



New York University
A private university in the public service

Office of Sponsored Programs

15 Washington Place, Apt. 1-H

New York, NY 10003

Telephone: (212) 998-2121

FAX: (212) 995-4029

July 7, 1993

Proposal Processing Unit
National Science Foundation
1800 G Street, N.W., Room 223
Washington, D.C. 20006

Attention: Division of Networking & Communications
Research & Infrastructure, CISE

Ladies and Gentlemen:

We are pleased to submit on behalf of New York University the enclosed proposal for support of research entitled, "Internet Access to Large Government Data Archives: the Direct Edgar Access System," to be undertaken by Professor Ajit Kambil in the Department of Information Systems at the Stern School of Business, in cooperation with the Internet Multicasting Service.

Funds are requested in the amount of \$508,927 for the two-year period commencing October 1, 1993; first-year costs are budgeted at \$292,540.

We appreciate the Foundation's consideration of this proposal and are hopeful that it will meet with approval. Please let us know if there are any questions.

Sincerely,

Ann H. Greenberg
Assistant Chancellor

AHG:rl
enclosure

cc: Dr. Donald R. Mitchell
NCR, National Science Foundation
(with list of proposed reviewers)

Dr. Carl Malamud
President, Internet Multicasting

cc: A. Kambil ✓
R. Brief
P. Cardin

1. Suggested Reviewers

Professor Michael Schwartz
Department of Computer Science, University of Colorado
Boulder, Colorado
schwartz@cs.colorado.edu

Professor Larry Landweber
Department of Computer Science, University of Wisconsin-Madison
1210 West Dayton Street, Madison, Wisconsin 53706
lhl@cs.wisconsin.edu

Dr. Clifford Lynch
Director of Library Automation, University of California
300 Lakeside Drive, 8th Floor, Oakland, California 94612-3350
calur@uccmvsa.ucop.edu

Dr. Vinton Cerf
Vice President, Corporation for National Research Initiatives
1895 Preston White Drive, Suite 100
Reston, VA 22091
vcerf@cnri.reston.va.us

Dr. Lyman Chapin
Chief Network Architect
BBN Communications
150 Cambridge Park Drive
Cambridge, MA 02140
lyman@bbn.com

George H. Brett II
Director, Clearinghouse for Networked Information
3021 Cornwallis Road
Research Triangle Park, North Carolina 27709
George.Brett@cnidr.org

Lewis M. Branscomb
Director, Science Technology and Public Policy Program
John F. Kennedy School of Government, Harvard University
79 John F. Kennedy Street, Cambridge MA02138

Dennis E. Egan
Director, Information Sciences Research
Bellcore,
445 South Street, Morristown, New Jersey 07962-1910

APPENDIX III

INFORMATION ABOUT PRINCIPAL INVESTIGATORS/PROJECT DIRECTORS

Submit only ONE copy of this form with your proposal. Attach it on top of the cover page of the copy of your proposal that bears the original signatures. Leave the back of the page blank. *Do not include this form with any of the other copies of your proposal, as this may compromise the confidentiality of the information.*

Please check the appropriate answers to each question for all principal investigator(s)/project director(s) listed on the cover page, using the same order in which they were listed there:

	Principal Investigator/ Project Director	First Additional PI/PD	Second Additional PI/PD	Third Additional PI/PD	Fourth Additional PI/PD
--	---	---------------------------	----------------------------	---------------------------	----------------------------

- | | | | | | |
|---|-------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 1. Is this person | | | | | |
| Female | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Male | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Is this person a | | | | | |
| U.S. Citizen | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Permanent Resident | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Other non-U.S. Citizen | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 3. Which one of these categories best describes this person's ethnic/racial status? (If more than one category applies, use the category that most closely reflects the person's recognition in the community.) | | | | | |
| American Indian or Alaskan Native | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Asian | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Black, not of Hispanic Origin | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Hispanic | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Pacific Islander | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| White, not of Hispanic Origin | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 4. Does this person have a disability* which limits a major life activity? | | | | | |
| Yes | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| No | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Check here if this person does not wish to provide some or all of the above information | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Required: Check here if this person is currently serving (or has previously served) as PI, Co-PI or PD on any Federally funded project. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

AMERICAN INDIAN OR ALASKAN NATIVE: A person having origins in any of the original peoples of North America and who maintains cultural identification through tribal affiliation or community recognition.

ASIAN: A person having origins in any of the original peoples of East Asia, Southeast Asia or the Indian subcontinent. This area includes, for example, China, India, Indonesia, Japan, Korea and Vietnam.

BLACK, NOT OF HISPANIC ORIGIN: A person having origins in any of the black racial groups of Africa.

HISPANIC: A person of Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin, regardless of race.

PACIFIC ISLANDER: A person having origins in any of the original peoples of Hawaii; the U.S. Pacific territories of Guam, American Samoa, and the Northern Marianas; the U.S. Trust Territory of Palau; the islands of Micronesia and Melanesia; or the Philippines.

WHITE, NOT OF HISPANIC ORIGIN: A person having origins in any of the original peoples of Europe, North Africa, or the Middle East.

*Disabled: A person having a physical or mental impairment that substantially limits one or more major life activities; who has a record of such impairment; or who is regarded as having such impairment.

WHY THIS INFORMATION IS BEING REQUESTED:

The Federal Government has a continuing commitment to monitor the operation of its review and award processes to identify and address any inequities based on gender, race, ethnicity, or disability of the proposed principal investigators/project directors and co-principal investigators. To gather the information needed for this important task, you should submit a single copy of this form with each proposal; however, submission of the requested information is not mandatory and is not a precondition of award. Any individual not wishing to submit the information should check the box provided for this purpose. (The exception is information about previous Federal support, the last question above.)

Information from this form will be retained by Federal agencies as an integral part of their Privacy Act Systems of Records in accordance with the Privacy Act of 1974. These are confidential files accessible only to appropriate Federal agency personnel and will be treated as confidential to the extent permitted by law. Data submitted will be used in accordance with criteria established by the respective Federal agency for awarding grants for research and education, and in response to Public Law 99-383 and 42 USC 1885c.

COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

APPENDIX IV

For Consideration by NSF Organization Unit(s) <small>(Indicate the most specific unit known, i.e. program, division, etc.)</small> <div style="text-align: center; font-weight: bold;">NCR</div>		For NSF Use Only <div style="text-align: center; font-weight: bold;">NSF Proposal Number</div>														
Program Announcement/Solicitation No./Closing Date																
Date Received	Number of Copies	Division Assigned	Fund Code	File Location												
Employer Identification Number (EIN) or Taxpayer Identification Number (TIN) 1-135562308-A1	Show Previous Award No. If This is: <input type="checkbox"/> A Renewal or <input type="checkbox"/> An Accomplishment-Based Renewal		Is This Proposal Being Submitted to Another Federal Agency? YES___ NO <u>X</u> If YES, List Acronym(s)													
Name of Organization to Which Award Should be Made: New York University		Address of Organization, Including Zip Code Washington Square New York, New York 10003														
Institutional Code (if known) 0027854000																
Is Submitting Organization: <input type="checkbox"/> For-Profit Organization; <input type="checkbox"/> Small Business; <input type="checkbox"/> Minority Business; <input type="checkbox"/> Woman-Owned Business																
Branch/Campus/Other Component (Where work is performed, if different) Leonard N. Stern School of Business				Institutional Code (if known)												
Title of Proposed Project Internet Access to Large Government Data Archives: The Direct Edgar Access System																
Requested Amount \$ 508,927	Proposed Duration (1-60 months) 24 months	Requested Starting Date October 1, 1993														
Check Appropriate Box(es) if This Proposal Includes Any of the Items Listed Below: <table style="width: 100%; border: none;"> <tr> <td><input type="checkbox"/> Vertebrate Animals</td> <td><input type="checkbox"/> National Environmental Policy Act</td> <td><input type="checkbox"/> Facilitation for Scientists/Engineers with Disabilities</td> </tr> <tr> <td><input type="checkbox"/> Human Subjects</td> <td><input type="checkbox"/> Proprietary and Privileged Information</td> <td><input type="checkbox"/> Research Opportunity Award</td> </tr> <tr> <td><input type="checkbox"/> Historical Places</td> <td><input type="checkbox"/> Disclosure of Lobbying Activities</td> <td><input type="checkbox"/> International Cooperative Activity:</td> </tr> <tr> <td colspan="3"><input type="checkbox"/> Research Involving Genetically Engineered Organisms</td> </tr> </table> <div style="text-align: right; margin-top: 10px;"> Country/Countries _____ </div>					<input type="checkbox"/> Vertebrate Animals	<input type="checkbox"/> National Environmental Policy Act	<input type="checkbox"/> Facilitation for Scientists/Engineers with Disabilities	<input type="checkbox"/> Human Subjects	<input type="checkbox"/> Proprietary and Privileged Information	<input type="checkbox"/> Research Opportunity Award	<input type="checkbox"/> Historical Places	<input type="checkbox"/> Disclosure of Lobbying Activities	<input type="checkbox"/> International Cooperative Activity:	<input type="checkbox"/> Research Involving Genetically Engineered Organisms		
<input type="checkbox"/> Vertebrate Animals	<input type="checkbox"/> National Environmental Policy Act	<input type="checkbox"/> Facilitation for Scientists/Engineers with Disabilities														
<input type="checkbox"/> Human Subjects	<input type="checkbox"/> Proprietary and Privileged Information	<input type="checkbox"/> Research Opportunity Award														
<input type="checkbox"/> Historical Places	<input type="checkbox"/> Disclosure of Lobbying Activities	<input type="checkbox"/> International Cooperative Activity:														
<input type="checkbox"/> Research Involving Genetically Engineered Organisms																
P/VPD Department Information Systems		P/VPD Postal Address 44 West 4th Street Management Education Center, Suite 9-82 New York, NY 10012-1126														
P/VPD Fax Number 212-995-4228																
Names (Typed)	Social Security No.*	High Degree, Yr	Telephone Number	Electronic Mail Address												
P/VPD Name Ajit Kambil	019-64-2144	Ph.D 1993	212-998-0843	akambil@stern.nyu.edu												
Co-P/VPD																
Co-P/VPD																
Co-P/VPD																
Co-P/VPD																
NOTE: The FULLY SIGNED Certification Page must be submitted immediately following this Cover Sheet.																
*Submission of social security numbers is voluntary and will not affect the organization's eligibility for an award. However, they are an integral part of the NSF information system and assist in processing the proposal. SSN solicited under NSF Act of 1950, as amended.																

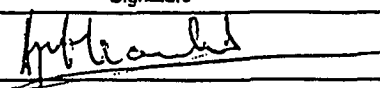
CERTIFICATION PAGE

Certification for Principal Investigators and Co-Principal Investigators:

I certify to the best of my knowledge that:

- (1) the statements herein (excluding scientific hypotheses and scientific opinions) are true and complete, and
- (2) the text and graphics herein as well as any accompanying publications or other documents, unless otherwise indicated, are the original work of the signatories or individuals working under their supervision. I agree to accept responsibility for the scientific conduct of the project and to provide the required progress reports if an award is made as a result of this application.

I understand that the willful provision of false information or concealing a material fact in this proposal or any other communication submitted to NSF is a criminal offense (U.S. Code, Title 18, Section 1001).

Name (Typed)	Signature	Date
PI/PD Ajit Kambil		July 5, 1993
Co-PI/PD		
Co-PI/PD		
Co-PI/PD		
Co-PI/PD		

Certification for Authorized Institutional Representative or Individual Applicant:

By signing and submitting this proposal, the individual applicant or the authorized official of the applicant institution is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding Federal debt status, debarment and suspension, drug-free workplace, and lobbying activities (see below), as set forth in Grants for Research and Education in Science and Engineering (GRESE), NSF 92-89. Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, Section 1001).

Debt and Debarment Certifications

(If answer "yes" to either, please provide explanation.)

Is the organization delinquent on any Federal debt?

Yes ☐ No ☒

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency?

Yes ☐ No ☒

Certification Regarding Lobbying

This certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding \$100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding \$150,000.

Certification for Contracts, Grants, Loans and Cooperative Agreements

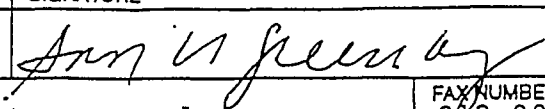
The undersigned certifies, to the best of his or her knowledge and belief, that:

(1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, "Disclosure Form to Report Lobbying," in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than \$10,000 and not more than \$100,000 for each such failure.

AUTHORIZED INSTITUTIONAL REPRESENTATIVE		SIGNATURE	DATE
NAME/TITLE (TYPED) Ann H. Greenberg, Asst. Chancellor Office of Sponsored Programs			
TELEPHONE NUMBER 212-998-2121	ELECTRONIC MAIL ADDRESS greenberga@acfccluster.nyu.edu		FAX NUMBER 212-995-4029

Internet Access to Large Government Data Archives: The Direct Edgar Access System

2. Project Summary

This project will develop and demonstrate ways to post large government data archives on the Internet for access by researchers and the general public. The two-year project will use the Securities and Exchange Commission (SEC) EDGAR Database as a test database and will develop and implement techniques to provide low-cost management and dissemination of these data archives. The goal of the project is to develop techniques that can be used directly by government agencies and other information providers, giving these organizations the tools and knowledge to allow them to place government information on the Internet for direct access by researchers and the general public. Secondary goals of the project are to make key financial archives from the SEC available to researchers and the general public who would not otherwise have access to the information.

The Direct EDGAR Access System will be under the direction of principal investigator Dr. Ajit Kambil of the New York University Stern School of Business. The project starts with the EDGAR Dissemination Service, a raw data feed intended for information wholesalers and retailers to process. The project will develop techniques for automatic reformatting of this information into file systems appropriate for dissemination on the Internet and will maintain an Internet-based server capable of supporting 1.0 Mbps of application level data throughput.

The project will then work with public domain tools to adapt them to the specific needs of large government data archives. Custom WAIS vocabularies, support for multicast based data dissemination, and semi-automatic structuring of hypertext and Gopher links will all be investigated. The project will obtain advice from researchers in the Finance department of the Stern School to identify ways to provide semi-automatic value-added cataloging and data abstraction.

The data conversion, access and management techniques learned in this project will be fully detailed in project reports and all software and techniques will be placed in the public domain.

3. Table of Contents	
1. Cover Sheet (NSF Form 1207).....	1
2. Project Summary.....	2
3. Table of Contents.....	3
4 Project Description.....	4
4.1. Access to Government Data.....	4
4.2.The SEC's Edgar Database.....	4
4.3.Data Conversion and Maintenance.....	5
4.4. Data Analysis and Access.....	7
4.4.1. Network Based Tools.....	7
4.4.2. Data Administration and Aggregation.....	8
4.4.3. Advanced Topics.....	9
4.5. Timelines.....	9
4.6. Project Reporting.....	9
4.6.1. Oversight Committee.....	10
4.6.2. Timely Notification of Significant Problems.....	10
4.6.3. Verbal Reports and Liaison.....	10
4.6.4. Monthly Letter Progress Reports.....	10
4.6.5. Quarterly Status Report.....	11
4.6.6. Final Report.....	11
4.7. Project Benefits.....	11
4.8 Letter of commitment	
Internet Multicasting Service.....	13
5. Bibliography.....	15
6. Biographical Sketches.....	17
7. Summary Proposal Budget.....	20
7.1 Summary Budget Year 1.....	21
7.2 Summary Budget Year 2.....	22
7.3 Budget Explanation.....	23
8. Current and Pending Support.....	25

4 Project Description

4.1. Access to Government Data

Federal and state government agencies are all tackling the very difficult issue of making key information available to the public over computer-based information systems. This project develops tools and techniques that enhance the ability of government agencies to make their information available to the public over the Internet computer network.

The Internet computer network is a key U.S. priority, receiving substantial backing from agencies ranging from the National Science Foundation to NASA to the Department of Defense. This infrastructure is widely available in university, research, and government institutions. The Internet and related initiatives are a key part of the emerging National Research and Education Network.

While the network is already in place, there is considerable work to be done to make large data archives available over this infrastructure. This project investigates techniques by which the government can make large archives available to the public over the Internet. A key goal of the project is to provide an example for agencies from which they can learn to directly provide information to the Internet in a cost-effective manner.

4.2. The SEC's Edgar Database

EDGAR is the Security and Exchange Commission's Electronic Data Gathering, Analysis and Retrieval System. The EDGAR Dissemination Service is operated by Mead Data Central, Inc. as the SEC's dissemination subcontractor. Under the EDGAR rules and phase-in schedule, financial firms must begin filing key SEC forms electronically. That information becomes available through EDGAR as either a live broadcast-service or a delayed-dissemination tape service.

Under the EDGAR schedule, the initial project began July 15, 1992 with 450 voluntary filers. Beginning April 26, 1993, additional groups of filers joined the program as follows:

April 26, 1993	Group 01,	450 Filers
July 19, 1993	Group 02	1203 Filers
October 4, 1993	Group 03	749 Filers
December 6, 1993	Group 04	1000 Filers

By December 1993, there will be a total of 3,402 filers. At that point, a six month test and evaluation period will be conducted. If successful, final EDGAR rules will be adopted and all other SEC filers will be brought on-line according to the following schedule:

August, 1994,	Group 05,	2299 Filers
November, 1994	Group 06,	2297 Filers
May, 1995,	Group 07	1496 Filers
August, 1994	Group 08,	2299 Filers
November 1995	Group 09	2299 Filers
May,1996	Group 10	808 Filers

Mead Data provides several levels of dissemination of data. The Level 1 Broadcast Service is an on-line stream of data using X.25 or TCP/IP protocols. It requires subscribers to maintain seven 56 kbps lines (or presumably a single T1 1.544 Mbps line) into the Dissemination Service mainframe. Lower-volume subsets of the data as well as overnight courier services are available.

The data in the EDGAR system consists of the following key subsets:

Subset 1	Periodic filings by companies on the New York Stock Exchange including forms 10-K,, 10-Q,, 8-K,, 10 and all Definitive Proxies.
Subset 2	Periodic filings by companies on the American Stock Exchange including all forms in Subset 1.
Subset 3	Periodic filings by companies traded over-the-counter including all forms in Subset 1.
Subset 4	Filings made pursuant to the Williams Act including Schedules 12-D,, 14D,, 14D-9,, 13-G,, 13-E1,, 13-E3,, and 13-E4.
Subset 5	Annual and quarterly reports of all companies pursuant to the Securities and Exchange Act of 1934.
Subset 6	Securities Act of 1933 Registration Statements including S series
Subset 7	Stock ownership reports including forms 3,, 4,, 144
Subset 8	Periodic reports filed under the Investment Company Act of 1940 including Forms NSAR-A,, NSAR-B,, and NSAR-U.

The data itself is formatted as a standard series of files with a Control Block header and a series of SGML-tagged data fields. The header indicates the form type, the company name, and an indication if the filing is a submission, correction, deletion, or other form of document transfer. The data itself consists of text files formatted with tags and data values.

4.3.Data Conversion and Maintenance

The project will establish a Level 1 Broadcast service feed to the Dissemination Service mainframe. Initial software development will write data parsers that translate the incoming data stream into a standard set of files. The initial data conversion will use standard tools such as PERL and the C Programming Language and will result in a standard file system structure appropriate to this data set.

The data will then be maintained on a Sparcstation server at Internet Multicasting Service with a large disk farm. This machine will be connected to the Internet and will be configured to support an application level data stream of 1.0 Mbps, equivalent to approximately 50 simultaneous anonymous FTP sessions.

The data conversion and maintenance part of the project will be divided into two phases:

- Establishment of the data feed
- Ongoing processing of the data during the project

In the initial phase, lasting for the first six months of the project, a dedicated T1 line will be established to the EDGAR dissemination service using the TCP/IP protocol option. This link will terminate directly in the EDGAR dissemination mainframe.

The other end of the link will terminate on a DSU/CSU, which will connect to a router on a dedicated LAN at the Internet Multicasting Service. That LAN will also house a Sparcstation which will provide the main processing power for the project. The Sparcstation will be configured with an 8 mm backup device and 20 gigabytes of disk space. ¹

During the initial project phase, we will establish a procedure for incoming data from the Edgar dissemination system to be spooled on the Sparcstation disk farm. The initial spooling will be in the raw format defined by the dissemination service.

Using PERL scripts, C programs, and other Unix-based processing tools, the project will then begin the initial reformatting of the data. The exact format of the file system and the data will be determined by the project team, attempting to balance processing requirements, ease of use, completeness, and other conflicting factors.

¹See letter of commitment from INTERNET MULTICASTING SERVICE at the end of this project description.

We anticipate the file structure to consist of a series of directories, one for each type of form. In that directory would be a series of files, each coded by company name, date of filing, and type of filing (e.g., initial filing, amendment, etc...).

A series of indices and symbolic links will be automatically generated to allow direct access to all forms a particular company has filed, or by a particular date, or other cross-sections of the data.

In addition to the initial restructuring of the data, the project will also establish verification procedures to determine that data integrity has been maintained. The procedures will include statistically valid samples of the data on the Sparcstation which will be compared with source documents, as well as an analysis of the resulting file structure against the initial data feed.

Once the initial procedures are put in place, the system will stabilize and will continue the processing of information throughout the life of the project. This base data will form the platform for public access and will be supplemented throughout the project by additional data generated by students at New York University.

4.4. Data Analysis and Access

The project team will start with the basic data conversion then apply the skills of information systems, finance and accounting faculty and students at the New York University Stern School of Business. Students will interview faculty in management, finance and accounting departments using structured techniques to identify specific data access and search requirements to facilitate management research as well as classroom, educational and public use of the EDGAR data. The student groups under the direction of the principal investigator will work with the base data to apply modern resource discovery tools and will examine techniques for providing value-added information on the base data.

4.4.1. Network Based Tools

The project will develop and modify common public domain tools such as WAIS, World Wide Web, Gopher, Veronica, Viola, Mosaic and others as applicable to work within the specific constraints of large government databases. The project will evaluate techniques such as custom WAIS vocabularies to aid in searching.

In addition to the standard resource discovery tools, the project will use anonymous FTP, e-mail based information servers, and other low-tech methods to make the data available. The project will investigate how to

make a large, complex mass of information such as the EDGAR database easily accessible by low-bandwidth users with tools such as basic e-mail.

The project will begin by immediately establishing access to information with anonymous FTP and a public domain e-mail-based server. This basic, raw access will allow immediate access to Edgar information for researchers willing to work with the data in raw format.

The next level of value-added service to be provided will index the information for inclusion in a WAIS database. The commercial version of WAIS will be used because the indices are more efficient and the server code has higher performance. However, any WAIS client will be able to access the information.

Access by Gopher and hypertext links in the World Wide Web system will take a little longer to establish since these systems establish structure on the data. We anticipate a rough version of a Gopher server to be immediately established, but a final version of the menus will be developed during the data administration and aggregation portions of the project.

4.4.2. Data Administration and Aggregation

In addition to providing and modifying common tools for access to the data, the project team will investigate techniques for providing value-added information in a semi-automatic fashion. Data-specific summaries such as indices and abstracts can be automatically generated. Multiple iterations of data (original filings plus additional corrections) can be folded into a single combined piece of information. Aggregations of data can be provided on an industry, market, or other aggregation basis.

Initially, the program will start by determining the proper way to structure the information to allow simple, menu-based navigation of the Edgar data. Combined with the WAIS free-text search capabilities, a highly functional interface to the data can be established.

A more sophisticated interface into the information can be provided using hypertext links, linking one portion of the database to another portion (or to other databases on the Internet). The project will determine which pieces of information in the database can be related to other pieces, then link those pieces together.

In this stage of the project, students from finance and information systems will work closely together to combine the semantic content of the data with the advanced tools. The result will be Gopher, WWW, WAIS, and other interfaces into the information which will provide the general public and researchers with a powerful, flexible interface to the Edgar data.

4.4.3. Advanced Topics

In addition to basic research topics detailed above, the project will investigate several novel distribution techniques. First, the project will use IP multicasting to move the database around the world as very low volume (7-9 kbps) data streams. Software for receiving the data and setting complex traps for specific pieces of information will be developed and deployed and this software will be integrated into the Session Directory (SD) software running on the Multicast Backbone (MBONE).

4.5. Timelines

The project will begin on October 1, 1993 and will run for two years. At the close of the second year, the National Science Foundation may elect to continue the project with the specific aim of providing a direct transition of the system into the SEC information systems division. Such transition is dependent on SEC evaluation of the technical feasibility of operating a database server on the Internet.

In the first six months, the basic data structure will be established and the server will become operational. The second six months of the project will provide access to the data on the Internet using basic interfaces such as Gopher, anonymous FTP, and e-mail. Basic WAIS access will be provided during the stage.

In the second year of the project, the advanced interfaces based on Gopher, WWW, and WAIS will be established and the multicast based deployment of the data will be initially deployed.

Throughout the project, frequent briefings will be provided to interested parties, including officials of the Securities and Exchange Commission. Such briefings, supplemented by regular project reports, will allow the SEC to monitor the progress of the project and to evaluate the potential for establishing a similar service within the SEC.

4.6. Project Reporting

Grantee shall provide periodic project briefings to interested parties, including government agencies, researchers in academic institutions, librarians, and others as appropriate. These briefings will detail the status of the project and identify key issues.

Principal Investigator and project participants shall provide full information to interested government agencies including the Securities and Exchange

Commission (SEC) with a view of providing interested agencies with the tools to perform similar functions internally.

4.6.1. Oversight Committee

New York University and the Principal Investigator shall establish a project oversight committee to provide advice on project progress and plans. The committee shall be chaired by a professor from the Stern School of Business and will include representatives from key interested parties, including the academic and research community, the financial services industry, the retail information industry, the Network Information Center (InterNIC) and others as appropriate.

The Oversight Committee shall meet twice annually. Principal Investigator and project participants shall present an overview of project progress and the Oversight Committee will make recommendations. The results of these meetings shall be included in the next reports to the NSF Program Director including a summary of items discussed and action items to be taken.

4.6.2. Timely Notification of Significant Problems

The Grantee will inform the NSF Program Official (either by e-mail or in writing) in a timely manner of any significant problems or events that could affect the overall schedule or progress in the program.

4.6.3. Verbal Reports and Liaison

The Grantee will meet on an informal basis, as necessary or requested, with the NSF Program Director to review progress to date and to exchange views, ideas, and information concerning the program.

When requested by the NSF Program Director, Grantee will arrange to have its subawardees in attendance at meetings which deal with their areas of activity. In addition, at the request of the NSF Program Director, the Grantee will arrange on-site meetings for the Program Officer, other Federal staff and/or representatives of the world-wide Internet community and the Grantee's professional personnel, and/or those of its subawardees.

4.6.4. Monthly Letter Progress Reports

Monthly letter progress reports will be submitted electronically to the NSF Program Official and NSF Administrative Official at the address shown on the cover page. These monthly letter progress reports will be submitted in such detail and format as required by the Foundation's Program Director and will contain statistical and narrative information on the performance of the Grantee during the preceding month.

4.6.5. Quarterly Status Report

Grantee will prepare and furnish electronically and in four hard (4) copies quarterly letter status reports; the first quarterly status report will be for the period from October 1, 1993, through December 31, 1993. These reports will show the status of all major events and summaries and major work performed during the quarter, including technical status, accomplishments, problems, collaboration activities, changes in future plans, and any pending requests for NSF approval and will be fully reconciled with the information, goals and projections contained in the Annual Report and Program Plan. The report will also include a summary of award expenditures and line charges both cumulative and for the current quarter.

The report will be prepared on a quarterly basis and shall be submitted within thirty (30) days after the reporting period ends. No quarterly report will be submitted for the quarter in which the Annual reports are submitted, but, Awardee will insure that any germane information for the quarter not contained in the Annual Report (i.e, list of pending requests for NSF approval) are submitted by separate letter.

4.6.6. Final Report

The Grantee will submit electronically and in three (3) hard copies a final report to NSF at the conclusion of the Cooperative Agreement. The final report will contain a description of all work performed and problems encountered (and if requested a copy and documentation of any and all software and data generated) in such form and sufficient detail as to permit replication of the work by a reasonably knowledgeable party or organization.

4.7. Project Benefits

The project will provide the following primary benefits:

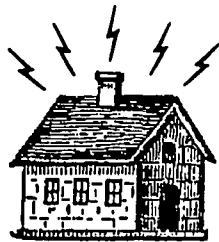
- Demonstration of how to convert large government data bases for public and research access through the Internet.
- Demonstration of new techniques for data distribution in the Internet.
- Modified resource discovery tools and management techniques for converting government archives for public and research data access.

- Improved public, student and faculty access to information resources for research and teaching in finance, accounting and management.

- Provision of a large data archive for information systems and computer science research on knowledge discovery in temporal databases.

Finally, the NYU Stern School of Business provides an especially useful infrastructure and resource base for implementing this project. The Information Systems Department and the Finance departments of the Stern School are both nationally ranked as among the top ten nationwide in terms of research and scholarship. The school, faculty and students have excellent relations with, and are very knowledgeable about the financial services industry. The school has excellent Internet access and computing infrastructure, including a backbone FDDI ring and ethernet connections to desktop computers in the Stern School of Business, Management Education Center. These resources will be utilized in this project.

INTERNET



**TOWN
HALL**

June 30, 1993

Carl Malamud
Executive Producer

Professor Ajit Kambil
Department of Information Systems
Leonard N. Stern School of Business
New York University
Management Education Center, Suite 9-170
New York, NY 10012-1126

Dear Dr. Kambil:

I am writing at your request to propose a subcontract for your National Science Foundation proposal on the new "Internet Direct EDGAR Access System." The Internet Multicasting Service would be pleased to provide core data processing functions and to work with you and your students as an integral part of this exciting research project.

Specifically, we propose a subcontract for the amounts of \$168,800 for FY94 and \$83,000 for FY95. Under this contract, we would procure, operate, and support a Sparcstation system with 20 Gigabytes of disk space and Internet connectivity capable of serving data to the Internet at a sustained throughput rate of 1.0 Mbps at the application layer. This is roughly the equivalent of 50 simultaneous File Transfer Protocol (FTP) sessions.

In addition, we will actively work with you and other project participants to secure funding for the EDGAR Dissemination Data Feed. Once the data feed is established, we will coordinate the technical establishment of that data feed and will process the incoming data in a manner suitable for access by the Internet community. We will perform such tasks for a two-year period, as per your proposal to the National Science Foundation.

For purposes of the subcontract, I will be considered a "key personnel" and you will have the right to approve any changes of key personnel. In addition, I will devote additional staff of the equivalent of 1.0 FTE as needed to perform data processing and support tasks. We agree to participate in project briefings, to contribute to project reports, and to otherwise work with you and your students to support this project under the scope of work as outlined in your proposal to the National Science Foundation.

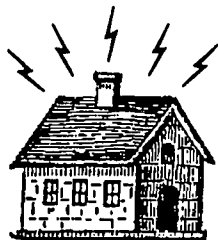
A PRODUCTION OF THE INTERNET MULTICASTING SERVICE

SUITE 1155 • THE NATIONAL PRESS BUILDING • WASHINGTON, D.C. 20045
TEL: (202) 628-2044 • FAX: (202) 628-2042 • EMAIL: Carl@Radio.COM

PARTICIPANTS

ARPA
National Science Foundation
National Press Club
Sun Microsystems
O'Reilly & Associates
Cornell University
WAIS, Incorporated
Xerox PARC
UUNET Technologies

INTERNET



**TOWN
HALL**

Professor Ajit Kambil, Page 2

The Internet Multicasting Service is a Delaware Corporation with offices in Washington, D.C. We have filed for tax-exempt status with the IRS and D.C. authorities and our by-laws are drafted as a public benefit corporation. We will be contributing cost sharing to this project in the form of personnel (\$20,000 per year), subsidization of Internet connectivity (\$25,000 per year), and additional equipment needed for the project with a value of \$100,000.

Work on this subcontract can commence upon approval by the National Science Foundation of the project, our successful establishment of funding for an EDGAR Dissemination Service data feed from the private sector or other sources, and the disbursement of funds by NYU to enable us to procure the base hardware system needed for the project.

I look forward to a long and fruitful collaboration with you and the Stern School of Business.

Sincerely,

A handwritten signature in cursive script, appearing to read 'Carl Malamud', written in dark ink.

Carl Malamud

5. Bibliography

Deering, S., Multicast Routing in Internetworks and Extended LANs,-, Proceedings of the ACM Sigcomm Conference, April, 1988.

Demarco, T., "Structured Analysis and System Specification, New York, Yourdon Press 1978.

Erickson, T., and G. Salomon. "Designing a desktop information system: Observations and Issues", Proceeding of Human Factors in Computing: CHI'91 Conference, pp 49-54, April 1991, ACM, New York

Kahn, R., and V. Cerf, The Digital Library System, Corporation for National Research Initiatives (Reston, VA).

Malamud, C., STACKS: Interoperability in Today's Computer Networks, Prentice Hall (1992)

Malamud, C., Exploring the Internet, Prentice Hall (1993).

Mead Data Central, EDGAR Dissemination Service Documentation, Release 3.0, Mead Data Central (April, 1993)

Schwartz, M. F., Resource Discovery and Related Research at the University of Colorado.- ConneXions, Vol. 5, No. 5, May 1991, pp. 12-20.

Schwartz, M. F., and Tsirigotis, P. G., Experience with a Semantically Cognizant Internet White Pages Directory Tool.- Journal of Internetworking: Research and Experience, 2(1), pp. 23-50, March 1991. Available for anonymous FTP from latour.colorado.edu in the file pub/RD.Papers/White.Pages.ps.Z.

Schwartz, M. F., and Wood, D. C. M., A Measurement Study of Organizational Properties in the Global Electronic Mail Community.- Technical Report CU-CS-482-90, Department of Computer Science, University of Colorado, Boulder, Colorado, August 1990. Submitted for publication. Available for anonymous FTP from latour.colorado.edu in the file pub/RD.Papers/RD.For.Anon.FTP.ps.Z.

Schwartz, M. F., et al., Supporting Resource Discovery Among Public Internet Archives Using a Spectrum of Information Quality.- Proceedings of the Eleventh IEEE International Conference on Distributed Computing Systems,

pp. 82-89, Arlington, Texas, May 1991. Available for anonymous FTP from
latour.colorado.edu in the directory pub/RD.Papers/Email.Study.

N. Venkatraman.and Kambil, A."Strategies for Electronic Integration: Lessons
from Electronic Filing of Income Taxes," Sloan Management Review, Winter
1991

6. Biographical Sketches

Ajit Kambil

Assistant Professor
Department of Information Systems
Leonard N. Stern School of Business
New York University
email: akambil@stern.nyu.edu

Management Education Center
44 West 4th Street, Suite 9-82
New York, NY 10012-1126
Office: [212]-998-0843
FAX: [212]-995-4228

Education

Massachusetts Institute of Technology

Ph.D Management (Major Area: Information Technologies)
S.M Management
S.M Technology & Policy Program
S.B Electrical Engineering and Computer Science

February 1993
June 1989
June 1989
September 1985

Positions Held

1992-present

Assistant Professor of Information Systems
Leonard N. Stern School of Business
New York University

6/1991-8/1991

6/1990-8/1990

Lecturer in Telecommunications, Master of Science in MIS Program
Graduate School of Management
Boston University

6/1990-8/1990

Adjunct Faculty, Simmons Graduate School of Management

5/1987-8/1987

Consultant in the Telecommunications Practice, Arthur D. Little Inc.

5/1985-8/1985

Research Intern, INTELSAT Advanced Satellite Systems Planning

Research

My research is focused on three related themes: the effects and use of information technology to transform industrial structure and competition; processes and methods for alignment of business and information technology strategies; and the effective design and use of telecommunications and collaborative technologies as a business resource. My current projects include: an extended study of the impact of electronic tax filing on structure and competition in tax preparation and related markets, an assessment of outsourcing practices and its use to align business and information systems strategies, and a study of techniques for the effective use of the Internet as an educational and business resource.

Recent Articles:

"Information Technology and Vertical Integration: Evidence from the Manufacturing Sector," in M.G. Calvert and S. Wildman (eds), *Electronic Service Networks: A Business and Policy Challenge*, Praeger Publishing Co., December 1990.

"Strategies for Electronic Integration: Lessons from Electronic Filing of Income Taxes," Sloan Management Review, Winter 1991, with N. Venkatraman.

"Electronic Integration: A Critical Review and Network Extensions," in R.J Streng et al (eds) *Scientific Research on EDI*, Samsom Publishing, 1992

"Strategic Management of Information Technology Investments: An Options Perspective" in R.D Banker, R.J Kauffman, M.A Mahmood (eds), *Strategic Information Technology Management: Perspectives on Organizational Growth and Competitive Advantage*, Idea Group Publishing, 1993

Collaborators

My dissertation committee was chaired by Jack Rockart and included N. Venkatraman, John Henderson, and Benn Konsynski

My collaborators also include Jim Short, Henry Lucas, Jon Turner, Ken Laudon. I am supervising research with three doctoral students: Luis Martins, Michael Fish and Alan Zimmerman.

Carl Malamud
President
Internet Multicasting Service
Suite 1155
The National Press Building
Washington D.C 20045

Carl Malamud is President of the Internet Multicasting Service, an organization that is providing large-scale dissemination of data onto the Internet. The Internet Multicasting Service is a public-interest corporation with headquarters in the National Press Building in Washington, D.C.

Malamud is the author of seven professional reference texts in the fields of networking and database systems and is active in the Internet Engineering Task Force. He was project coordinator for conversion and posting of technical standards developed by the International Telecommunication Union, a system that successfully used the Internet for dissemination of large data archives. The ITU has since successfully taken the function over and is providing distribution of key technical standards over the Internet.

Malamud has served as a consultant to many government groups including Argonne National Laboratory, Lawrence Livermore National Laboratory, the Joint Chiefs of Staff, the Board of Governors of the Federal Reserve System, the Office of the Secretary for Defense, and the Defense Technical Information Center. He has served as a consultant to many firms in the computer industry, including Sun Microsystems, Control Data Corporation, Hewlett-Packard, Xerox, Digital Equipment Corporation, and many others.

NYU will cost-share in accordance with NSF policy.

SUMMARY PROPOSAL BUDGET

APPENDIX V

ORGANIZATION New York University				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (MONTHS) Proposed Granted	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Ajit Kambil, P.I.				AWARD NO.		
A. SENIOR PERSONNEL: P/PI, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-mos.	Funds Requested By Proposer	Funds Granted By NSF (If Different)
1. Ajit Kambil				CAL	3	1
2.						
3.						
4.						
5.						
6. () OTHERS (LIST INDIVIDUALLY ON BUDGET EXPLANATION PAGE)						
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)				3	1	30,203.
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. () POST DOCTORAL ASSOCIATES						
2. () OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)						
3. (5) GRADUATE STUDENTS 1 Research Asst. @100% + 4 Grad. Assts @50%						68,224.
4. () UNDERGRADUATE STUDENTS						
5. () SECRETARIAL - CLERICAL						
6. () OTHER						
TOTAL SALARIES AND WAGES (A + B)						98,427.
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						8,989.
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						107,416.
D. PERMANENT EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$1,000.)						
TOTAL PERMANENT EQUIPMENT						12,000.
E. TRAVEL 1. DOMESTIC (INCL. CANADA AND U.S. POSSESSIONS)						12,000.
2. FOREIGN						
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____						
2. TRAVEL _____						
3. SUBSISTENCE _____ 8,000						
4. OTHER _____						
() TOTAL PARTICIPANT COSTS						8,000.
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						2,000.
3. CONSULTANT SERVICES						
4. COMPUTER (ADPE) SERVICES						
5. SUBCONTRACTS						251,800.
6. OTHER Tuition Remission						34,795.
TOTAL OTHER DIRECT COSTS						288,595.
H. TOTAL DIRECT COSTS (A THROUGH G)						428,010.
I. INDIRECT COSTS (SPECIFY RATE AND BASE)						
TOTAL INDIRECT COSTS						80,917.
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						508,927.
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPM 252 AND 253)						
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						\$ 508,927 \$
PI / PD TYPED NAME & SIGNATURE Ajit Kambil		DATE July 5, 1993	FOR NSF USE ONLY			
INST. REP. TYPED NAME & SIGNATURE		DATE	INDIRECT COST RATE VERIFICATION			
			Date Checked	Date Of Rate Sheet	Initials - DGC	

NYU will cost-share in accordance with NSF policy.

10/1/93 - 9/30/94

SUMMARY PROPOSAL BUDGET

APPENDIX V

				FOR NSF USE ONLY		
ORGANIZATION				PROPOSAL NO.		DURATION (MONTHS)
New York University						Proposed Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR				AWARD NO.		
Ajit Kambil, Ph.D.						
A. SENIOR PERSONNEL: PI/PI, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-mos.		Funds Requested By proposer
				CAL	ACAD	SUMR
1.	Ajit Kambil, P.I.			1		\$ 7,278.
2.						
3.						
4.						
5.						
6.	() OTHERS (LIST INDIVIDUALLY ON BUDGET EXPLANATION PAGE)					
7.	(1) TOTAL SENIOR PERSONNEL (1 - 6)			1		7,278.
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1.	() POST DOCTORAL ASSOCIATES					
2.	() OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)					
3.	(3) GRADUATE STUDENTS 1 Res. Asst. @100% (16,640) + 2 grad. assts @ 50%					33,280.
4.	() UNDERGRADUATE STUDENTS					
5.	() SECRETARIAL - CLERICAL					
6.	() OTHER					
TOTAL SALARIES AND WAGES (A + B)						40,558.
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS) 29%						2,111.
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						42,669.
D. PERMANENT EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$1,000.)						
IBM Compatible Intel 486/DX - 3,000						
Color Monitor - 1,500						
Gigabyte SCSI Hard Disk - 2,000						
Hublet - 500						
WAIS Server Software - 5,000						
TOTAL PERMANENT EQUIPMENT						12,000.
E. TRAVEL 1. DOMESTIC (INCL. CANADA AND U.S. POSSESSIONS)						6,000.
2. FOREIGN						
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ _____						
2. TRAVEL _____						
3. SUBSISTENCE _____						
4. OTHER _____						
(5) TOTAL PARTICIPANT COSTS						4,000.
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						1,000.
3. CONSULTANT SERVICES						
4. COMPUTER (ADPE) SERVICES						
5. SUBCONTRACTS Internet Multicasting Service						168,800.
6. OTHER Tuition Remission @ 51% of Graduate Student Salaries						16,973.
TOTAL OTHER DIRECT COSTS						186,773
H. TOTAL DIRECT COSTS (A THROUGH G)						251,442
I. INDIRECT COSTS (SPECIFY RATE AND BASE)						
@ 55.04 MTDC (Excludes Equipment Participant costs: TOTAL INDIRECT COSTS subcontracts > 25,000: Tuition = 74,669						41,098.
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						292,540.
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPM 252 AND 253)						
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						\$ 292,540 \$
PI/ PD TYPED NAME & SIGNATURE* Ajit Kambil				DATE July 5, 1993		FOR NSF USE ONLY
INST. REP. TYPED NAME & SIGNATURE*				DATE		INDIRECT COST RATE VERIFICATION
						Date Checked Date Of Rate Sheet Initials - DGC

10/1/94 - 9/30/95

NYU will cost-share in accordance
with NSF policy.

SUMMARY PROPOSAL BUDGET

APPENDIX V

				FOR NSF USE ONLY			
ORGANIZATION				PROPOSAL NO.	DURATION (MONTHS)		
New York University					Proposed	Granted	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR				AWARD NO.			
Ajit Kambil							
A. SENIOR PERSONNEL: P/VPD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-mos.		Funds Requested By proposer	Funds Granted By NSF (If Different)
				CAL	ACADISUMR		
1. Ajit Kambil, P.I.				3		\$ 22,925.	\$
2.							
3.							
4.							
5.							
6. () OTHERS (LIST INDIVIDUALLY ON BUDGET EXPLANATION PAGE)							
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)				3		22,925.	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. () POST DOCTORAL ASSOCIATES							
2. () OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)							
3. (3) GRADUATE STUDENTS 1 Research Assistant @ 100% + 2 Grad. Assts. @ 50%						34,944	
4. () UNDERGRADUATE STUDENTS							
5. () SECRETARIAL - CLERICAL							
6. () OTHER							
TOTAL SALARIES AND WAGES (A + B)						57,869.	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS) @ 30%						6,878.	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						64,747.	
D. PERMANENT EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$1,000.)							
TOTAL PERMANENT EQUIPMENT							
E. TRAVEL 1. DOMESTIC (INCL. CANADA AND U.S. POSSESSIONS)						6,000.	
2. FOREIGN							
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ _____							
2. TRAVEL _____							
3. SUBSISTENCE _____							
4. OTHER _____							
(5) TOTAL PARTICIPANT COSTS						4,000.	
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						1,000.	
3. CONSULTANT SERVICES							
4. COMPUTER (ADPE) SERVICES							
5. SUBCONTRACTS Internet Multicasting Service						83,000.	
6. OTHER Tuition Remission @51% of Grad Student Services						17,822.	
TOTAL OTHER DIRECT COSTS						101,822.	
H. TOTAL DIRECT COSTS (A THROUGH G)						176,568.	
I. INDIRECT COSTS (SPECIFY RATE AND BASE) @55.5 MTDC (71,747)							
TOTAL INDIRECT COSTS						39,820.	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						216,388.	
K. RESIDUAL FUNDS (IF FOR FURTHER SUPPORT OF CURRENT PROJECTS SEE GPM 252 AND 253)							
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						\$216,388.	\$
PI / PD TYPED NAME & SIGNATURE* Ajit Kambil				DATE July 5, 1993	FOR NSF USE ONLY		
INST. REP. TYPED NAME & SIGNATURE*				DATE	INDIRECT COST RATE VERIFICATION		
					Date Checked	Date Of Rate Sheet	Initials - DGC

7.3 Budget Explanation

Personnel: One month summer salary for the first year and 1/3 academic year salary for the second year is requested by the principal investigator. Support for one principal graduate assistant on a full-time basis and two other graduate students on a part time basis (0.5 FTE) is requested for two academic years and the summer.

Wage increases are budgeted at 5% per year. The current rate of fringe benefits is 29%, increasing by 1% per year. Graduate student tuition is budgeted at 51% of salary.

Equipment:

The Principal Investigator has a Mac Quadra 700 as his primary machine provided by the academic institution. We request an IBM compatible Intel 486, machine with 200 MB of disk, 20 MB of RAM, with a 16 " color monitor. Given the dominance of the Windows and DOS environments, this will be the primary platform for developing and testing Window's compatible user interfaces to EDGAR data. A 1 Gigabyte SCSI hard disk is required for local storage and a hublet is also required to enable multiple machines to connect to the network.

In addition a licence for WAIS server software for a New York University Sparcstation is required to provide a local WAIS server site for testing WAIS vocabularies on sample data from Edgar.

Travel:

Funds are requested for the Principal Investigator and students to visit the SEC and the Internet Multicasting Service in Washington D.C for project coordination and work at least three times a year. In addition the Principal Investigator requests travel funds to the International Conference on Information Systems (Florida '93 and Vancouver '94) and INET and National Net Conferences to disseminate research results and reports.

Other Direct Costs

We have included a small sum for printing to primarily support notification of business school faculty of the availability of new data on the Internet for research as well as dissemination of preliminary documentation for access and research results.

Money was also allocated for a project oversight committee to visit and advise the principal investigators and their team on the progress of the

project. This money will enable up to 5 committee members to visit New York from out of state twice a year for two days.

Internet Multicasting Service - Fixed Price Contract

Internet Multicasting Service will provide data processing and maintain the Sparcstation and disk farm to support the project. In addition they will provide initial conversion and formatting of the data, maintain a server capable of 1 Mbps applications level throughput, and implement multicast based data distribution. They will also provide technical assistance for the project and participate in project meetings, reporting and other activities as per the letter of commitment attached to this proposal.

Other Costs

Total NSF costs do not include cost-sharing contributions. New York University will cost share in accordance with NSF policy and contribute graduate student support valued at \$40,000. The fixed price subcontract from Internet Multicasting Service includes cost sharing of Internet access facilities, hardware, and personnel valued at \$55,000. Additional cost sharing of \$138,000 per year will be contributed by other organizations for the Edgar Dissemination Service data feed.

8. Current and Pending Support

Ajit Kambil

None