The Information Content of Corporate Websites

Bradford Lynch-Levy and Daniel Taylor

NBER Big Data, December 10, 2021
An efficient and economical method for companies to make information available about themselves to many investors is through their Internet websites.

"-

Securities and Exchange Commission
(Release 33-8128, 2002)
Motivation

- The disclosure literature is vast

- The information effects of SEC filings, management forecasts, analyst reports, and press releases are well documented

- A survey of the literature reveals that comparatively little is known about how firms and shareholders use corporate websites
Background - Regulation

- In 1995, the SEC highlighted the growing prevalence of the PC in households and the low cost of electronic communication*

- In 2000, the SEC recommends all firms use websites for investor relations**

- In 2002, the SEC asks all firms to
  1. Disclose their web address in their annual report
  2. Firms without a website must explain why they do not have one***

- As of August 7, 2008, the SEC allows companies to substitute website disclosures for 8-K filings on EDGAR****

Background – Adoption of websites

Dow’s website circa 1996, archived by Alexa Internet.
Background – Adoption of websites

Dow’s website circa 1996, archived by Alexa Internet.
But are websites more than EDGAR?

“This code of ethics and the Company’s corporate governance policies are posted on the Company’s website at http://www.JohnDeere.com.” (Deere & Co 2003 10-K)

“We make our website content available for information purposes only. It should not be relied upon for investment purposes, nor is it incorporated by reference into this Form 10-K.” (Telos 10-K, 2003)

“In addition, the Company has posted the charters for its Audit Committee, Compensation Committee, and Nominating and Corporate Governance Committee, as well as the … on its website.” (Ansys’s 2018 10-K)
But are websites more than EDGAR? - Kiora

➤ Kiora (Formerly EyeGate) never posts their earnings announcements to EDGAR. Instead, they keep records of the PRs on their website.
But are websites more than EDGAR? - Ansys

- 10+ presentations containing hundreds of slides by top management were posted to Ansys’s website and not mentioned elsewhere.
Research Questions

- Do firms use their corporate website to provide pertinent new information to investors?

- If so, what are the broad topics of that information?
Contribution

1. **Investor relations**
   - Corporate website disclosure is economically significant
   - Websites contain *more* business operations content than EDGAR

2. **Information processing costs**
   - Websites alleviate information asymmetry between investors

3. **Effects of new technologies on markets**
   - Websites are largely unregulated but have significant capital market effects

These results hold across a variety of research designs, identification strategies, and cross-sectional tests
Related Literature on Corporate Websites

- Boulland et al. (working paper) construct a **quarterly** measure of website content using data from the Internet Archive (IA)
  - Find it is positively correlated with the provision of earnings forecasts, EDGAR filings, and EDGAR file size

- The IA, or Wayback Machine, scrapes webpages and makes historical archives available to the public

- **Caution:** IA endogenously chooses which pages to scrape and when to scrape them based on internet attention
  - IA data poses key identification challenges
Related Literature – Celcuity Inc.

Saved 214 times between July 24, 2014 and November 27, 2021.

Note
This calendar view maps the number of times celcuity.com was crawled by the Wayback Machine, not how many times the site was actually updated. More info in the FAQ.

Note
This calendar view maps the number of times celcuity.com was crawled by the Wayback Machine, not how many times the site was actually updated. More info in the FAQ.
Related Literature – Cel cuity Inc.

Note
This calendar view maps the number of times cel cuity.com was crawled by the Wayback Machine, not how many times the site was actually updated. More info in the FAQ.
Related Literature – Celcuity Inc.

Bubble size represents count of scrapes that day
Sample Selection

- We collect daily observations of all corporate websites listed in Compustat every day at 9:15AM from July through October, 2019
  - Minimal selection bias
  - Includes the homepage and all inner pages

- Sample is ~150 Terabytes

- We collect data on control variables from Compustat, CRSP, RavenPack, SEC EDGAR, and NYSE TAQ

- 220,505 firm-date observations
Measuring Website Content

- In principle, there are numerous ways to measure content
  - Text, images, videos, tables, external data referenced via hyperlinks
Measuring Website Content

- In principle, there are numerous ways to measure content
  - Text, images, videos, tables, external data referenced via hyperlinks

- We measure the length of the document, the number of HTML formatting tags, links, rich content, and dynamic elements

<table>
<thead>
<tr>
<th>Variable</th>
<th>N-obs</th>
<th>Mean</th>
<th>Std</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>220,505</td>
<td>2791.14</td>
<td>4335.93</td>
<td>1773</td>
</tr>
<tr>
<td>Formatting</td>
<td>220,505</td>
<td>372.84</td>
<td>381.47</td>
<td>281</td>
</tr>
<tr>
<td>Links</td>
<td>220,505</td>
<td>123.59</td>
<td>144.95</td>
<td>89</td>
</tr>
<tr>
<td>Rich</td>
<td>220,505</td>
<td>32.92</td>
<td>71.84</td>
<td>16</td>
</tr>
<tr>
<td>Dynamic</td>
<td>220,505</td>
<td>20.44</td>
<td>18.72</td>
<td>16</td>
</tr>
</tbody>
</table>
We construct a composite measure, \textit{WebContent}, using factor analysis. The first principal factor explains 92\% of the variation in our five proxies for content, loads positively on all measures, and is the only factor with an eigenvalue greater than unity.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Proportion of the variation explained</th>
<th>Cumulative proportion of the variation explained</th>
<th>First Factor Loading Variable</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1\textsuperscript{st}</td>
<td>2.28</td>
<td>92.4%</td>
<td>92.4%</td>
<td>\textit{Length}</td>
<td>0.527</td>
</tr>
<tr>
<td>2\textsuperscript{nd}</td>
<td>0.16</td>
<td>6.6%</td>
<td>99.0%</td>
<td>\textit{Formatting}</td>
<td>0.948</td>
</tr>
<tr>
<td>3\textsuperscript{rd}</td>
<td>0.03</td>
<td>1.0%</td>
<td>100.0%</td>
<td>\textit{Links}</td>
<td>0.845</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>\textit{Rich}</td>
<td>0.471</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>\textit{Dynamic}</td>
<td>0.833</td>
</tr>
</tbody>
</table>
Daily Event Study

- Focus on “events” where the change in WebContent is in the top tercile and there are no EDGAR filings or PRs in the [-1,+1] window around the change.
- Examine changes in capital market outcomes in a tight window around the event.
- Daily nature of the data allows for firm-quarter fixed effects.
  - Tight event study design to test whether websites contain information content incremental to SEC filings and PRs.
Results

\[ \text{Outcome}_{i,t+1} = \alpha + \gamma \cdot \text{Controls}_{i,t} + \varepsilon_{i,t} \]

\[ Plot_t = \frac{E[\varepsilon_t] - E[\varepsilon]}{\text{Var}(\varepsilon)} \]

Absolute Returns

Trading Volume
Results

\[ \text{Outcome}_{i,t+1} = \alpha + \gamma \cdot \text{Controls}_{i,t} + \varepsilon_{i,t} \]

\[ \text{Plot}_t = \frac{E[\varepsilon_t] - E[\varepsilon]}{\text{Var}(\varepsilon)} \]

Amihud Illiquidity

Effective Spread
Dissemination Versus New Information

- It is possible that corporate websites merely serve as a dissemination tool rather than providing new information.
- To distinguish between the two roles, we re-estimate our previous models while varying the width of the non-disclosure window.
  - Require that there are no SEC filings or PRs in the \([-s, +s]\) window around the change in website content.
  - Where \(s \in \{5, 10, 15\}\).

\[
\Delta \text{Content} = \begin{cases} 
\text{DailyPR} = \text{DailyEDGAR} = 0 & \text{if } s \in \{5, 10, 15\} 
\end{cases}
\]
### Dissemination Versus New Information

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EventDay[0,10]</td>
<td>Coeff. on</td>
<td>t-stat on</td>
</tr>
<tr>
<td>Excluding observations with disclosures [-5,+5]</td>
<td>0.60***</td>
<td>(5.82)</td>
</tr>
<tr>
<td>N-obs = 444,446, N-events = 7,214</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excluding observations with disclosures [-10,+10]</td>
<td>0.63**</td>
<td>(5.00)</td>
</tr>
<tr>
<td>N-obs = 307,112, N-events = 4,987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excluding observations with disclosures [-15,+15]</td>
<td>0.53***</td>
<td>(3.44)</td>
</tr>
<tr>
<td>N-obs = 228,548, N-events = 3,712</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Controls: Yes
Date FE: Yes
Firm-Year-Qtr FE: Yes
Types of Content in Corporate Websites

- We classify changes in WebContent according to file type, e.g., Office documents, PDFs, and engineering drawings are Productivity related.
- Productivity related files make up roughly 10% of the content in WebContent changes.
Types of Content in Corporate Websites

- We classify changes in WebContent using the topic model of Dyer et al. (2017) and compare to EDGAR.
Conclusions

- Using a tight event study design, we find that websites contain information content incremental to traditionally studied channels.
- Notably, the market reaction to website changes is of similar economic significance as SEC filings and press releases.
- Additionally, websites contain more business operations content than EDGAR and the market reaction to such information is heightened.