**Final Report: NSF Award 9319331**
**Internet Access to Large Government Data Archives:**
**The Direct EDGAR Access System.**

Ajit Kambil
Assistant Professor of Information Systems
New York University Stern School of Business
44 West 4th Street, Suite 9-82
NY, NY 10012
akambil@stern.nyu.edu

July 29, 1996

**Abstract**

The purpose of this project was to demonstrate the feasibility and effectiveness of disseminating information from large government databases over the Internet. Using the SEC EDGAR (Electronic Data Gathering, Analysis and Retrieval) text database as an example, the project successfully demonstrated the feasibility of using the Internet to effectively disseminate large government databases to the public. The project educated a large number of users to the potential of the Internet and demonstrated its usefulness for communications and government data dissemination. In October 1995, the Securities and Exchange Commission took over the core systems developed in this project and currently uses it to disseminate EDGAR documents to the public.

The project demonstrated dissemination using electronic mail, file transfer protocol (ftp), gopher and world wide web (WWW) as primary methods. A number of programs were developed to create specialized searches on the data and provide more useful user interfaces. During the course of the project over 3 million documents were transferred to the public by October 1995. Currently the NYU server receives 15,000 to 20,000 hits per day generating requests for filings. With continued private sector funding the project continues to develop new tools for improving free public access to the EDGAR database.

**1. Introduction: What is EDGAR?**

EDGAR is the U.S. Securities and Exchange Commission's (SEC) Electronic Data Gathering, Analysis and Retrieval system. This system is designed to electronically receive various corporate disclosures of publicly traded companies. These filings include annual and quarterly reports, proxy statements, current reports (8K) and various change in ownership reports. Since April 1993, companies have been gradually phased into the system. Since May 1996, all publicly traded companies must file electronically via the EDGAR system.

EDGAR filings consist of text files with limited SGML type tags. Tags at the beginning of documents help to identify: the company name, filing type, the CIK number (the SEC's unique numerical identifier for a filing company or entity) and other limited information. Other tags within the document are typically used to support formatting such as page numbers and tables. Thus EDGAR documents form a **largely unstructured** text database with very little tagging to identify the useful contents of the data set. The major exception is the new financial data schedule introduced in 1995 which for the first time clearly associates clear identifier tags with various key financial values disclosed in the annual and quarterly financial statements. However there is still little tagging of text content. In 1995, the EDGAR system received over 10 million pages of text. In 1996, with all companies filing electronically, we estimate that 70 GB will be required, for storage indicating the large size of the database.

EDGAR is a critical database for business professionals and for investors. Its filings help investors monitor the performance and actions of management, as well as understand management's assessment of key prospects and risks confronted by the firm. The filings also disclose information useful for monitoring competitor performance and actions, and filings such as those disclosing 5% ownership of stocks, highlight key investment patterns or takeover interests.

Thus corporate disclosures included in the EDGAR database provide key information to participants in U.S. Securities markets helping to increase the transparency and liquidity of these markets.

**2. Traditional Dissemination of EDGAR Data**

Prior to this project, data from EDGAR was disseminated to the public through five nationwide Securities and Exchange Commission reading rooms, and by private companies that provide on-line, CD-ROM or paper versions of EDGAR data at a cost. Specifically, Mead Data Central had the primary dissemination rights to the EDGAR data. Mead disseminated the raw filings via: direct connection to customers (Level 1 broadcast service using leased X.25 or TCP/IP services), an overnight tape service, and on its Lexis Nexis service. Fees to the overnight services were approximately $75,000 for 1994. Subscription to Lexis Nexis varied substantially in costs. Third party vendors would purchase one of these data feeds to create value added products and specialized databases on the data.

End users could purchase the data from Mead Data's Lexis service or from other third party vendors such as Disclosure, Moody's etc. Value added services offered by third parties include mailing or fax to customers of well formatted individual filings for prices of $30 or more per document. Users could also purchase CD-ROM versions of the filings, or get on-line access to filings. In all these options the individual investor incurred substantial costs or delays in retrieving filings. Thus timely and quality access to this data was primarily limited to professionals in organizations that could afford to purchase this data. Access to filings by the broader groups of individual or foreign investors, or other parties interested in using the data was limited. This also increased the costs to the investor relations group of the disclosing company who typically had to mail paper versions of the disclosure on request to its shareholders.

**3. The Base Internet EDGAR system**

The basic EDGAR system initially consisted of a server hosting the EDGAR files on the Internet. This server was maintained by Internet Multicasting Services (IMS) and enabled electronic mail, ftp, gopher and World Wide Web (WWW) access to files. The EDGAR files were extracted from tapes purchased from Mead Data. These files were parsed into a main filing containing the full text of the filing, as well as a small header file containing basic information on the filing such as: filing type, company addresses and contact information, etc. In addition a digital signature was added to the filings using the RIPEM protocol. This enables users to verify or authenticate that the document was not changed in the transfer to the user. Users of the WWW would have to

use the ftp protocol for retrieving and saving the filings for successful authentication.

The files were then stored as text files into specific subdirectories denoted by the Central Index Key (CIK) in the "data directory". The CIK is the SEC's unique identifier for each company. For the most part all files were correctly filed into CIK directory of the filing company. In some cases where two companies were noted in the filing header such as the subject company (in a 5% acquisition) and different filing company (such as mutual fund), the actual text file was located in the CIK directory of the subject company. However, the indexing system would identify the filing under searches for either the filing or subject companies. Thus from a search perspective the storage of some files in directories of the subject vs. those of the filing company is not particularly significant.

On a daily basis three types of indexes were created on the data. These included an index sorted by company name, one sorted by form type and the master index sorted by the central index key. On a quarterly basis the daily indexes were consolidated into a quarterly index. These index files allow different users to construct their own search tools on the EDGAR data stored on the IMS and now the SEC servers. On a daily basis specialized indexes were also updated to support user searches for specific filings.

Initially electronic mail and ftp access to the data were implemented. Electronic mail and ftp accesses required the user submit a request retrieving the index files or portion of the index file. Once the required file was identified a second electronic mail or ftp request could be made to retrieve the

text file. This was rather cumbersome for users. Electronic mail processing was also resource intensive, and users often found the files truncated to 32 or 64 kbytes.

Next gopher and World Wide Web access to data were implemented. These systems provided more simplified access to EDGAR data. Essentially a WAIS index was set up on the indexes previously defined. The WAIS index permitted users to search for any corporate filing by entering a portion of the company name. The WAIS index was later extended to include all contents of the header files. This extended the categories of search to include portions of the address or filing type. However, variations in separators and formatting in the source document could confuse the search engine into not returning a result.

This base EDGAR system, along with some of the searches developed at NYU was transferred to the Securities and Exchange Commission by Internet Multicasting Services in September and October 1995. This became the core technology underlying the SEC's own web site and current EDGAR service providing ftp, gopher and WWW access to the data and fulfilling one of the objectives of the original project of transferring technology back to the government for making EDGAR data widely accessible to the public.

**4. Enhanced Access and Research at NYU**

To add to the development of the base system, we undertook a number of development activities at New York University.

At the beginning of the project in 1993 and early 1994 we conducted a number of focus groups as well as interviews with different types of users of EDGAR data. These included: analysts, lawyers, accountants, librarians, individual investors, faculty and students. Except for lawyers and analysts at major firms, many of the above users did not have timely and convenient access to the data[1]. While we identified major use categories of the data, we found that users were not sufficiently familiar with the Internet to envision or convey interesting applications on the data. Thus while these sessions educated users about the Internet, they were less useful from an application design perspective.

Initially most users were most concerned with formatting and printing issues and less with retrieval and applications on the text. In essence users were so used to existing formats of the data the focus group meeting did not immediately result in their imagining higher level searches. We learned quickly that users wanted to be presented the data in terms they were most familiar with. These included searching for documents using exchange ticker symbols, or company names in the format that they are most familiar with (for example, IBM, not International Business Machines). Next, through supplementary questionnaires and on-line feedback mechanisms we were able to learn substantially more about users and their needs during the project. First, the most important types of documents to users were: the 10-K annual report, 10-Q quarterly report, 8-K current reports, DEF 14A (proxy statements), registrations, and acquisition and ownership change reports (Schedule 13 forms). Within these documents specific areas were more

[1]Indeed even those who had access to EDGAR data at their firms found Internet access useful for work from remote locations or from home.

important than others such as the management discussion and analysis, or notice to shareholders. In addition many users requested Forms 3,4, 5 and 144 -- security ownership and transactions filed by corporate insiders and notice of proposed sale of securities. These forms can now be filed voluntarily over EDGAR but were not available during the period of this project.

Users use the EDGAR database for a wide variety of purposes. Investors use the data for prospective information as presented in management discussion and analysis sections, or for monitoring the use of assets and performance of management. Beyond investors, numerous others use the EDGAR data. For example, EDGAR data may be used to monitor competition, the actions of firms in regard to their investment, executive compensation and officers, location of activities, potential liabilities, and even the impacts of unusual events such as weather that affects the performance of the firm. Indeed managers and employees often track the filings of their own firm to learn more about the firm's disclosures and activities. Some of these applications are easy to serve. Other applications require full text search combined with users' deeper knowledge about the content of the data.

For example a simple application would be finding a specific filing tied to a company within a specific time period. More complex queries would include searching across many different documents to find specific pieces of information, aggregating results and summarizing findings. An example would be all fees paid to a competing investment bank in 1995. Other queries would track all appearances of a specific person in the EDGAR database. Given the constraints of our equipment and resources we were not able to give users extensive full text search, and were only able to satisfice the

simpler queries. Full text search on EDGAR would have required substantially more disk space on the back end server. Originally we had budgeted for 20GB of data. Full text search on all documents over the course of this project would have required more than double this resource.

In response to users we developed a number of search tools and interfaces to better support user searches of EDGAR. In addition, as a demonstration project we sought to illustrate a number of potential applications of WWW technologies in disseminating EDGAR data. These are discussed below:

**4.1 Search Tools**

We developed a number of search tools to support user searches of data. These include:

- Search by ticker symbol (currently with over 8000 companies)
- Search by limited fragments of a company name
- Search by industry grouping (Zacks major and minor codes)
- Search by form categories

We also provided ways of limiting searches to specific date ranges, and company and form types. These searches provide more precision than an ordinary WAIS search. Adding more structure to the query form also forces users to defining more precise queries. All the search tools were developed as CGI scripts using PERL and can be accessed from the NYU EDGAR home page: http://edgar.stern.nyu.edu

To support the searches we created specialized indexes. In addition we had to create a ticker to name and CIK code mapping utility. Due to differences between the SEC and the exchanges in naming companies, we had to reconcile differences manually. Given limited staff resources, this is costly from a system update and maintenance perspective.

**4.2 Database Integration**

As a demonstration project we also developed applications to illustrate more complex searches on the data, integrating database applications with the Web. We wrote applications to track the major equity holdings as disclosed by mutual funds across different time periods. These applications automatically extracted data from Forms 13 E and F, stored the data in a database, and compared data across different time points. The mutual fund tracking application allows investors to track significant fund holdings and their performance across time. This application also graphed the performance of equity holdings in different industry sectors.

Programs for extracting the data were written in PERL. A donated copy of the Illustra database was used for storing data and supporting the application. While Illustra is an excellent database for time series, it has a substantial learning curve. Thus maintenance of these and similar database applications was difficult for us due to turnover in student staff as they graduated or were hired by companies. Future versions of this application will use a more standard database product.

There is substantial interest in mutual fund data. Although funds have to file various reports electronically, they can do so on magnetic tape. This data does not necessarily appear on the EDGAR tapes, unless the fund files using the EDGAR system. Given the growing importance of mutual funds, the SEC should encourage all to file through a consistent method so that the data is widely available through the Internet EDGAR system.

**4.3 Automatic Extraction of and Tagging of Data from Text**

Given the large size of the database it was infeasible within the constraints of this project to full text index the data. Thus we focused on creating a corporate profile consisting of select sections of the 10-K, proxy statements, 8-K reports, and registration statements. Programs were written to automatically extract the data from multiple documents using PERL. These included the description of the business, management discussion and analysis, notice to shareholders, executive compensation, and other select portions of the database. Programs were also written to update the profile and illustrate the performance graph in the document.

Profile extraction was based on the structure of the document, standard headings and other delimiters as noted in the various SEC filers manuals. (We used the Bowne version of the filer's handbook, combined with analysis of the documents). However, automatic profile extraction was not widely successful because of substantial variations in documents of a specific type. Besides variations in formats, incorporation of key information by reference in different sections of a report made it difficult to automatically extract the information close to 100% of the time. To get close to a 100% success without

human intervention will require a substantial natural language programming effort. Our most recent testbed of profiles maintains about 150 companies for test and illustrative purposes. Profiles can be less than 300 kilobytes in size per company, in contrast to many megabytes for the key documents. For reference purposes the profiles also maintain hypertext links to the source documents so that users can verify the information, and review the original materials. This is especially important given the use of footnotes and other incorporation by reference in source documents. We continue to investigate low cost ways of automatically extracting and automatically maintaining profiles for a small subset of major companies for research purposes.

The SEC's requirement for providing a well formatted and tagged financial data schedule since 1995 has begun to improve the structure of presenting key numerical data and has facilitated their extraction. While we have extraction tools for this data we will wait till fall 1996, when all companies file financial data schedules, to provide a comprehensive database of this category of data. This will help users select companies based on revenues and other categories. Tagging more categories of data, and ensuring that filings conform to specific formats will vastly improve the extraction of useful data from the text.

**4.4 Indexing and Retrieval**

We tested a number of public domain full text index search tools including freeWAIS, Glimpse and Excite for their suitability to full text index the profiles. FreeWAIS SF was tested for suitability for both full text search and fielded searches. More recently we have been using Glimpse which was the

more efficient in terms of storage requirements. We are currently testing Excite and its suitability as a search engine.

Full text indexing increases storage requirements by at least 1.5 times the original file sizes. As mentioned earlier we did not have sufficient storage space to full text index a large and useful subset of the source documents. As storage prices fall we are now acquiring the disk to store and index a subset of the documents.

Despite the possibility of full text indexing using WAIS or similar search engines, the precision of searches is likely to decrease as the number of files increases. Instead of constructing indexes for search ex post the filing into the EDGAR database, specific tags can be inserted in the documents to catalog the document at the point of creation. One alternative to consider in the next iteration of EDGAR is the possibility of creating more meaningful header documents which include information that catalogs key events in the form.

Our research also examined ways of improving information retrieval through partial coordination techniques. As more and more people publish information on the Internet, both computerized post coordinate and full text search begins to result in unmanageably large retrieval sets and less precision. Indeed full text search can lead to spurious matches of terms out of context unless users form very well structured queries. However, cataloging the documents can be expensive and traditional precoordination using keywords for broad categories and imposing a standard and intelligent order on the keywords is also unlikely to vastly improve search. We propose partial coordination as a means of improving search and increasign precision. Partial

coordination, provides a new way of cataloguing which combines benefits of new computerized search with traditional benefits of pre-coordination to limit spurious partial matches and thereby enhance precision. This is described in the NYU Center for Research in Information Systems working paper IS-96-1: *Pre-coordination + Post-coordination = Partial Coordination.*.

**4.5 Help System and Reciprocal Links**

We collected and responded to a number of user questions, creating and maintaining a frequently asked questions list to support users. In addition we provided further resources such as the R.R Donnelley Filers' Manual and point to various financial resources on the Internet.

We also responded to as many users as we could who provided bug reports. We typically receive five to ten bug reports or user queries per day. As consequence of these feedback systems, we adapted our software to support different commonly used web browsers or access mechanisms, including lynx for low end users. As the market selects de facto standards such as the Netscape browser, our need to support variations in different client software have diminished and access to the EDGAR database has greatly improved.

**4.6 Server Redundancy**

We have implemented two servers at NYU as front ends for data access. These include a SparcLX that was the original EDGAR development machine, and a Sparc5 donated by FAME. In addition as Disclosure Inc. implemented a replica of the EDGAR back-end we now point to the Disclosure server and the SEC EDGAR database. This now provides for a fully redundant source of free EDGAR data if either the SEC or Disclosure site is down.

Besides maintaining servers we maintain access logs to the data. Currently the NYU systems receive 15,000 to 20,000 hits daily. Since users bookmark pages, this translates into approximately 5000 to 10,000 EDGAR files transferred from the SEC or free Disclosure service based on the NYU server on a daily basis. About 10 percent of requests are from foreign countries.

**4.7 Advanced Development**

With the advent of Java and other client side programming languages we have been investigating ways of creating more advanced interfaces to the EDGAR system. For the last six months our major development effort was on a Java front end to the EDGAR data as a first step to developing a new architecture for dissemination which would push more and more processing to the client side. Our basic premise underlying this effort is that both client processors and communications infrastructure and bandwidth to desktop will continue to improve dramatically. In this environment either the data or indexes to data can be moved to the client desktop. Value added processes and applications will then be performed on the indexes or data at the client side rather than at the server side. For public dissemination, this enables maintenance of low cost server solutions, while providing users with varying degrees of functionality based on the capability of their processors.

Our initial application which uses these principles is a Java interface to the NYU EDGAR site which can be accessed from our homepage: http://edgar.stern.nyu.edu/. This application provides a clean interface that uses a ticker metaphor in displaying daily filings. By clicking the mouse on

the name of a filing displayed in the ticker, the filing is directly retrieved from the SEC or Disclosure site and displayed. Users can sort the index by form type or company name, or find the filing of their choice in the daily index by entering the first few letters of the company name. The interface also provides access to various other search tools available on the NYU EDGAR server, as well as stock quote information.

A key feature of this interface is that processing of the daily index is completely done on the user's own computer. This way the NYU server is only accessed to transfer the applet and the index file on a daily basis but not for processing searches on the daily index. Next the user can select the filings he or she wants from the daily index and retrieve them directly from the SEC or Disclosure site. This reduces the server loads from a processing perspective at NYU. The applet once loaded on a client computer generally responds to search requests on the daily index far more quickly than accessing our current servers with similar requests. This illustrates the potential to distribute some of the search activities to the user machine, lowering the server system requirements of the data disseminator.

Currently the applet and daily index file ranges in size from 200 to 300 kbytes. This is a rather large file for dial up access and thus this interface is more suitable for users with fast Internet connections. We are currently constrained by various security mechanisms implemented by browsers running Java. For example a socket connection is required to load and process the index data. For users of the Netscape browser, this means both the index and other parts of the applet must be located on the same server (i.e., the NYU EDGAR server). In the near future we will release the source code in a way that the

applet and the daily index can be stored on servers within a corporate intranet. This way the user will be able to access the applet on a local server on a LAN and does not have to come to the NYU site for invoking the applet or for retrieving the index. However, we will really be able to push processing to the client side, and provide greater functionality when the security issues surrounding JAVA or similar client side programming tools are better resolved.

The Java applet interface has a number of unique features. First, the ticker speed is controllable by the user, so they can view index contents at a speed they desire. Second, the interface provides context sensitive help for each of the different search tools. This is helpful for end users and is again built into the client side.

We also provide users of the Java interface the capabilities of maintaining preferences for filings by ticker symbols or company name. These preferences can be written into "cookie" files for users of Netscape, or on a homepage specifically written for the user. Thus when the user loads the applet from the NYU server, the data from the "cookie" file or homepage is also loaded and processed, so that the filings from the users' preferred companies appear first on the ticker. While we can provide a preference search for other users we have chosen to initially test this with advanced users before having to support large numbers of users. As Java and browser security issues are resolved we believe maintenance and processing of recurring user preferences can be completely done on the client side without having to come to a central server. Such a solution will provide greater privacy to users on their preferences. Customization of the interface to user search preference as

illustrated above, simplifies data retrieval, minimizes effort and increases user convenience.

**5.0 Extending the Current System**

Given the commitment of Disclosure and other sponsors to continue funding of the NYU EDGAR site for the next year, we are able to continue developing the system. Our specific goals are to rationalize and stabilize the system, taking advantage of emerging Internet tools. We intend to take advantage of these new tools to lower the cost of software maintenance and to provide new functionality to the public.

Specifically we will focus on three distinct areas:

- Rationalizing indices and creating database applications
- Developing client side tools and interfaces
- Analyzing queries and improving search tools

By using a commercial relational database such as Microsoft SQL server we can store our various indexes and write simpler applications using SQL queries for data retrieval. A relational database also enables us to efficiently store and execute more complex queries on financial data schedules or mutual fund information. We anticipate it will also simplify the maintenance of ticker and other database files. We will migrate database services to a dual Pentium server running WindowsNT to simplify and lower the costs of maintenance.

We will continue developing the Java interface as well as add more tools for customizing searches and client side processing of documents. These may include simple functions like KWIC at the client side or more complex functions calculating financial ratios. In addition we will upgrade the help systems to the data.

Finally we have acquired the disk space to provide full text indexing of one year's worth of 10-Ks, and to full text index the latest few days worth of filings on a rolling basis. We will study the queries[2] people ask of these filings to identify potential ways of better tagging and cataloging the underlying data. In addition we will work with existing freeWAIS or other public domain full text search tools and modify them to improve their performance in searching through the selected portions of the EDGAR database.

The above should result in substantial improvements to the current free access system to the EDGAR database.

**6.0 Public Access**

All software developed at NYU during the period of the NSF grant will be made available to the SEC for use in their system. We will assist the SEC in implementing any of the software. In addition we will continue to improve the systems as described above and make these improvements also available for free public access for the coming year. We will also release the source code

[2]Given privacy concerns we did not log specific queries matched to specific users in this project. Given alternate sites with free EDGAR data, we are in a position now with user permission to log queries. This will help us to better understand how specific users utilize the EDGAR data.

by setting up an ftp site for most applications, although we will not be able to provide support for them.

**7.0 Challenges**

There were a number of challenges to implementing this project. First, the project was undertaken during a time of rapid innovation and change in the Internet. Thus we were constantly working with the latest technologies and adapting to new innovations. We constantly focused on designing the system to anticipate and take advantage of new developments (e.g., Java) while providing a public service. Nevertheless, some of the techniques for dissemination early in the project like electronic mail and gopher were rendered obsolete in a very short time. Today, newer tools for WWW development, maintenance and database to WWW integration tools, make it much easier to accomplish higher value added dissemination as well as maintain a large access system at low cost.

Second, given the widespread publicity of the project we were inundated with both press and public requests for information and help in using this system. While the SEC was invited to participate early in the project they declined to do so, forwarding all information requests by investors or users directly to us. We responded as best we could to these requests which were distracting and took substantial time. Greater cooperation from the SEC early in the project would have helped us immensely in making this a better system.

Third, given normal student turnover, and hiring by local firms who sought Internet expertise, we had to constantly bring new staff down the learning curve. However, all students who worked on this project have been hired by

major financial services or financial information provider firms. (See project team on the EDGAR homepage).

Fourth, there were difficulties coordinating and scheduling advisory meetings, and the administration of this public demonstration project was both costly in time and effort, distracting us from developing applications and research projects.

**8.0 Conclusions**

This project demonstrated that it is fairly straightforward to widely disseminate large government data sets via the Internet and that there is widespread public demand and interest in using the data. Indeed widespread public interest and support for this project led to the adoption of the basic system and some specialized search applications by the SEC. Over 3 million documents were transferred to the public during the project in a very cost effective and timely manner.

We provided a number of useful search and retrieval tools on the EDGAR database. However, it is important to recognize that downstream value addition to EDGAR data is greatly simplified if the input documents have better structure. While we automatically extracted certain portions of the data from the text filings to create profiles, this effort illustrated the difficulty of automatically adding value to documents that do not follow standard formats or have limited tags to identify key data. Thus we recommend the following steps the SEC can undertake to simplify value addition, and enhance services to the public. First, the SEC should encourage better conformance to specific formats and structuring of data before permitting filing of a document on the

EDGAR system. The SEC should also monitor if documents conform to standards, and that key classification data such as SIC codes and section headers are correct. Second, the SEC should ensure that a more comprehensive group of filings is available on-line, including insider purchase and sale information, and all mutual fund and municipal bond information. Third the SEC should specify more categories of tags to impose better structure on the documents. Specific tags and document summary information could be specified in a more meaningful header file that is smaller in size. Header files can help users to automatically filter and more easily select relevant documents . Fourth, permitting document creators and filers, or auditors to add new tag categories that flow through the EDGAR system can also help reduce downstream data extraction costs.

Emerging tools like Java for platform independent or client side programs promise the potential of radically improving both the filing, retrieval and document value addition processes. While few users have computers and communications to support the Java applications we provide, we expect the current growth of the Internet will result in widespread adoption of these tools within six months. The emerging computing environment with distributed client side applications will allow a government agency to maintain low cost Internet distribution of the data, and then rely on both the for profit and non-profit sector to provide new software applications to add value to the data.

By providing free access to the raw data, and ensuring that it is well defined and structured, the goverment will lower entry barriers into the information marketplace. It will encourage more data vendors to enter the marketplace,

innovate, add value and resell the information. This will better serve customers of the information. A basic government system will not compete against existing vendors. It will only encourage greater competition in this marketplace. The Internet EDGAR project did not necessarily adversely affect the existing vendors of corporate data. It educated more users, and encouraged the purchase of higher valued added data from the information vendors. All the vendors are now working to implement Internet access mechanisms to their value added products.

Besides serving thousands of users worldwide by providing reliable and daily access to valuable EDGAR data, this demonstration project educated many users about the Internet and the EDGAR database. It also provided many students with opportunities and access to infrastructures to build Internet applications. We are gratified that the SEC has adopted this system and would like to thank the NSF and other sponsors for their support during the project. Various users also contributed in helping us better understand their needs as well as the content of the documents. Finally I would like to acknowledge the leadership of Carl Malamud, President of Internet Multicasting Service, who conceived of the project and provided very effective leadership in implementing this system in the public domain.

Appendix 1: Publications:

" The EDGAR Internet Project: Web Application Development Considerations," Mark Ginsburg and Ajit Kambil, New York University, Proceedings of the Association for Information Systems, Americas Conference on Information Systems, August 1995.

"Pre- and Post Coordinate Indexing = Partial Coordination," David Bodoff and Ajit Kambil. NYU Center for Research in Information Systems Working Paper IS-96-1 (under revise and resubmit to the Journal of the American Society for Information Systems).