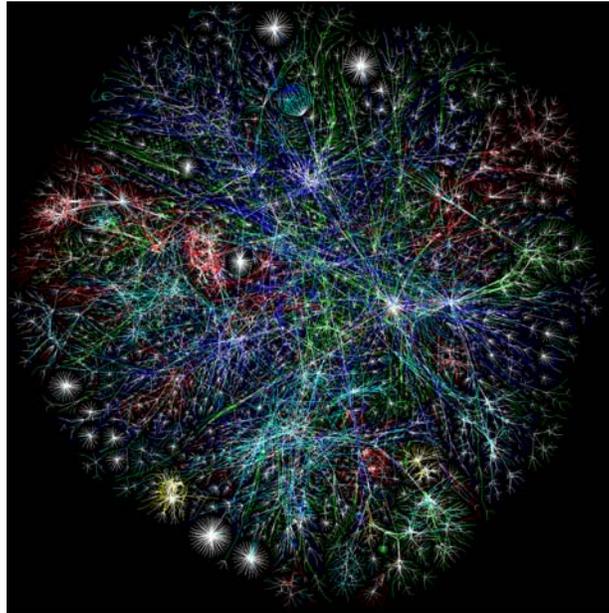


Fig 1: Conceptual structure of Internet (www.web-opte.org)

2.2 Bow Tie Theory and Emergence of Deep Web

Though unrelated, the newer revelations of the structure of the internet-challenging the very existence of the web structure- namely the bow tie structure as proposed by Border et.al (2000) and presence of deep web (<http://library.albany.edu/internet/deepweb.html>) has phenomenal impact on the search, access and positioning strategy development for IS professionals.

According to a recent research conducted by IBM, Compaq and Altavista, the Internet is like a bow-tie (see fig 2 below) with only around 56 million web pages (about 25% of total websites) are in the Strongly Connected Core that follows the Small world theory. In most of the other cases the probability of finding a path from one site to other, without accessing using an explicit URL address is almost zero. This finding has got significant relevance in the design of digital libraries as well as in adopting search strategies. If a website is hosted in a *tuber* or *tendrīl*, there is little chance of accessing their content using normal search strategies available with commercial virtual library software. With IS professionals adopting a vanilla approach for hosting the information repositories, there is a very high probability that a site with very useful content is residing in a remote corner of the bow-tie. Thus the IS professionals should restructure their search/access strategies as well as to adopt extreme programming measures to ensure that the content, when uploaded is residing on a site that forms the part of Strongly connected Core. The difference between Small world web structure and Bow-Tie structure is summarised in table 1

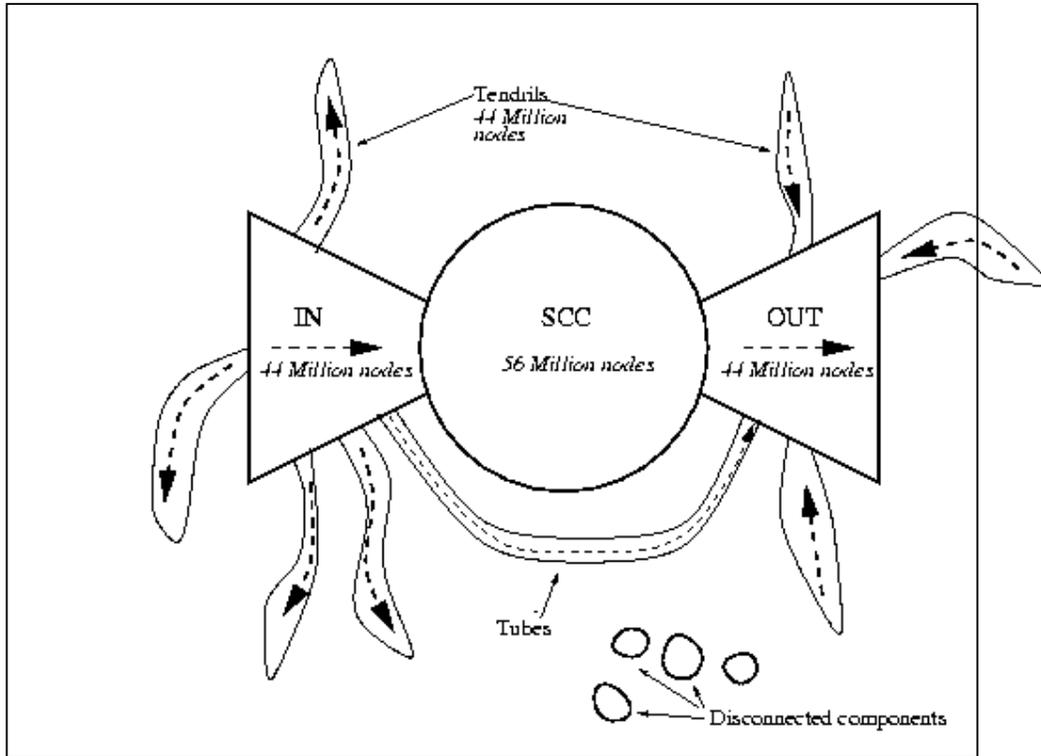


Fig 2: The new Bow-Tie Structure of Internet (conceptual mapping)

Table 1: The transition from Old Structure to New Structure of Internet

Item	Old Theory	New Theory
Conceptual Structure	Web	Bow-Tie
Interconnectivity	All nodes(web sites) are interconnected	Only the nodes in Strongly connected Core is connected. There are many IN –with connection to SCC and many OUT-connection from SCC sites. There are also Disconnected sites and interconnected tubers and tendrils
Accessibility	Any site can be accessed from any other site within a limited (usually less than 20) number of clicks or transitions	There is a high probability (almost 0.75) that there is no connection exists between any randomly selected pair of websites
Search Strategy design for a digital (virtual) library	Simple vanilla approach is enough as all sites are connected. The emphasis is on deep search within a site and meta data harvesting	Search strategy is quite complex. The exact position of the sites have to be obtained. While deep search within a site is necessary, newer strategies to locate the site is to be included

Hosting and Promotion	Simple. It is irrelevant where it is hosted as all sites are interconnected. The guiding factor is the bandwidth speed design simplicity for faster access	Need to be very choosy and careful. It is important that the information repository lies in a web site in SCC. Look for ward inward and outward references before hosting. The bandwidth speed, while is important is not as crucial as the position of the site in the Bow-tie
Challenge for the IS professional	Need not have to bother about the structure of the web and exact position of hosting point Need for tracing the search path is not very useful as the benefit is marginal The need for Technology knowledge for IS professional is not crucial	Have to be extremely careful while choosing site for hosting the site as well as to adopt newer analytical techniques into search strategies Archiving the sites and deployment of high power robot programmes is very important to ensure better spanning of the Internet. The IS professional should be more knowledgeable in IT tools and Techniques as site design is very crucial in enhancing the effectiveness of the virtual library.

While Bow-Tie theory, discussed above throws open new challenges due to the changing nature of the interconnection between the sites, the methods of storage of information in the repositories is also creating new problems for the Information Science professionals. With the realisation that substantial (over 80%) of the information repositories are lying submerged in the Deep Web³, which is seldom indexed by popular search engines, even if the site hosting such repositories form a part of SCC mentioned earlier, the IS professionals are forced to develop specific search strategies to get the information of the databases that form the part of the Deep Web. Due to commercial interests (discussed in section 4) the many players in the information industry had initially adopted proprietary standards for packaging and storing information. This has necessitated the need for understanding the syntax and semantics of such data sources to develop specific search strategies (www.completeplanet.com) to mine out the vast amount of data. While many IS professionals are working on developing specific search strategies for well known databases in medicine, science and humanities, a generic model is yet to emerge to solve this problem. In the absence of such a generic model, the amount of data accessible to a common digital/virtual library design is only around 20% of that available in the net. This percentage will be further reduced substantially, when one considers the revelations of the Bow-tie theory explained earlier. Hence in order to make the searches effective, the IS professional should not only concentrate on analytical techniques to spot the site lying on a distant-inaccessible- corner of the bow-tie but also should be capable of developing site specific search strategies to tap the data that lie invisible in the site. On the development front, they should be ensure that the site is hosted with in the SCC and the content should be following standard protocols so that it could be indexed and searched using common search engines.

2.0 Theme 2: Transition from Information to Knowledge

The changes in the structural nature of the web provides fresh challenges to the IS professionals in formulating search and positioning strategies. Along with this, the changes happening in the use of information is also throwing open a large number of new issues. It is

³ Deep Web is sometimes known as Invisible web also

widely recognised that a major advantage of digital form of accessing data is the flexibility of processing data and providing information relevant to the end user. For instance, while a physical data source like a text book can provide facts and figures, an e-book is capable of taking inputs from user and present the data in a way that useful and relevant to the user. Hence most of the digital data accessing devices are equipped with such information processing systems embedded in them. It could be a simple programme that categories the data or could be a very complex one that provide simulation capabilities. As shown in Figure 3 below, the cross references happening in research is on the rise. In other words the demand for content that spans multiple disciplines is on the increase creating new challenges for the managers of the data resources. Moreover the demand of the users for “knowledge” rather than for information is also on the rise.

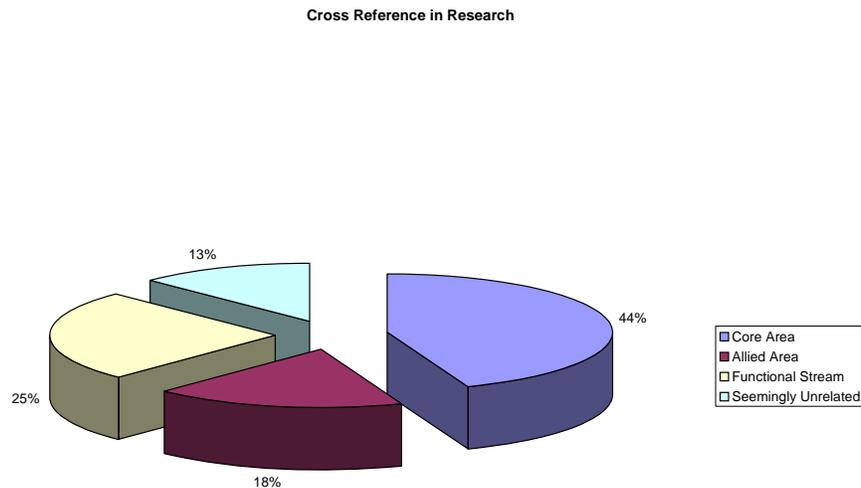


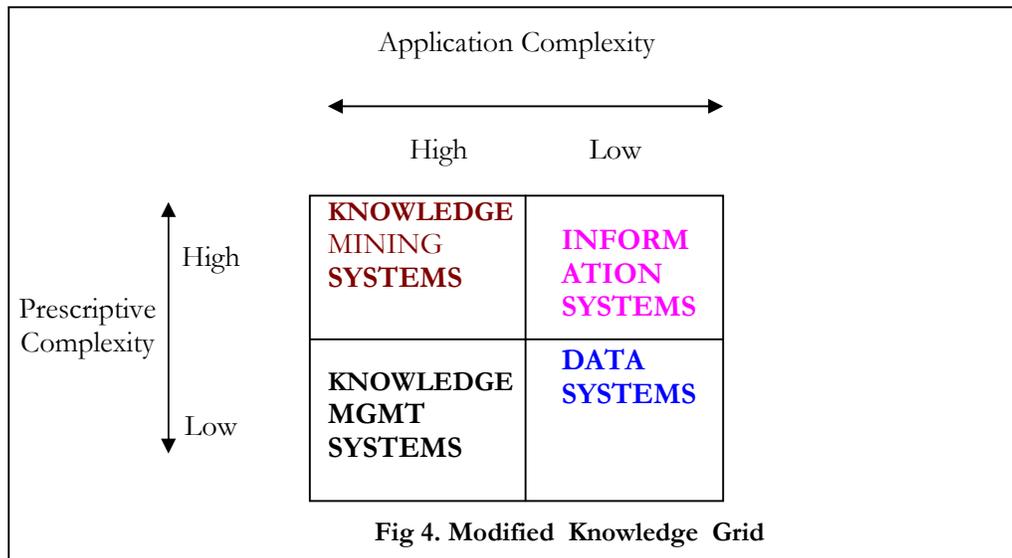
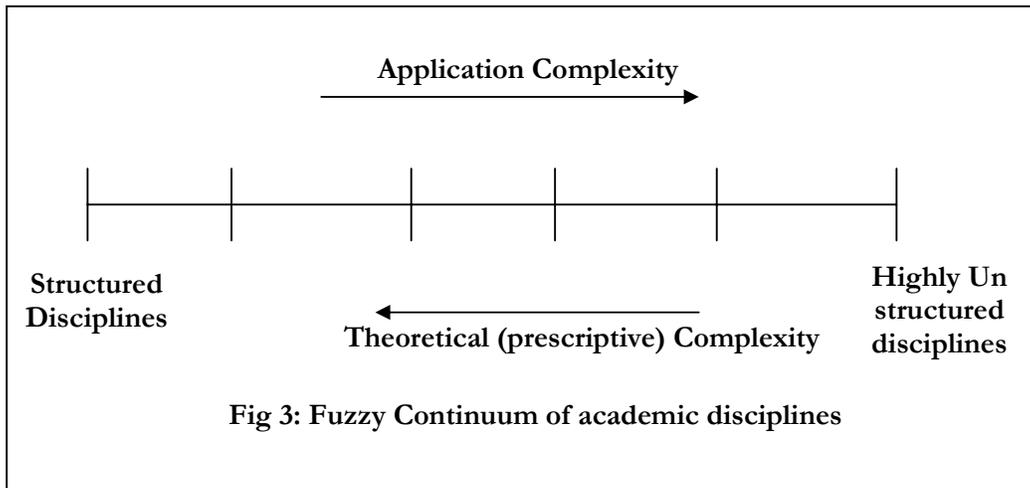
Fig 3: Cross References in a typical research study

3.1 Information and Knowledge

Information is defined as processed data while Webster defines knowledge as *all that has been perceived or grasped by the mind*. While the proponents of Knowledge theories, are increasingly arguing for the transition of IS professionals to Knowledge Service professionals, there is also a strong counter argument that most of the disciplines at best require only processed data. Knowledge Management, the later argues, is more of canning the information for archival purposes and seldom include the feelings and perceptions of the generators of information (www.knowledgemanagement.org). It is true that a large number of Knowledge management Systems are just a hype, an incremental improvement over the conventional Information Systems. However it also important to understand that for a large number of academic disciplines, mere information is not enough to do meaningful tasks. Gopinath and Sahadev (2004) had tried to resolve this paradox in the context of Executive Learning. Based on their research findings the various subjects of study could be plotted on a continuum based on the prescriptive and application complexity⁴. It is seen (see figure4) that

⁴ This classification is derived based on the Kol's learning Cycle (1995). For details see Gopinath (2003)

as the implementation complexity increases, the need for “knowledge” component also increases proportionately. The modified knowledge requirement grid is given in Fig 4



From the Figure 4, it is clear that the level of Knowledge Management increases with level of Application complexity. For instance, in a discipline like Mathematics, which is having high level of theoretical (prescriptive) complexity (High level of AC as per Kolb’s model (1995)) the users need more of Information Systems that simplify the data-the facts and figures- so that learning process will be effective. On the other hand, if a discipline like Management is considered, the theoretical complexity is low in comparison with application complexity (High level of AE-CE as per Kolb’s model). In such disciplines, mere supply of information won’t suffice the learning and more complex Knowledge management systems have to be developed. With the users adopting to multi-disciplinary learning, the complexity of learning process is increasing in both dimensions resulting in the need of yet another class of systems that provide high level of information derived through complex analytical models along with knowledge mined by creating seemingly unrelated patterns within the information and data elements.

As providers of the information resources and managers of data repositories, the modified Knowledge Management grid is posing some new challenges to the information scientists and professionals. With users moving into multi-disciplined studies, the data resources have to be converted into knowledge estates, where in the data, information, knowledge management and knowledge mining systems have to co-exist as per the user need. The IS professional, have to reposition themselves from their classical gate keeper role to the role of a facilitators, who could mine the knowledge out of the vast repositories of data that exist in distributed sources. This is no easy task given the syntactical and sematic differentials existing across the distributed data points. During the recent past several attempts have been done to create standard protocols for data (information) aggregation. However most of these are limited to the data packaging (web services for instance) and very little attempt is made to standardise the content, which is mandatory for the development of any knowledge management systems. For developing knowledge mining systems, IS professionals should acquire new skills in identifying the hidden patterns and be proficient in the domain of data mining and business intelligence. One of the major stumbling blocks is prevalence of multiple standards adopted in developing and storing the data repositories. Even though one can argue for standardisation and open software movements, there are serious financial flaws in such arguments. This form the third theme of this paper

3.0 Theme 3: Moving from Proprietary to Open Standards

A major chance that is happening in the Information landscape in the recent past is the emergence of Open Standards as against the proprietary standards for data creation and management. As shown in fig 5, the recent developments in Technologies are all XML based standards. Proponents of Open standards, argue that this development is very beneficial to ensure seamless integration of information/data repositories across the world.

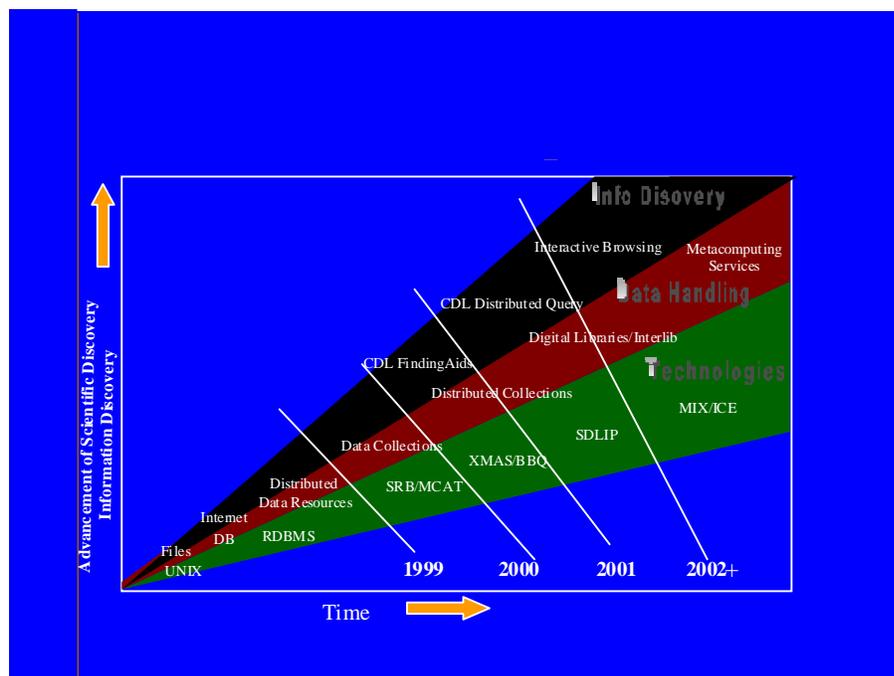


Fig 5. Emergence of Technologies in the recent past

While, the open standards are necessary for integration and are having a social significance, it is important to understand the financial and strategic implications of this movement. Ever since the seminal article of Nicholas Carr (2003) had challenged the strategic advantages of

the Information technology, the debate between the strategic and operational advantages of proprietary and open technologies are the focus of discussion in IT circles. The dilemma of the adoption of these technologies is summarised in figure 6

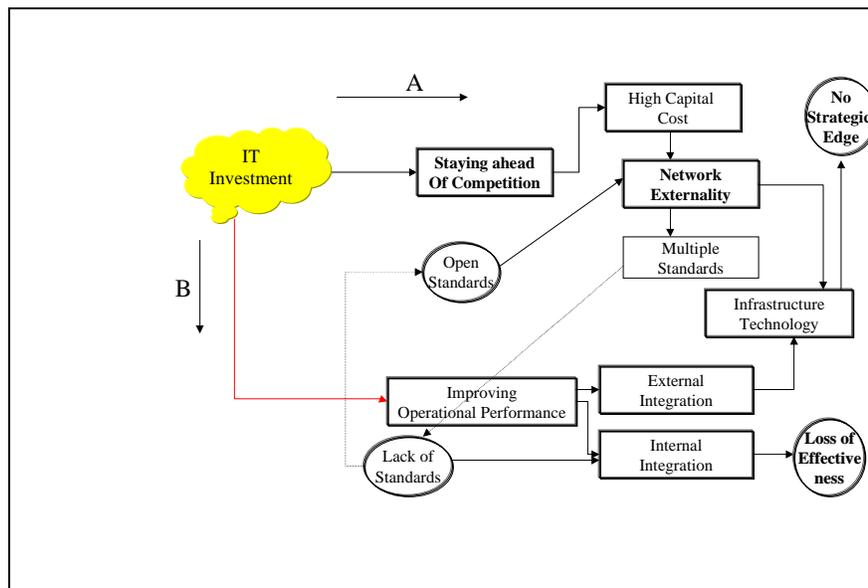


Fig 6. Impact of Open Standards on Strategic relevance of IT investments

In the figure, the two paths (A and B) are shown. If Investment is made on developing Information resources, huge capital costs involved will force the investors to adopt proprietary standards so that the amount could be recovered through resulting network economics. However this leads to multiple standards. If one follow path B, the IT investment for increasing operating effectiveness, the prime objective will be to ensure internal and external integration, for which multiple standards form as stumbling block. The Open Standards movement have emanated out of this need. However this work against basic principles of network economics and would erode the strategic edge of the IT investment. This dilemma which Information technology industry is facing today, with resultant diminishing returns on IT investments (www.mckinsey.com) is yet to affect the IS domain, mainly due to the large level of State funding existing in the sector. However, over the period the State funding to digital library projects are on the decline and the financial viability of such projects are causing increasing concerns for the investors and IS managers. With corporate sector getting into Information Industry in a big way, the major challenge which IS professionals are going to face in the near future is to resolve the investment paradox of Information Technology.

4.0 Conclusion

This paper presents the structural and conceptual changes happening in the information landscape. The paper discusses the changes in the structure of IT infrastructure-the web- and explores how information professionals should re-orient themselves to get the best out of the changed scenario. As a second theme, the paper focus on the movement of users into Knowledge management domain and stresses the need for Information professionals to transform into Knowledge mining experts. The stumbling block in this transition, the paper argues, is the presence of multiple standards, which were introduced for commercial reasons. The Movement to Open Standards is not easy, with drying down of Government Funding in the Digital Library projects as Return on Investments in such projects may cause severe questions on the financial viability of such ventures. Solving this paradox will be the major challenge for the IS professionals in the near future.

References

The references of various websites are provided in the body of the text

1. Carr, Nicholas.G (2003) *"IT Doesn't Matter"* Harvard business Review, May 2003
2. Gopinath.S, (2003) *" Pedagogical paradigms for E-learning"* Proceedings of International conference on E-learning, Chennai
3. Gopinath S and Sahadev S (2004), *" E- Learning : Assessment of pedagogical effectiveness in on-line management education"* Proceedings of CELDA 2004 Lisbon, Portugal
4. Kolb. D, Rubin. I, & Osland, J (1195) *"Organisational Behaviour : an experimental approach"* Edn 6, Prentice Hall