

White Paper: Technology Issues Associated With Modernization of the Securities and Exchange Commission's Electronic Data Gathering, Analysis and Retrieval (EDGAR) System

Computer Science and Telecommunications Board
Commission on Physical Sciences, Mathematics, and Applications
National Research Council
Washington, D.C.
February 1996

CONTENTS

[ABSTRACT](#)

[INTRODUCTION](#)

Increasing Demand
Changes in the Policy Environment
New Technologies

[SYSTEM ARCHITECTURE](#)

[SYSTEM AND DATA SECURITY](#)

Confidentiality, Authentication, and Access Control
Data Integrity

[DATA STANDARDS](#)

Alternative Formats
Multiple Filings
Architectural Solutions

[DATA STORAGE AND ANALYSIS](#)

[DISSEMINATION](#)

[MANAGING FUTURE TECHNOLOGICAL CHANGE](#)

[POSSIBILITIES FOR FUTURE CSTB STUDY](#)

APPENDIXES

A [Invited Meeting Participants](#)

B [Government Participants and Observers](#)

ABSTRACT

This paper, summarizing a one-day brainstorming session, was prepared by CSTB Program Officer [Jerry Sheehan](#). An earlier draft was reviewed by the CSTB chairman and other non-SEC participants in the brainstorming session. It does not represent the output of a formal National Research Council project, but it includes content that may help in planning such a project.

The U.S. Securities and Exchange Commission (SEC) plans to embark on a program to modernize the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system it uses to collect, process, and disseminate information regarding securities offered in U.S. markets. This effort is intended to improve the current system by incorporating new technologies into EDGAR's major subsystems and to respond to changes in the operational environment, including calls to privatize elements of the system. As such, the SEC is considering many topics in the redesign process: new architectures that take advantage of the capabilities offered by the Internet; practical alternatives for submitting documents to the SEC in formats

other than ASCII; off-the-shelf technologies for conducting full-text searches across millions of pages of documents; and options for increasing competition in the dissemination of EDGAR information.

At the request of the SEC, the Computer Science and Telecommunications Board (CSTB) of the National Research Council (NRC) convened a one-day meeting on December 1, 1995 in Washington, DC to help the SEC prepare for the redesign of EDGAR. The brainstorming session explored SEC business processes and requirements, technologies applicable to the EDGAR system, and more formal ways in which CSTB could help the SEC plan and manage the EDGAR upgrade. It provided an opportunity for representatives of the SEC to meet with members of the technical and investment communities and examine candidly issues relevant to the modernization program (participants are listed in the Appendixes A and B). This document summarizes the results of that meeting, as well as subsequent informal discussions among participants and CSTB staff. It outlines many possible options for addressing technical issues raised by the SEC and reveals the close linkages between many of the SEC's objectives.

INTRODUCTION

The U.S. Securities and Exchange Commission (SEC) is an independent, nonpartisan regulatory agency with responsibility for administering federal securities laws. These laws were designed to ensure fair and proper operating of securities markets in the United States and to ensure that investors have access to disclosure of all material information concerning publicly traded securities. This latter mandate is discharged primarily through the SEC's public disclosure programs, which require publicly traded firms and organizations preparing public offerings of securities to disclose financial and other significant information about them. The SEC is responsible for disseminating this information to the investment community.

The SEC uses an Electronic Data Gathering, Analysis and Retrieval (EDGAR) system to facilitate the collection, analysis, and dissemination of filing information in electronic form. The system consists of three primary components: Receipt and Acceptance (R&A), Analysis and Review (A&R), and Dissemination. The R&A subsystem accepts documents from filers via modem in an ASCII text-only format. The SEC provides filers with a free customized software package, EDGARLink, to facilitate transmission of data and aid in tagging documents with Standard Generalized Markup Language (SGML) tags. The A&R subsystem, currently operated by LEXIS-NEXIS (LEXIS-NEXIS is a division of Reed Elsevier, Inc.), provides for electronic storage of documents on a centralized computer system and allows SEC staff to conduct full-text searches across 20 million pages of archived information. The dissemination subsystem consists of two parts: a World Wide Web site maintained by the SEC that provides low-volume users with raw filing data with a 24-hour delay; and a high-speed, real-time dissemination system, operated by LEXIS-NEXIS, that distributes information to commercial resellers, who, in turn, sell value-added services and information to investors. Nine resellers currently subscribe to the high-volume service at a cost to each of \$210,000 per year.

Approximately three-quarters of all publicly traded companies use EDGAR to submit filings to the SEC, though many companies hire third-party intermediaries, such as financial printing companies, to prepare and submit their filings. In total, the SEC receives roughly 300,000 documents per year from some 14,000 filers (including intermediaries). Documents average 42 pages in length, bringing the total volume of filings received annually to 12 million pages. The SEC estimates that, on average, most companies submit five to seven filings a year, including an annual report (Form 10-K), quarterly reports (Form 10-Q), a proxy, and several reports of current events (Form 8-K). A small number of highly active companies submit a much larger number of filings.

The SEC plans to modernize the EDGAR system, beginning in January 1997. While the timing of the modernization program is linked to the expiration of the existing contract for operating EDGAR, the program is being undertaken to respond to larger changes in the operational environment, including increasing demand for EDGAR services on behalf of filers and users; changes in the policy environment that include a call for greater privatization of EDGAR; and new technological capabilities that could allow EDGAR to accomplish its objectives in ways not possible when the system was originally designed. These changes influence not only the requirements placed on EDGAR, but the ways in which those requirements can be met.

Increasing Demand

Over the next several years, the SEC expects the number of EDGAR filers to increase dramatically. Part of the increase derives from the growing market for mutual funds. Mutual fund companies continue to devise

new series of offerings and revised fee structures, each of which requires a filing with the SEC. While it is possible that the mutual fund industry will undergo dramatic consolidation in the near future, the trend is toward increased filings. An expansion in the number of different types of securities offered in the marketplace is also contributing to the increase in filings. Offerings such as asset-backed securities and derivatives create new issuers and filers who must submit information to the SEC. Furthermore, international trading of securities is likely to become more prevalent. As international accounting standards and trading rules become more consistent across borders, U.S. exchanges are likely to become more closely linked with foreign exchanges, increasing the number of documents the EDGAR system would need to track.

Demand for EDGAR information is also likely to grow, especially among individual investors who access the SEC's World Wide Web site. In the first two months since the Web site became operational, investors accessed 4 million EDGAR files (approximately 2 million files each month) and downloaded 100 gigabytes of data. The SEC's World Wide Web site became operational in September 1995, following a trial period in which the site was jointly operated by New York University and the Internet Multicasting Service. About 60 percent of the Web site users are from the commercial sector, with the balance split between military and educational users. The SEC anticipates much higher use as more homes become linked to the Internet. Estimates of the number of Internet users in the United States range from about 9 million to 24 million. See "Internet Survey Estimates 9.5 Million Users in U.S.," *Wall Street Journal*, Jan. 12, 1996, p. B2. At the same time, demand by professional analysts who access EDGAR data through subscription services is likely to grow. These analysts support investment decisions by large brokerage houses and financial institutions and, as such, have a need for real-time, in-depth information that allows them to conduct cross-company comparisons. As more information becomes available on EDGAR and searching routines become more sophisticated, the SEC expects these professional analysts to increase their demand for EDGAR information.

Changes in the Policy Environment

Changes in the policy environment will also impose new requirements on EDGAR. Congress is considering legislation that would require the SEC to solicit proposals for privatization of EDGAR. Though not specifying the type of privatization desired, the legislation implies that the SEC should consider expanding the role of private sector actors in operating or managing parts of the system. There may be some limits on the extent to which EDGAR can be privatized. For example, the SEC may need to remain part of the system in order to receive and review filings and perform its internal analysis, or to lend a greater degree of credibility to the filing process than if filers were allowed to submit documents to a private entity. The issue appears to be one of finding an appropriate balance between private and public control of EDGAR that ensures that the system is both responsive to the SEC's mission and needs and provides the proper incentives for the private sector to play a greater role in operating the system.

Budgetary constraints will also affect EDGAR. In light of recent debates over the federal budget and attempts to reduce the deficit, projects like EDGAR will be under even greater pressure to demonstrate their benefits and justify their costs. Congress has proposed a new financing structure for EDGAR that would shift its support from filer fees to a federal appropriation. Though designed to provide more consistent funding for EDGAR, the measure may subject EDGAR to more intense congressional scrutiny and control, making cost-effectiveness a more crucial part of the modernization program.

New Technologies

New technologies will offer EDGAR new ways of meeting these challenges. The rise of client/server environments and the expansion of network infrastructure, for example, might allow the SEC to consider more distributed architectures for EDGAR that could give the private sector a greater role in operating or managing the system. Advances in database management and data standards may also enable EDGAR to perform new tasks or to handle existing tasks more efficiently. At the same time, the rate of change in computing and communication technologies creates a challenge in developing and deploying EDGAR. The SEC must try to ensure that EDGAR is designed in a way that can accommodate future technologies and will not constrain the types of uses analysts and filers can make of the system. To the extent the SEC tries to minimize costs by incorporating commercial off-the-shelf technology into EDGAR, greater efforts will need to be made to build flexibility into the system.

SYSTEM ARCHITECTURE

The EDGAR system currently follows a traditional system architecture, in which documents are submitted to a single, centralized location and are disseminated from a single distributor that taps into the central database. In light of the recent growth in distributed, networked computing, the SEC has expressed interest in evaluating alternative architectures for the EDGAR system that might take a more distributed form, perhaps even using the Internet for information gathering and dissemination. Such a change could allow companies to post documents on their own servers rather than transmit them to the SEC. In order to explore these issues, the SEC asked meeting participants to discuss the technical feasibility of alternative architectures for EDGAR and the relative merits of centralized versus distributed architectures.

Meeting participants agreed that existing technology could support any of a wide range of architectures for EDGAR. These range from strongly centralized systems, like the one in use today, to highly decentralized systems in which filers post documents on their own World Wide Web servers linked to the Internet. In between lie hybrid architectures that have varying degrees of centralization and distribution, such as systems in which companies would file documents with a limited number of repositories in addition to the SEC. Each of these architectures offers a different set of capabilities in terms of cost, accessibility, and system response time, and each creates a unique set of challenges for the SEC. Participants suggested that selection of an architecture for EDGAR may be driven less by technological constraints than by the capabilities the SEC wants to build into the system and the resources the SEC can allocate to it.

In general, distributed architectures would allow the SEC to share responsibility for operating and maintaining EDGAR with a number of private sector actors, but would give the SEC less control over the system. For example, in a distributed architecture, system reliability hinges on the reliability of each node in the system as well as the network connecting them-properties that are controlled by independent systems managers at each of the nodes. Problems with any of the nodes-or with the network-can degrade system performance, but typically with localized effects (i.e., the whole system will not necessarily crash, nor will all system functions necessarily be lost). In a centralized system, reliability is determined more by the reliability of the central hub, which can be managed by a more centralized authority. Problems with the central hub can, however, have more widespread effects on system availability, terminating more functions and locking out a greater number of users.

Other tradeoffs exist between distributed and centralized computer systems. For example, while distributed systems can make large-scale database searches more difficult by requiring users to direct searches across a large number of sites or construct elaborate directory structures, they can help circumvent the potential bottlenecks caused by centralized systems that direct searches through a single node. Of course, such bottlenecks can be overcome through the use of additional hardware or second-tier suppliers, but with added cost and complexity.

Many participants agreed that some degree of centralization may be necessary in the EDGAR system, if for no other reason than to act as an official archive of information. In a fully distributed system, for example, archival information could be lost if a company went out of business and took down its server. Further, the prospect of searching across some 14,000 filer sites for information is daunting for analysts. Many analysts lack the software to perform searches across distributed sites. Similarly, issues of bandwidth and system availability could become problematic: the number of searches performed could increase as the square of the number of sites on the system, since each site could initiate searches of all other sites. A centralized system would minimize this problem by limiting the number of independent sites that would need to be searched.

Some of the advantages of both centralized and distributed systems could be gained by using hybrid system architectures. For example, by using a "logically" centralized system, information could be physically distributed across a large number of sites, but accessed through a central computer that would direct queries to the appropriate site. This approach could facilitate database searches by allowing analysts to direct their searches toward a single site, while actual searching and retrieval would be distributed among a number of sites. Another possibility would be to use a centralized SEC server as an adjunct to a more distributed architecture. The centralized server could provide system backup and archiving capabilities and could facilitate database searches. Other combinations could also help compensate for some of the disadvantages of either a strongly centralized or strongly distributed architecture.

SYSTEM AND DATA SECURITY

Security considerations are central to the EDGAR redesign process. As SEC Commissioner Wallman stated, one of EDGAR's functions is to ensure the confidentiality and integrity of SEC filings by controlling

who can access the system and submit or read information and by ensuring that information is entered or changed only in a specified and authorized manner. For a more complete description of computer security, see Computer Science and Telecommunications Board, *Computers at Risk: Safe Computing in the Information Age* (Washington, DC: National Academy Press, 1991). Though the SEC views U.S. securities laws as a potential deterrent to an attack on EDGAR and abuse of the information contained in the system, meeting participants suggested that additional mechanisms may be necessary to bolster security. As several participants noted, information can be changed relatively easily as it is being transmitted across computer/communications networks from one site to another, or while it is stored in computerized databases. Such manipulation does not happen inadvertently, however; it must be done willingly.

Potential threats to EDGAR-and the vulnerability of the system to these threats-will depend partly on the architecture of the system. As with system reliability, many security tradeoffs exist between centralized and distributed architectures. While centralized architectures create a single target for attacks that can jeopardize the viability of the entire system and the data stored on it, they also allow security measures and administration to be centralized. Distributed systems, in contrast, can help prevent security problems at a single site from affecting the entire system; however security is highly dependent on the way in which security is administered at each of the distributed sites. In either case, mechanisms will need to be put in place to maintain confidentiality and data integrity.

Confidentiality, Authentication, and Access Control

The existing EDGAR system relies upon user authentication techniques to control access to the system and ensure confidentiality. The system uses four types of codes to make certain that access to the system by filers is authorized: 1) a Central Index Key (CIK) code, which serves as a unique identifier for each filer; 2) a CIK Confirmation Code (CCC) for filer authentication; 3) a user password (PW) for login and for modifying the CCC; and 4) a password modification authorization code (PMAC) for modifying the password. CIK codes are assigned by the SEC and are publicly available. CCC and PW codes are initially assigned by the SEC, but can be modified by filers. User passwords expire after 12 months. PMAC codes are assigned by the SEC and can be changed only by submitting a new Form ID to the SEC. CCC, PW, and PMAC codes each consist of eight-character strings containing at least one number and one special character (such as @, #, \$, or *). All three are case-sensitive.

This system does not allow the SEC to identify individual users logged onto the system, but it can identify their corporate affiliations. SEC representatives believe that the system has performed well to date: while there have been occasional mismatches between CIKs and CCCs, the SEC does not know of a case in which a file has been submitted to EDGAR from an improper or unknown location.

Despite this record, several meeting participants cautioned the SEC to examine system security concerns more closely. They pointed out that passwords are limited in their ability to prevent access by unauthorized users and can be circumvented or compromised in a number of ways. Though the SEC is unaware of any instances in which the system has been spoofed or broken into, potential attackers may already have "sniffed" the system to intercept user passwords. As more filers and analysts begin to make use of EDGAR, authentication problems could increase, as could opportunities for misusing the system. Meeting participants suggested that more formal risk analysis using Tiger Teams may be necessary to uncover the vulnerabilities of the system so that proper security measures can be developed.

Data Integrity

Currently, the SEC has few mechanisms for ensuring the integrity of EDGAR data. The SEC performs limited review when filings are received to make sure they comply with certain disclosure requirements, but it cannot verify the accuracy of the information submitted. Filers can request that the SEC send an acknowledgment of receipt, along with an accession number, via regular mail or via CompuServe electronic mail if filers have a CompuServe account. The SEC has contracted with CompuServe to provide E-mail and bulletin board services that users can access through EDGARLink software. Information on subscribing to the EDGAR E-mail service is included in an information packet that CompuServe sends to each registrant or third-party filer who receives EDGAR access codes. They can also ask for paper or electronic copies of the documents to check their accuracy, but such checks are not mandatory. The SEC hopes that companies will take the initiative to correct filings they see that contain incorrect information. On the dissemination end, the SEC has no specific mechanisms for guaranteeing the validity of information distributed by LEXIS-NEXIS or other value-added resellers of EDGAR information. So far, neither LEXIS-NEXIS nor its subscribers have asked for guarantees that the information they receive is valid or authentic.

Meeting participants noted that digital signatures would provide one means of helping to ensure the authenticity of EDGAR filings on both the input and dissemination ends of the system, but recognized that they provide only a partial solution to the problem. Digital signatures can be effective only if both senders and recipients of information use and check the signature, something a few participants doubted EDGAR users would do. Moreover, the way in which value-added resellers cut up and repackage EDGAR filings to suit the needs of different users would defeat the purpose of digital signatures. Digital signatures can verify the authenticity of the documents to which they are attached, but cannot vouch for information abstracted from them. Digital signatures may, therefore, be useful in the dissemination of entire filing documents (via Internet or LEXIS-NEXIS, for example), but would not necessarily help users of repackaged information verify the authenticity of the data they receive.

Several meeting participants suggested that more formal analyses may be necessary to determine whether numerous small changes to data in SEC filings could induce significant changes in the market. Though the SEC believes a linear correlation exists between the degree of data manipulation and its effects on the market, the subject has not been extensively studied. It is possible that malicious parties could surreptitiously enter the EDGAR system and change data in consistent ways that are hard to detect, but would have more noticeable effects on the market. As one participant commented, finding such vulnerabilities and learning how best to exploit them are the objectives of so-called "information warfare."

DATA STANDARDS

Data standards have become a major issue with regard to EDGAR. The SEC currently requires filers to submit documents in an ASCII text-only format and to tag certain data items using SGML to facilitate machine processing. Filers view these mandates as an unnecessary burden because they must convert documents to ASCII from the word processing formats in which they are written; remove control codes for document style and formatting such as pagination, footnotes, boldface, and underlining; eliminate charts, graphs, and photographs from their documents and attempt to summarize them in the text; and manually insert SGML tags into the text to identify portions of the document and information describing it. Because most filers prepare documents with a standard word-processing system and wish to incorporate graphics into documents they send to potential investors, they often end up producing two documents: an ASCII document for submission to the SEC, and an enhanced document, in a word-processing format, for distribution to the investment community.

The SEC also has concerns about data formats. While recognizing the benefit of ASCII and SGML in facilitating machine processing of EDGAR documents, some SEC representatives do not believe they can fully discharge their duty to review investment-related information if they can see only the textual information that has been abstracted from filers' documents (e.g., without graphics). At the same time, they do not want to restrict the ability of companies to experiment with new ways of presenting data by limiting the types of information companies can incorporate into their documents. Meeting participants attempted to outline ways in which the SEC could address these sometimes conflicting goals. The options they generated ranged from suggestions for specific data formats, to multiple document solutions, to architectural innovations.

Alternative Formats

One option would be for the SEC to accept documents in the word-processing formats in which they are written and accept graphics (as many filers have suggested). Such a system would relieve filers from having to translate their documents into ASCII and would allow graphics to be included in the filings. While technically feasible, this solution has several drawbacks. To facilitate database searching and the eventual hand-off of older filings to the National Archives and Records Administration, the SEC will need to convert EDGAR documents into a standard format. The SEC retains disclosure filings for varying periods of time, extending up to 30 years for some documents, before sending them to the National Archives and Records Administration for long-term storage. While conversion programs can be developed and have been developed for most major word processors, keeping up with the continued evolution of word-processing packages and upgrading conversion tools would impose an additional burden on the SEC, one it is not necessarily equipped to handle. In effect, this option would shift the burden of conversion from the filers to the SEC.

Alternatively, the SEC could define a single data format, such as a Portable Document Format, as the single standard in which all filers would be required to submit documents. The advantage of such a system is that it would keep the process of receiving, storing, and searching files simple for the SEC since they would have to handle information in one format only. This approach could reduce the burden on some filers if the SEC

adopted a common text format and a common graphics format as its standards, though filers that use other formats to create their documents would still need to convert their documents to the new format. The problem with such a solution is that most data and graphics formats change over time. Such changes would not only require the SEC to continually change its standard, but would make it more difficult to access older SEC filings.

By the end of the meeting, several participants saw an advantage in keeping ASCII as the data format for EDGAR because of the flux in data standards over time. Unlike other data standards, ASCII has not changed over time. Given the long periods of time for which the SEC must maintain EDGAR information, ASCII may therefore be the most efficient filing format for electronic documents. The primary drawback to ASCII is its inability to accommodate graphical information. Options for overcoming this deficiency include adopting a separate graphics format for accepting charts, tables, photographs, and figures; continuing to require filers to strip graphics from their filings; or restricting companies from filing non-text items. Many SEC representatives expressed a desire to avoid using these last two alternatives because they want to ensure that SEC analysts can review all information contained in investment documents, and that companies can experiment with new ways of presenting information that might prove more useful or accessible to potential investors. Other options for overcoming the limitations of ASCII without restricting the SEC's ability to review filings or companies' ability to present relevant information would entail multiple filings or changes in the architecture of EDGAR.

Multiple Filings

Some of the concerns over data standards could be addressed by requiring firms to submit two separate documents to the SEC: one formatted for machine processing and analysis and one for human processing. The first document could be designed to meet the SEC's basic information requirements; the other could include enhancements such as graphics to make the document more readable and more attractive to potential investors. Filers could submit a basic filing to the SEC and post more extensive documentation on a company server. This would provide a centralized location for analytical assessments by the SEC and other users, as well as multiple distribution sites for information.

For example, the SEC could dictate that filers submit a scanned image of the page from an enhanced document (with graphics and other features that do not convert easily to ASCII) and a heavily tagged ASCII text document that includes all the words, numbers, and tags in the original image. This approach would: allow users to use the ASCII file for searching documents and use the scanned image for displaying information in a reader-friendly format, complete with graphics; permit tagging be optimized for computer use to facilitate searching, retrieval, and analysis of EDGAR information; and provide a limited number of formats in which the SEC, and eventually the National Archives, would collect information.

This specific implementation option poses several obstacles. Several participants noted that the character recognition capabilities of optical scanning technologies are not fully reliable. The SEC could not be assured of receiving an accurate electronic copy of the original document using existing systems. Furthermore, filers may develop new ways of presenting information that cannot be scanned. Documents that incorporate video clips or Java programs, for example, could not be scanned. Moreover, this solution may increase the burden on SEC staff by requiring them to review two sets of documents (the ASCII version and the enhanced version) if they want to make sure they have examined all information made available to investors.

Architectural Solutions

Some participants suggested that data standards issues be addressed through architectural changes in EDGAR's design. One suggestion that received considerable attention was for the SEC to allow either a single intermediary or a set of designated intermediaries to accept filings from users and translate them into formats that the SEC can read. In this plan, filers would submit documents to the intermediaries in any format, and the intermediaries would reformat the data so that the SEC would have a core of information that it could search and machine process. The intermediary would also be required to provide the SEC with a means of viewing other non-standard (i.e., non-ASCII) information in the original document. As such, the intermediary would make the full content of the issuers' documents available to the SEC. Intermediaries could perform additional reformatting and repackaging to provide the information to various subscribers who would pay for the service.

This approach offers several advantages. First, it would provide for greater privatization of both the input and output functions of EDGAR, in keeping with the desire of Congress. Second, it would relieve issuing companies of the burden of preparing two separate documents, a basic one for the SEC and a more

complex one for investors. Instead, it gives responsibility for document preparation to the intermediary. Third, it allows the SEC to review all the information contained in the original document, without limiting the ability of companies to include enhanced information in their filings. Finally, this scheme shifts the burden of developing or procuring new conversion software to the intermediaries, rather than the SEC.

Several issues must be resolved before such an approach could be considered. For instance, it is not yet clear whether companies could make money as intermediaries, or whether they could gain some other competitive advantage that would entice them to enter the market. Nor are all the technical challenges resolved. Could firms find cost-effective ways of putting all the information contained in filings into a form SEC users can view? What would be the implications for the SEC in terms of capacity, connectivity, and security? Similarly, the SEC, Congress, and the investment community must determine whether it is desirable for private companies to make money performing an intermediary service. Some meeting participants questioned whether the SEC should set up a system that requires filers to pay some intermediary in order to submit a filing, rather than submitting it themselves. Others questioned the wisdom of establishing a requirement for intermediaries using technologies widely recognized as enabling the demise of intermediated access to information. Still others expressed concerns over the fact that this solution creates a group of intermediaries that would have privileged access to EDGAR filing information.

An alternative architectural solution would allow companies to post basic EDGAR filings on corporate World Wide Web sites that include additional information about the company and are designed for investor relations purposes. The Web site could include graphics, sound, video imagery, and Java programs, or any other type of information the company would like to present to investors, but cannot include in a standard EDGAR filing. Filers could then notify the SEC that a new document is available, and the SEC could download it for review, analysis, and archiving. Users of investment data could download information directly from the Web site, or the SEC could maintain a separate dissemination system to distribute the basic EDGAR filings. The primary advantage of such an approach is that it would allow the SEC to view all information contained in official SEC filings and investor documents, but would allow filers to place additional information adjacent to-rather than inside-the document, overcoming their current objections to ASCII. This approach, however, presupposes widespread use of the World Wide Web by investors and filers alike, and might place an additional burden on SEC staff by requiring the agency to download each filing individually from corporate Web pages. Long-term availability of Web-based documents could also become an issue as standards evolve for posting information on the Web and the capabilities of Web browsers change.

DATA STORAGE AND ANALYSIS

In redesigning EDGAR, the SEC is interested in identifying possible alternatives to the existing LEXIS-NEXIS data management service, including the use of commercial off-the-shelf systems for text search and retrieval. The SEC has paid LEXIS-NEXIS \$13 million over the past 8 years for database management services; nevertheless, approximately 80 percent of all searches occur within the SEC on SEC computers. SEC workers have noted that other organizations have developed search and retrieval capabilities that seem to rival those of LEXIS-NEXIS and could represent an alternative to the existing system. For example, the SEC reports that one company, Global Securities, offers a product over Internet called "Live EDGAR" that can conduct sophisticated searches of filings with remarkable speed.

Meeting participants knew of few commercial products available today that could handle large database management tasks on the scale of EDGAR. Note that participants were selected for their expertise in fundamental computing and communications technologies, rather than current product offerings. According to participants, the technology to perform rapid search and retrieval is currently in the research or advanced development stage, but has not yet been translated into new products. As a result, the cost of such systems is high, as service providers charge users for the cost of bridging the gap between research and development and product development.

Many meeting participants expect to see growing commercial interest in the development of database search engines as the technical capabilities of system components such as memory and processors continue to improve. With older technology, systems developers had to invest considerable resources in the development of efficient search algorithms that would minimize the number of disk accesses performed during a single search. Limited disk space and 20- to 100-millisecond access times meant that searches of 1,000 or 10,000 records could take substantial amounts of time. As the size of disk storage has increased and disk access times have fallen, the need to develop efficient search algorithms has decreased, reducing the cost and complexity of development efforts. Systems developers can now buy a gigabyte of RAM to hold search indexes and use 10 gigabyte disks for storage. With affordable processors that run 200 million

instructions per second and disks that boast microsecond and nanosecond access times, searches can be conducted much faster, even if the software directing the search is less efficient.

Demand for such systems is also rising. Many organizations other than the SEC have a strong interest in developing or procuring search engines for large databases. Therefore, it is unlikely that the SEC will need to develop-or pay for the development of-its own search engine, but may soon be able to buy an off-the-shelf system or a system that will require only minor modification to suit its needs. In fact, the fast pace of technological development suggests that the capabilities of commercially available search engines would likely surpass that of any "home-grown" system the SEC could develop. Vendors who are far along the path to developing a suitable search engine should be able to adapt the system to the SEC's needs in a period of months; those attempting to sell a developmental system will likely need many months or years to fully implement the system.

Most participants agreed that the SEC's choice of search engines is not constrained by the operating system the SEC uses for its in-house computers. Nor is it clear that any one operating system has an intrinsic advantage over another for the types of applications the SEC runs. The most relevant criterion for selecting an operating system might instead be longevity, though it is hard to estimate the length of time a developer will support and extend a particular operating system in such a fast-moving industry. As with other elements of the system, the SEC will periodically have to change to new operating systems, based on advances in technical capability and reductions in cost. The SEC needs a process and an architecture that allows such change. Portability between operating systems and modularity may therefore be factors to consider in the redesign of EDGAR.

DISSEMINATION

The redesign of EDGAR may also provide the SEC with an opportunity to revamp the dissemination subsystem in ways that would increase competition among distributors of EDGAR information. Many people in Congress, the SEC, and the investment community believe that the SEC's contract with LEXIS-NEXIS gives the company an unnecessary monopoly over real-time dissemination, and hope that competition could serve as a means of driving improvements in the system and helping to reduce costs.

Several meeting participants agreed that long-term exclusive contracts like the one currently in place with LEXIS-NEXIS can be incompatible with the desire to promote innovation and encourage the rapid adoption of new technology. Many felt that EDGAR would be able to take advantage of new technology faster and more efficiently if dissemination were structured as an open market that any viable competitor could enter. Participants suggested that by making raw EDGAR data more widely available and, by lowering the cost of acquiring raw EDGAR data, the SEC might encourage new entrants to enter the dissemination market and find new ways to add value.

The SEC has a range of options for encouraging greater competition in the dissemination of EDGAR information. The least radical approach would be to contract with several wholesale distributors to provide real-time EDGAR data to the investment community. These distributors would compete with one another (and LEXIS-NEXIS) on the basis of price and performance. This strategy would create some incentives for wholesalers to improve the dissemination process and would allow the SEC to take full advantage of its current investments in the EDGAR system. It would, however, entail the development of redundant distribution systems by the additional contractors. The SEC might also have to upgrade its dissemination system in order to channel EDGAR information to multiple wholesalers.

A more radical alternative would be to construct a system of private escrow agents who would both collect information from companies and distribute it to the investment community. Filers could send documents to both the SEC and a limited number of private escrow agents who would bid for filers' business and, in turn, sell the information to third-party users and resellers. This type of system would allow greater privatization of EDGAR, but maintain a central role for the SEC in accepting and archiving filings, checking them for completeness and consistency, and enforcing disclosure requirements. The SEC database could also be used as a centralized source for large-scale searches and would provide a backup of the information sent to escrow agents. Such a system would, however, require substantive changes in EDGAR's architecture to incorporate escrow agents into the role of receiving filings. Digital signatures may also be needed to ensure that filers transmit the same document to both the SEC and the escrow agent(s).

Alternatively, the SEC could opt to use the Internet as its primary mechanism for dissemination of EDGAR information. In order to make Internet dissemination more attractive to professional investors who currently

subscribe to LEXIS-NEXIS, meeting participants suggested that the SEC eliminate the 24-hour delay on information posted to the World Wide Web. By using the Internet, the SEC could make EDGAR documents available at a low marginal cost to virtually all investors (or potential investors) with access to Internet, and, in effect, turn the SEC's Web server into the central repository for raw EDGAR data. While this structure would clearly raise a new set of security issues that would need to be addressed, it would also allow the SEC to use the Internet to collect EDGAR information from filers. Filers could submit documents containing raw data directly to the SEC Web site. The SEC would use these filings to support its internal analysis, as could other users who require only raw data-most likely casual, low-volume users. Value-added resellers (VARs) could also download data from the Web site and reformat it to add additional information or features, changing the formatting as necessary, and perhaps purchasing additional information from the issuer with which to enhance the document. VARs could even ask issuers to provide them, for a fee, with information formatted in specific ways. VARs could reformat information in different ways for different users, tailoring it to their specific needs. This system would use the one Web site to satisfy the needs of the SEC, other users of raw EDGAR information, and VARs.

Some participants believed that fundamental limitations on the Internet's reliability might restrict the SEC's ability to rely solely on the Internet for dissemination. Because of the rapid rate at which Internet usage has grown, sites are often down and access can be slow. As a result, Internet cannot currently guarantee service with 100-percent reliability. Some participants, therefore, cautioned against putting critical "mission traffic" on the Internet, noting that while the Internet may be fine for the casual user, it is not necessarily reliable enough for many critical business applications.

Other participants, however, believe that availability and quality limitations can be overcome. To ensure that hardware is available to service requests, for example, they suggested that the SEC keep its server on-line for longer periods of time and build in sufficient bandwidth-perhaps by using redundant servers-to satisfy expected demand. They also speculated that companies might soon be able to "buy" greater reliability and quality of service by paying a premium to their service provider to make sure additional paths exist between them and the nearest SEC node.

In addition to reliability concerns, a number of business and economic issues need to be resolved. Because all potential competitors would have equal access to raw EDGAR data, it is not clear how individual firms would establish a competitive advantage in the marketplace. The primary advantage LEXIS-NEXIS now has is its real-time access to data. Without this advantage, firms will have to devise new ways of adding value to the raw EDGAR filings, much as the nine subscribers to the existing LEXIS-NEXIS service seem to have done. By aggregating data in useful ways, presenting it in more readable formats, or conducting some simple sorts of simple analysis, firms may be able to attract customers and develop new businesses. Additional distinctions may also be achievable in terms of quality or reliability of service. At this point in time, however, it is not clear that potential profit margins will justify the cost of entering the market and repackaging the EDGAR information. Nor can one estimate the number of firms that could compete viably in this market. Such considerations would require more analysis and input from the business community.

MANAGING FUTURE TECHNOLOGICAL CHANGE

Meeting participants agreed that, in general, the redesign of EDGAR does not appear to present any insurmountable technical challenges. Current technology can support different architectures under consideration-or proposed during the meeting-for a redesigned EDGAR system. The greatest technical challenge lies in designing a system that can accommodate change and incorporate future technologies. As one participant remarked, the greatest danger to the success of the EDGAR modernization effort is that the SEC will design a good system with existing technology that will not be able to incorporate technological change.

Meeting participants generally agreed that the success of the EDGAR redesign may hinge on the SEC's ability to establish a formal mechanism for identifying new technologies and incorporating them into the system over time. New technologies will undoubtedly change the environment in which EDGAR operates. For example, documents are likely to become more active entities that include video clips or embedded programs to perform certain calculations; wireless systems could eventually replace copper wire and optical fiber in some communications and networking applications. Predicting such trends and the implications they may have for EDGAR is difficult, however. Technology changes quickly enough that new products and capabilities will likely be developed in the time it takes the SEC to evaluate proposals and implement a new EDGAR system. The key element for planning, therefore, is not necessarily to know what new technologies will be available in the next five to ten years, but to develop a management plan-and allocate sufficient

funds-for adopting new technologies as they become available.

Development of the Internet may provide a useful analogy. Many experts claim the Internet has been successful not because it was planned and designed as a good system from the start, but because a process was developed that allowed the system to evolve. Designers created a system of protocols that gives each packet of information passed along the Internet all the information it needs to get to its destination. The network across which packets move and the technologies that perform the necessary switching and routing tasks can change without affecting the underlying ability of the packets to navigate the network.

Meeting participants suggested that EDGAR be designed with an open architecture that would allow the system sufficient flexibility to meet both its tactical needs and long-term strategic needs without another major redesign. By building EDGAR upon a minimal set of system requirements, rather than a specific set of technologies, the system may be enabled to evolve as underlying technologies change. As one participant commented, the key idea is one of abstraction, in which requirements are distinguished from their implementation, and the system is designed around the requirements. Rather than thinking about specific tags put into an ASCII file, for example, the SEC could think about dates, company names, and identifiers as abstract entities that can be represented in any number of ways, using any of a number of different formats or tagging schemes. Following this approach, the SEC may be able to develop a system with an internal organization that suits a variety of changing requirements and allows changes to occur at the interface between subsystems, rather than disrupting the design of the entire the system.

POSSIBILITIES FOR FUTURE CSTB STUDY

The CSTB brainstorming session was intrinsically limited in its scope, size, and period of preparation. Nevertheless, as this summary demonstrates, the SEC has many issues to resolve with regard to the modernization of EDGAR. While some of these issues need to be dealt with before embarking on a modernization program, others will require continued thought and analysis as the program unfolds. For example, the SEC must evaluate different system architectures that span a spectrum from highly centralized to widely distributed systems and that might incorporate intermediaries on the input and/or output side of the system. It must consider the effects of these architectures on system security, functionality, and reliability, and understand the ways in which decisions made today will affect the ability of the system to evolve over time. Additional evaluation of technologies or software packages for database management is also needed. Conducting such analysis will require a broad range of expertise to help understand the technical, business, and policy issues associated with EDGAR.

CSTB has conducted several previous studies to assist federal agencies in the redesign and development of their information systems. This work includes studies for the Social Security Administration and the Internal Revenue Service on systems modernization. Computer Science and Telecommunications Board, *Continued Review of the Tax Systems Modernization of the Internal Revenue Service* (Washington, DC: National Academy Press, 1996); Computer Science and Telecommunications Board, *Review of the Tax Systems Modernization of the Internal Revenue Service* (Washington, DC: National Academy Press, 1992); Computer Science and Telecommunications Board, *Elements of Systems Modernization for the Social Security Administration* (Washington, DC: National Academy Press, 1991); Computer Science and Telecommunications Board, *Systems Modernization and the Strategic Plans of the Social Security Administration* (Washington, DC: National Academy Press, 1990). In both cases, CSTB not only provided a neutral meeting ground for different stakeholders in the analysis, but was able to help the agencies delineate the goals of their modernization efforts and develop implementation and management strategies for meeting them. The longer duration these studies (typically 12 to 18 months) allowed CSTB to maintain a sustained dialogue between an expert committee from industry and academia and representatives of the federal agencies, which provided sufficient time for the committee to more fully understand the missions and functions of the agencies, the ways in which they worked, and the complaints they had regarding information systems. With this understanding, the project committees were able to address questions relevant to the technology and technology management issues the agencies faced.

Such capabilities could benefit the SEC as it moves to redesign EDGAR. While the SEC intends to release a Request for Proposals (RFP) in early to mid-1996, implementation of the EDGAR system will occur over a longer time frame that would allow longer-term interaction with CSTB. Further evolution of the system could also benefit from such an effort. As the brainstorming session demonstrated, the SEC faces several critical decisions concerning system architectures and data requirements. Many options are possible, yet each architecture implies different system capabilities, constraints, and degrees of security. Each places different

requirements on the SEC, filers, and the investment community. Understanding the ways in which these elements interact is a task that will require the kinds of expert advice CSTB can provide, through its unique capability to convene experts from industry and academia for the purpose of advising federal entities on matters relating to computing and communications technologies.

APPENDIX A

INVITED MEETING PARTICIPANTS

Dr. William Wulf, Chair

Dr. Wulf is chairman of the Computer Science and Telecommunications Board
University of Virginia

Mr. James Branscome

Standard & Poor's

Ms. Cecilia Bryan

Alexander Brown and Sons, Inc.

Mr. Robert Collet

Commercial Internet eXchange

Mr. Joel Jacobs

MITRE Corp.

Mr. Richard Liebhaber

MCI (retired)

Mr. Steven Lipner

Trusted Information System, Inc.

Dr. Jerome Saltzer

Massachusetts Institute of Technology

Dr. Karen Sollins

Massachusetts Institute of Technology

Mr. Donald Sortor

Motorola, Inc.

Mr. Robert Swithers

Citicorp Global Information Network

APPENDIX B

GOVERNMENT PARTICIPANTS AND OBSERVERS

Participants:

Mr. John Avery

Office of General Counsel

U.S. Securities and Exchange Commission

Mr. Nicholas Balamaci

Office of the Chairman

U.S. Securities and Exchange Commission

Mr. Michael Bartell

Office of Information Technology
U.S. Securities and Exchange Commission

Mr. David Copenhafer
Office of Program Management and Analysis
U.S. Securities and Exchange Commission

Ms. Mauri Osheroff
Division of Corporation Finance
U.S. Securities and Exchange Commission

Mr. Michael Schlein
Office of the Chairman
U.S. Securities and Exchange Commission

Mr. Anthony Vertuno
Division of Investment Management
U.S. Securities and Exchange Commission

Mr. Steven Wallman
Commissioner
U.S. Securities and Exchange Commission

Dr. W. Duncan Wood
Office of Congressman Daniel Frisa
U.S. Congress

Observers:

Mr. R. William Barton
Office of General Counsel
U.S. Securities and Exchange Commission

Mr. George Brown
Office of General Counsel
U.S. Securities and Exchange Commission

Mr. Barry Miller
Division of Investment Management
U.S. Securities and Exchange Commission

Mr. Bradley Paulson
Office of Commissioner Wallman
U.S. Securities and Exchange Commission

Mr. George Pospolita
Office of Information Technology
U.S. Securities and Exchange Commission

Ms. Sylvia Reis
Office of Corporation Finance
U.S. Securities and Exchange Commission

Mr. Charles Woods
Office of Information Technology
U.S. Securities and Exchange Commission