

Public Access Web Information Systems: Lessons from the Internet EDGAR Project

Improving the dissemination of government data helps keep the lines of communication open between corporations, governments, and the private sector.

Governments collect, generate, and disseminate vast amounts of information. In a market-driven, democratic society, this information is vital to enhance the trust of citizens in their government and institutions, and critical to individual and organizational decision making.

To address this issue we examine the lessons learned from the EDGAR on the Internet (EOI) project, an early demonstration Web information system (WIS) for disseminating corporate disclosure documents filed with the U.S. Securities and Exchange Commission (SEC). We also present a framework for WIS strategic directions.

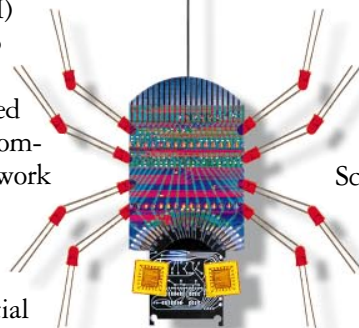
The SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system is responsible for regulating financial issuers and market makers, and mandating various financial disclosures critical to investor decision making. Developed in 1983, piloted in 1984, and fully phased in on May 6, 1996, the modem-based, non-Internet EDGAR system electronically receives all major corporate disclosure forms, and provides various dissemination mechanisms. EDGAR was introduced to improve the public's access to disclosure documents and the SEC examiners' substantive review of filings. Today, the system receives over 60GB of text data annually from 16,000 filing entities (corpora-

tions, mutual funds, major shareholders, directors, and managers) that file over 300,000 filings. Timely access is critical to investors and the general public.

The EOI initiative sought to provide widespread access to these government documents and demonstrate the Internet as a viable low-cost information distribution mechanism for public data access. Conceived by Carl Malamud, a leading Internet researcher, this project was sponsored by a two-year National Science Foundation (NSF) grant awarded in November 1993 to New York University's Stern School of Business in collaboration with Malamud's Internet Multicasting Service. This project was transferred to the SEC in October 1995, and today the SEC Web site (www.sec.gov/edgarhp.htm) transfers 400,000 filings daily, making it one of the most popular Internet sites.

Our team focused on developing value-added user access to filings such as:

- Specialized CGI scripts such as the selection of a specific form type for a specific company, within a limited date range or a report on all 5% or greater ownership of a filing company.
- Full-text search provided on the company annual



report to shareholders, Form 10K, using the Excite search engine.

- Automatic data extraction to create value-added reports such as 5% ownership by potential acquirers along with the name of the target companies.
- Aggregating and adding data where we mapped a list of 8,000 ticker symbols to their corresponding SEC Central Index Keys and company name.
- Client-side tools such as a Java applet (see Figure 1) to intelligently migrate value addition to client software. The applet provides a single entry point to all common NYU retrieval tools, allowing the regular EDGAR user to search the latest filings without having to make repeated search requests of the server.

Challenges and Lessons Learned

Interfacing to the legacy EDGAR application is the primary challenge to automatic value addition. Some of the obstacles include:

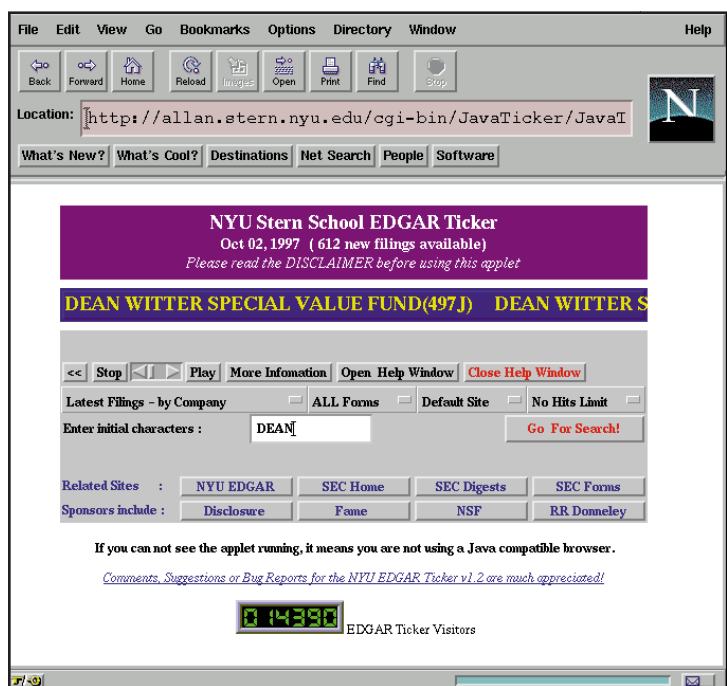
Incomplete information or data within legacy applications limits the ability of WIS designers to add value to data by automatically extracting key information. For example, many EDGAR filers are given the option to provide an electronic filing where key portions are incorporated by reference to previously filed documents. Filings may also miss vital header data such as the company's standard industrial classification (SIC) code which must be

corrected at the point of dissemination or use.

Poor identification of data objects. A key SEC priority was to create incentives and reduce the efforts of firms to comply with EDGAR-based filing, that is, modem filing versus tape or paper. Therefore, the SEC limited the number of items to be tagged to few interesting objects (for example, the financial accounting data table). Until the EDGAR systems is reengineered with a fuller and standard set of semantic tags strictly enforced by the SEC, it is a difficult technical process to extract information. For example, there is no consistent application of column tags, so a number like 1991 is ambiguous: it might be a value associated with a line item or it might be a year in a column heading. Even when a parsing software application is run against an original filing, the danger of creating errors in an extracted filing is introduced where none existed originally.

Reporting errors. Automatic extraction and data reporting can also lead to widespread dissemination of erroneous data. Initial filings are often inaccurate in reporting key accounting data, motivating the review by SEC examiners. In this case, the company must file an amendment in a timely manner (for example, a form 10-K/A amends a 10-K). There are presently no routines that automatically check the consistency of financial information in a document before they are disseminated. Thus, users and value adders should monitor for data amendments and the SEC should check data prior to publication and dissemination in a WIS.

Figure 1. Java applet interface (see edgar.stern.nyu.edu)



Terminology and nomenclature issues. WISs allow inexpensive publication of information, often reaching new user groups unfamiliar with specialized nomenclature. For example, the SEC company conformed name (CCN) and a company's common name in the marketplace are often different, confounding users' Web queries.

In addition to file names, the legal language of the filings is often confusing. Specialized legal or investment handbooks are not easily available; therefore individual investors often find these filings opaque. For example, if a company recently announced a major factory closing, a savvy EDGAR user would look at the 8-K (current report filing) whereas a less successful technique would be to wait for the next full quarterly accounting statement in a 10-Q.

Our experience with EDGAR suggests that value addition and public access to gov-

ernment data will be improved substantially if the government reengineers its methods of collecting and generating public data. Specifically, we encourage the SEC to investigate WIS for creating electronic disclosure filing mechanisms integrated with the workflow of disclosing organizations. The

SEC and other government organizations should also enforce stronger standards for data quality and specify tags to classify objects of interest in the data disclosed to the public. The Extensible Markup Language (XML) initiative to specify data formats for structured document interchange on the Web shows promise as a means of implementing such tags in documents at the point of document creation [4]. Such efforts reduce the costs of extracting and using relevant data from documents, help ensure data quality, and enable easier migration of the information to higher value-added products.

As government information becomes more accessible to the public, more effort is also required to simplify information where possible so that it is comprehensible to interested citizens. Reducing information overload (see Turoff and Hiltz in this section) is of paramount importance. In the case of EDGAR, such efforts should be closely integrated with efforts to specify standards to mark-up key data within documents submitted to the SEC.

Value Migration and Implications for Competition

In the market for disclosure documents, the Web and EOI have radically reduced the cost of disseminating EDGAR data to users. This diminishes the relative value of timely and convenient access to EDGAR documents as a key source of product differentiation. Previously the public often purchased SEC filing information through corporate resellers. As Lederer, Mirchandani, and Sims mention on the next page, vendors must provide ways to enhance their competitiveness. In EDGAR's case, market incumbents and new entrants are using WIS to modify and differentiate their products on different dimensions to create competitive advantage.

Figure 2 summarizes three strategic directions for information vendors in Web-enabled environments:

Table I. Some EDGAR resources on the Internet.

Site	Description
www.sec.gov/edgarhp.html	SEC EDGAR
Edgar.stern.nyu.edu	NYU EDGAR
www.motleyfool.com	Motley Fool Investors' Forum
www.disclosure.com	Disclosure, Inc.
bamboo.tc.pw.com/edgarscan.html	Price Waterhouse EDGARSCAN
www1.freeedgar.com/home.htm	FreeEDGAR
www.moody.com/fis/source/proddesc/edgarept.htm	Moody's EDGARReport
www.hottools.com	HotEDGAR Tool for Excel Conversion
www.edgar-online.com	EDGAR Online

from *technology leverage* to *intelligence leverage* (Y-axis), public data to proprietary data aggregation (X-axis), and a generic to a personalized and, ultimately, a community interface (Z-axis).

Technology leverage strategies [9] differentiate products by exploiting technology innovations to improve the performance of processing, storage, communications, interface, and programming technologies. For EDGAR these strategies include timely notification and document delivery, full text search, automatic data extraction and reporting, and data analyses.

The major commercial vendors such as Disclosure, Moody's, and new Internet entrants such as Edgar Online and Internet Financial Network leverage the Web to provide both alert mechanisms and timely document delivery for a fee. FreeEDGAR leverages the low-cost telecommunications and publishing capabilities of the Internet and Web to disseminate same-day EDGAR data free of charge.

A second technology leverage strategy is to add structure, format, and extract valuable data from filings. Disclosure and other for-fee vendors provide well-formatted filings. Financials Online, a small entrepreneurial venture, provides HOTEDGAR, that consists of software tools for parsing financial data from tables in the filings and reformatting it for use in Microsoft Excel spreadsheets. Similarly, Price Waterhouse's Edgarscan employs multiple parsing strategies to automatically extract major textual and numerical sections of filings. As discussed earlier, several difficulties inherent in the structure of the filings limit the effectiveness of automatic extraction of data from electronic filings. Thus, no method consistently extracts the data correctly without some human intervention.

Third, vendors can enable better search (full-text, for example) and value-added reporting of the data. As data warehouse software and online analytic processing tools for the Web improve, well-structured EDGAR data can

be presented to users in multiple ways to support peer group comparisons and other data analysis.

Technology leverage strategies as illustrated here are of limited strategic value as technology costs decline. These strategies are easily observed in the marketplace and generally easily imitated by well-financed competitors. As Web development and information dissemination tools improve, the entry costs for competitive vendors offering similar products continues to fall, eroding profit margins and making products differentiated solely by technology leverage strategies a low value commodity.

In contrast intelligence leverage strategies create new products and services based on applying intelli-

gence to data. These can be proprietary reports based on human interpretations or non-observable methods applied to the data. One example is Disclosure's construction of Insider Trading Indexes. This product aggregates information on a weekly basis from 5,000 insider trade SEC filings, then combines it with a proprietary weighting system and expert analysis to gauge executive sentiment across industry sectors and regions. Such a ratings system depends on the design of the rating algorithm, and human judgment to establish ratings with adequate predictive power to correctly signal market directions. If the rating establishes a reputation, Disclosure's proprietary product will be harder to replicate unless a

Using WISs to Enhance Competitiveness

Albert L. Lederer, Dinesh A. Mirchandani, and Kenneth Sims

What benefits do companies seek from their Web-based information systems? The rush to the Web has been enthusiastic. But what's motivated it?

To answer these questions, we emailed a survey to 846 companies with entries on the Internet Mall, an electronic shopping center. Of these, 25% percent (212 companies) answered, choosing from 1 to 7 (1 being very little and 7 very much) on Likert-type scales to indicate the importance of each of 33 commonly acknowledged benefits of computer-based information systems in their WIS implementation decision.

Table 1 lists the potential benefits and the subjects' ratings. It suggests ways for improving a WIS. We focus on the top benefit and we invite readers to consider their own interpretations.

In almost everywhere in our society, competition is seen as a core driver. Thus, it is no surprise that enhanced competitiveness tops the benefits sought from a WIS. If a WIS is to enhance competitiveness, then

it should support the traditional means by which companies compete: by lowering costs, focusing on groups of customers (such as, providing products that fulfill the needs of a select few buyers in a market), or differentiating their products and services from those of their competitors.

A WIS could lower costs by having customers rather than employees enter order information. It could eliminate sales people and retail outlets. It could replace expensive, proprietary electronic data interchange (EDI) systems that currently connect manufacturers to suppliers, retailers to wholesalers, and financial institutions to other institutions. Finally, it could better support using automated agents to search sites for valuable, but not explicitly requested information that can lower business costs. Ongoing efforts to improve ease of access, increase data transmission rates, enhance security, and provide more flexible development tools will facili-

tate these uses of WISs and enhance competitiveness.

To help companies focus on potential customers, the WIS should facilitate the individualized tailoring of outputs to users based on their recent purchases, information requests, and voluntary demographics. (As Kambil and Ginsburg explain here) The system should also develop mailing lists based on such criteria, and could use automated agents to search newsgroups for potential customers.

Companies may be challenged to deliver messages about products and services that encourage the fidelity rather than the resentment of these customers because customers may regard them as spam. In general, however, to enhance competitiveness, a WIS needs to facilitate the delivery of information while requiring users to make fewer searches and mouse clicks.

Finally, the WIS should help companies show how their products and services differ from their competitors' by

rating system with better predictive power and reputation can be established. This would be a costly undertaking by competitors, and illustrates a resource-based strategy [2]. Barney notes a firm's resources that are valuable, rare, inimitable, and nonsubstitutable provide the basis for sustained competitive advantage. Intelligence leverage strategies create such assets.

As EDGAR documents become a low-cost, low-margin item, EDGAR data vendors are including their product as part of more complex offerings, *data aggregating* EDGAR with proprietary information. Intellectual property protections such as copyright, patents, and trademarks make certain types of infor-

mation proprietary to owners, preventing others from consuming the information without purchase or owner's permission. Aggregating public information with proprietary information allows the aggregate product to be proprietary. This enables vendors to create unique products not easily replicable by competitors.

This data aggregation strategy is exemplified by Disclosure's combination of SEC data with the Articles Online collection of two million articles on companies and products from various newspapers, journals, and other sources. Aggregation of different proprietary data sources and integration with the nonproprietary EDGAR data enhances convenience

using standards for high-quality presentation. In fact, independent product comparisons could help consumers make better informed decisions and thus enhance competitiveness. Agents could accumulate real-time product and service information. Users would then select vendors and criteria, and display the information in an easily comparable format. This would facilitate differentiation and enhance competitiveness. An independent group similar to the Consumers Union or Better Business Bureau might develop and monitor such policies.

Companies are rushing to build their WISs to enhance their competitiveness. By concentrating on the traditional bases of competition, they might accomplish that goal. Consideration of the additional benefits in Table 1 might also help them use their WISs more effectively. **C**

ALBERT L. LEDERER (lederer@ukcc.uky.edu) is a professor at the University of Kentucky.
 DINESH A. MIRCHANDANI (dinesh0@sac.oku.edu) is an assistant professor at Grand Valley State University in Michigan.
 KENNETH SIMS (kenpc@cc.cumber.edu) is an assistant professor at Cumberland College in Kentucky.

Benefit	Mean
Enhance competitiveness or create strategic advantage	5.34
Enable easier access to information	5.24
Provide new products or services to customers	4.88
Increase the flexibility of information requests	4.68
Improve customer relations	4.66
Enhance the credibility and prestige of the organization	4.57
Provide better products or services to customers	4.49
Increase volume of information output	4.46
Align well with stated organizational goals	4.24
Enable the organization to respond more quickly to change	4.23
Enable faster retrieval or delivery of information or reports	4.21
Help establish useful linkages with other organizations	4.19
Save money by reducing communication costs	4.09
Change the way the organization conducts business	4.01
Increase return on financial assets	3.93
Enhance employee productivity or business efficiency	3.90
Speed up transactions or shorten product cycles	3.84
Improve the accuracy or reliability of information	3.70
Present information in a more concise manner or better format	3.55
Enable the organization to catch up with competitors	3.26
Allow previously infeasible applications to be implemented	3.06
Improve management information for strategic planning	3.05
Improve information for management control	2.89
Improve information for operational control	2.79
Allow other applications to be developed faster	2.78
Provide the ability to perform maintenance faster	2.47
Save money by avoiding the need to increase the work force	2.46
Save money by reducing travel costs	2.31
Save money by reducing the work force	2.25
Save money by reducing system modification or enhancement cost	2.24
Save money by reducing hardware use	1.85
Provide greater data or software security	1.82
Facilitate organizational adherence to government regulations	1.52

Table 1. Importance of each benefit sought from a WIS

and value by reducing their efforts in searching and acquiring related data.

Data aggregation also allows vendors to sell subscriptions more easily and reduce the buyer's adverse selection risks and transaction costs. The buyers confront adverse selection risks, as they often cannot know the relevance or quality of the information product without consuming it. By bundling multiple branded sources of information, the vendor increases the buyer's likelihood of finding information of value within the bundle. Thus, data aggregation strategies generally create more attractive products for customers and provide vendors with new dimensions for product differentiation. Data aggregation also creates new barriers to entry in the industry that arise from the high fixed costs of setting up a dissemination system for aggregate data: licensing content from multiple sources and effectively integrating the data into products. At the same time the low cost of communications and Web development tools facilitate ongoing data aggregation from multiple sources and dissemination of data, allowing the vendor to take advantage of economies of scale.

Over time, we expect information vendors on the Web to migrate their value toward more intelligent aggregation strategies that combine data aggregation with analysis and useful new signals on data generated by human judgment. These strategies, while costly to develop and implement, will also remain hard for competitors to effectively replicate.

WISs enable information vendors to transform interactions with customers to differentiate their products and services. Here, we identify three levels of interaction.

Generic interactions apply common user interfaces to all or large subsets of users. The EOI project generally provided generic interfaces to the EDGAR database with limited customization. In contrast, the Java Applet enabled a **personalized interaction** allowing users to define their preferences and presenting documents that matched their preferences first for review. This creates value by reducing the user efforts to search for documents on an ongoing basis. It also introduces a switching cost for users, who incur a cost to personalize their interface, and are thus unlikely to switch. Most for-pay EDGAR sites are adopting personalized interfaces for their clients.

Community interaction strategies leverage the Web to allow users with shared interests to interact. Some EDGAR vendors provide EDGAR data to various virtual communities on investment and finance. For example, EDGAROnline provides data to The Motley Fool (www.motleyfool.com), a popular virtual community for investors. The Motley Fool hosts a

number of bulletin boards that enable users to interact and discuss specific companies, investments, and disclosures such as earnings reports by companies.

There are two main advantages to information vendors implementing community interaction strategies. First, it generates stocks of new, unique and proprietary content that is hard for competitors to replicate. Second, as noted by Hagel and Armstrong [3], virtual communities can potentially create significant returns to scale in content production, reputation, and revenues for community organizers. As a site becomes more useful it can generate more subscribers who, in turn, can contribute new useful content, generating positive externality benefits. This creates additional opportunities for advertising or transactional revenue.

While there are many different ways to add value to data (see [8]), the three directions identified here succinctly capture the key value-addition and differentiation choices confronted by information vendors in markets where competitors apply WISs to deliver products to customers. While the low-cost communications infrastructure of the Internet has made access to SEC data a free commodity to customers, it has also provided information vendors with opportunities to create new products that better serve existing customers and reach new customers to expand markets.

Implications for Public Data Access

Government information is a valuable resource to the public for decision making, and a profitable resource for many information vendors. Beginning in the 1980s, the U.S. government increasingly privatized the dissemination of various types of electronic government information. (See [5, 7].) Office of Management and Budget Circular A-130 orders Federal agencies to try and recover the "cost of dissemination" and not "expend public resources filling needs which have already been met by others in the public or private sector." However, the 1993 version of A-130 provides government agencies with a means of directly disseminating data "if legal consideration requires an official government dissemination product."

The Internet and Web dramatically reduce the cost of dissemination to government and industry. While vendors provide free EDGAR systems that are substitutes, there are several reasons why the SEC and other government should provide information directly over the Internet. First, access to the SEC site continues to grow, illustrating demand for government dissemination from the official source. Second, only the government site can maintain a free and exhaustive repository without access charges to

the public in consistent formats. This enables specialized entrepreneurs to build diverse products and services. Third, similar public and private sector sites are not mutually exclusive on the Internet. The public site adds greater choice of information, and provides a backup to free resources elsewhere in the event of technical or network failure. Fourth, the interactive capabilities of the Internet

allow the agencies to get valuable feedback on the types of information collected and disseminated, and ways of maximizing the value of filings to customers.

The falling costs of dissemination suggest that the public data access debate should not center on whether the government should implement Internet dissemination systems; rather, how much added value should government owned or operated Internet sites provide? We believe government sites should at a minimum provide timely data with full-text indexing and retrieval of text documents. In the case of EDGAR, more timely access to documents (a 20-minute delay, as is common for stock price information) should be combined with search methods for documents indexed by ticker symbols, SIC code, and full-text search. We expect information vendors who provide value-added products will migrate their strategies toward proprietary data aggregation, intelligence leverage, and community strategies. Thus we do not see adverse impacts on innovations or markets in the information industry if governments leverage the Web to provide generic public interfaces to aggregate or single sets of government data through currently inexpensive and widely available software technologies. These services would effectively be limited to the bottom left quadrant on the cube in Figure 2.

Conclusion

EDGAR on the Internet clearly demonstrates the capacity of the Internet to effectively disseminate government information in a timely and inexpensive way. As tools for WIS development and the infrastructure for data dissemination improve, government-owned Web systems will satisfy the basic needs of the public for access to key government informa-

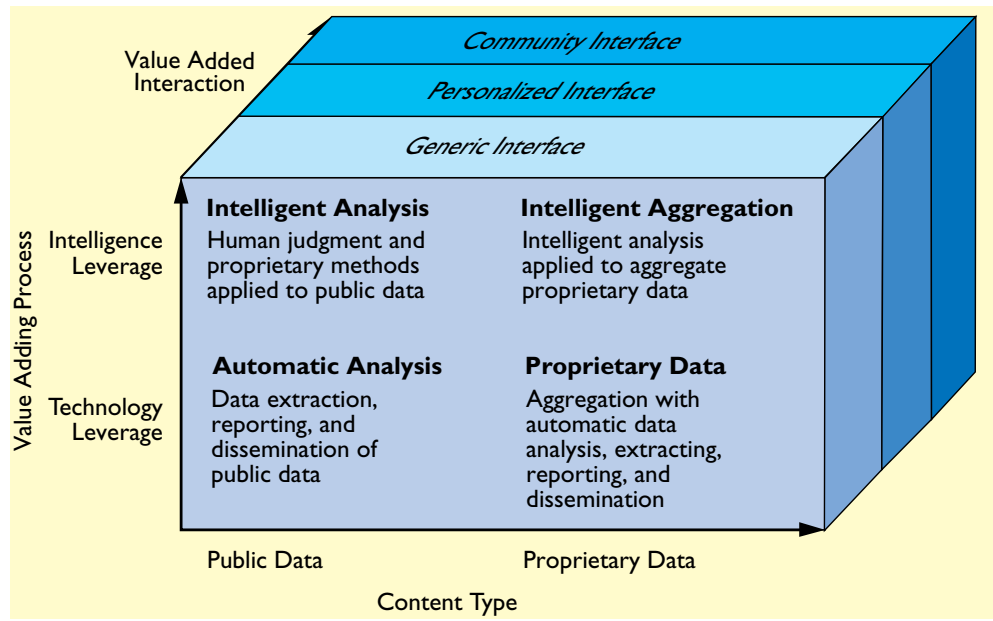


Figure 2. WIS-enabled information vendor strategies

tion. As WIS capabilities increase, data vendors will have to migrate value from simple value addition mechanisms that leverage the features of technology to competitive models which leverage intelligence applied to select aggregate bundles of data, customized to individual users and communities. These transformations will enhance customer value and create new markets for information vendors. **G**

REFERENCES

1. Akerlof, G. The market for lemons: Quality, uncertainty and the market mechanism. *Q. J. Econ.* 84 (1970), 488–500.
2. Barney, J. Firm resources and sustained competitive advantage. *J. Manage.* 17 (1991), 99–120.
3. Hagel, J. and Armstrong, A. *Net Gain: Expanding Markets through Virtual Communities*. (1997) Harvard Business School Press, Cambridge, Mass.
4. Khare, R and Rifkin, A. XML: A door to automated Web applications. *IEEE Internet Computer*. (May–June 1997).
5. Love, J. The marketplace and electronic government information. *Gov. Pub. Rev.* 19 (1992), 397–412.
6. Love, J. Pricing government information. *J. Gov. Info.* 22, 5 (1995).
7. Starr, P., and Corson, R. Who will have the numbers? The rise of the statistical services industry and the politics of public data. *The Politics of Numbers*. W. Alonso and P. Starr, eds. (1987) Russell Sage, New York.
8. Taylor, R. *Value Added Processes in Information Systems*. (1996) ALEX Publishing Corp., Norwood, N.J.
9. Venkatraman, N. and Kambil, A. The check is not in the mail: Strategies for electronic integration. *Sloan Manage. Rev.* (Winter 1991).

The EOJ Project was funded by NSF Grant Number 9319331, R.R. Donnelley and Sons, Disclosure, Inc., and Fame Information Services.

An expanded version of this article is available as a New York University Stern School of Business, Information Systems Department Working Paper, no. IS-98-13.

AJIT KAMBIL (akambil@stern.nyu.edu) is an assistant professor in the Stern School of Business at New York University.

MARK GINSBERG (mark@edgar.stern.nyu.edu) is a doctoral candidate in the Stern School of Business at New York University.