

Digital-Resource Disclosure and Subsequent Firm Outcomes: A Resource-Based Analysis of U.S. Incumbents

Haoyun Liu*

Litai College of Languages and Cultures, Shanghai International Studies University, Shanghai, China

**Corresponding author: Haoyun Liu.*

Abstract

Annual-report language may record a digital initiative, an aspiration, or a risk before a firm realizes an operating change. This paper studies whether digital-resource disclosure in Form 10-K filings is associated with subsequent outcomes for U.S. incumbents. The SEC source screen retains 39 issuers and 273 filing records; outcome availability yields 133 primary-ROA firm-years across 23 issuers from 2017 through 2022. A predeclared, length-normalized dictionary is joined to next-year SEC Company Facts and examined with controls, alternative dictionaries, a governance-language placebo, and firm-cluster bootstrap intervals. The full-dictionary ROA estimate is small and its bootstrap interval spans zero; the other outcome estimates are likewise imprecise. Adding disclosure to a controls-only held-out linear comparison does not improve mean absolute error. The contribution is a transparent filing-provenance protocol, reproducible public-data audit trail, and association-based account of disclosure and subsequent outcomes, not a causal estimate or a new prediction method. The findings support a cautious, outcome-specific interpretation.

Keywords

digital-resource disclosure, resource-based view, incumbent firms, digital transformation, firm outcomes

1. Introduction

Digital transformation concerns changes in work processes, organizational arrangements, and value creation, rather than the mere installation of an information system [1, 2]. For an established firm, a Form 10-K may mention cloud services, data, automation, or artificial intelligence well before related routines affect operating results. The timing is uneven across firms and functions [3, 4, 5].

Resource-based research helps explain that gap. A resource may be valuable in principle yet yield little until it is combined with the routines that make it usable; both prior assets and organizational learning shape that process [6, 7, 8]. Information-systems research similarly treats information-technology capability as a technical and managerial bundle, not a binary adoption state [9]. Incumbents therefore face a mixed inheritance: legacy resources can support digital change, but they can also limit it [10].

Filing language cannot settle whether a capability has been realized. A digital term in an annual report may refer to installed infrastructure, a proposed investment, competitive pressure, or cyber risk. The measure used here is consequently a trace of stated digital-resource emphasis in public disclosures. It does not observe implementation quality and does not identify the causal effect of an implementation decision.

The empirical question is narrower: within a public-data panel, how does Form 10-K digital-resource disclosure relate to next-year financial outcomes? The analysis estimates adjusted associations and checks the result against a narrow dictionary, a governance-language placebo, and firm-cluster bootstrap intervals. It does not treat the disclosure score as an intervention or present it as a new prediction method.

Regulatory filings offer dated, firm-specific narratives in a common repository, while their XBRL facts permit outcome matching without manual transcription. That convenience does not remove the ambiguity of phrase counts: "cloud computing," for example, can describe an operating system, an intended investment, or a risk exposure. The design retains that ambiguity rather than treating every occurrence as proof of completed transformation.

Each retained firm-year is linked to a ticker record, Form 10-K accession, cleaned filing text, SEC Company Facts observation, and source hash. The resulting provenance record exposes the issuer screen, fiscal-year mapping, vocabulary, and outcome construction. It documents how associations vary by outcome and dictionary while keeping the interpretation within what a filing-text proxy can support.

2. Related Work

Work on digital transformation joins organizational change, strategy, and dynamic capabilities. Field reviews stress that organizational design and complementary capabilities affect the returns associated with a digital initiative [3, 4, 5]. A resource-based reading is useful for that reason, but it offers no basis for assuming that every digital disclosure should precede a common financial return.

Several empirical studies draw on U.S. 10-K text. Zareie et al. form a digital-transformation score from terms in annual reports and relate it to corporate value; Gaspar et al. combine a 10-K-based measure with earnings-call evidence to examine performance and resilience during shocks [11, 12]. The panel here is smaller and purposively selected, so it is not a substitute for those broader designs. Its value lies in the documented chain from issuer eligibility and filing identity through text processing to next-year financial facts; the accompanying replication materials enumerate that chain.

Financial-text research cautions that a dictionary imported from outside reporting can misclassify 10-K language, which makes document selection and proxy construction consequential [13, 14]. The length-normalized score, together with its narrow and placebo forms, therefore measures disclosure rather than verified implementation. Information-technology and artificial-intelligence capabilities may matter for performance when firms combine them with routines and governance [9, 15]. That argument guides the interpretation of an observable disclosure trace; it is not used to infer an unobserved capability or a causal effect.

3. Data and Method

Before downloading filing text or Company Facts, the analysis script fixes a purposive list of 41 U.S.-listed issuers. The list is neither an index membership list nor a market-capitalization ranking or probability sample; no market-representative inference follows from it. Eligibility requires a Form 10-K at or before fiscal 2013 and a usable annual filing in every source year from 2017 through 2023. Although disclosure is analysed through 2022, the 2023 filing is needed to form the next-year outcome for a 2022 disclosure. The screen retains 39 issuers and 273 source records, and outcome availability leaves 133 primary-ROA firm-years from 23 issuers for 2017-2022. Ticker records, submission records, archived 10-K HTML, and Company Facts come from public SEC endpoints. Industry groups follow the first two digits of the issuer-level SIC code in the submissions record. Every exclusion is retained in the ledger; a retrieval or fiscal-year mismatch is not recoded as an absence of digital activity.

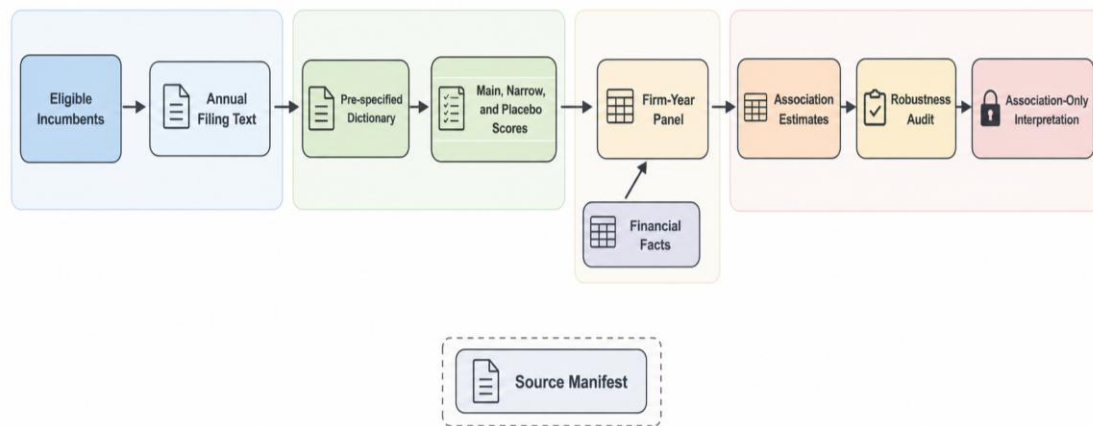
Fiscal years are matched by linking the 10-K report date to an annual Company Facts observation. The SEC fiscal label is used where available, with the report-date calendar year used only as a fallback. This prevents a

non-calendar year-end from being mislabeled, as in Johnson & Johnson's fiscal 2020 report, which ended in January 2021. Disney's 2013 legacy-CIK filing is checked for the incumbency screen alone, while source-year records remain attached to the current ticker CIK. Each retained source record carries its CIK, accession, filing URL, retrieval time, and content hash. The supplementary replication package contains the candidate list, screening ledger, source manifest, phrase dictionaries, derived panel, and analysis scripts.

The exclusion ledger has two cases. Disney satisfies the pre-2013 rule through its legacy filing, but the continuity record lacks the required 2017 and 2018 source filings; those panel years are excluded for missing source records, not for lack of incumbent history. Dow has no retained Form 10-K at or before 2013 and fails the pre-specified incumbency rule. Neither missing filings nor missing financial facts are imputed. An unavailable filing is a data-screening fact, not evidence about a firm's digital-resource disclosure.

Figure 1 shows the provenance sequence from a fixed issuer list to archived filings, a disclosure score, and subsequently reported SEC facts. The panel is descriptive. Neither the text score nor the record linkage establishes that a firm implemented a particular technology.

Figure 1: Filing-provenance panel protocol. Public SEC issuer records and Form 10-K filings yield length-normalized digital-resource disclosure measures that are joined to next-year SEC Company Facts outcomes



For firm i in fiscal year t , the digital-resource disclosure score is

$$S_{it} = 1000 H_{it} / N_{it}.$$

where H_{it} counts case-normalized phrase occurrences and N_{it} is the number of alphabetic word tokens after markup removal. The full vocabulary contains digital transformation, digital platform, cloud computing, artificial intelligence, machine learning, data analytics, cybersecurity, enterprise resource planning, internet of things, digital technology, and information technology. The narrow score uses the first five phrases. The placebo contains board of directors, shareholder, and audit committee. Because each score comes from the same stripped filing, differences among them reflect vocabulary choice rather than a changed set of documents.

The text routine removes script, style, and markup elements; decodes HTML entities; collapses whitespace; and counts the fixed strings and word tokens. Phrases are not selected after outcomes are inspected. Raw HTML is hashed before cleaning, which separates a changed source document from a changed parsing rule. Complete filings are used because archived HTML lacks uniform section markers for Business, Risk Factors, and Management Discussion across retained issuers. A section-restricted score would be a different measurement design and is not represented as an output of this routine.

Company Facts provides three next-year outcomes when the required tags are present. Next-year ROA is next-year net income divided by average current- and next-year total assets. Next-year operating margin is next-year operating income divided by next-year revenue. Next-year revenue growth is next-year revenue divided by current-year revenue minus one. Controls are current log total assets, the contemporaneous total-liabilities-to-assets ratio $D_{it} = \text{total liabilities}_{it} / \text{total assets}_{it}$, and fiscal-year and industry indicators.

For every outcome, the extractor keeps the source tag, report end date, filing date, and reported value with the firm-year record. A row may contain enough information for ROA but lack the revenue or operating-income tag needed for an auxiliary outcome. Operating-margin analysis therefore has 97 observations, whereas

the primary ROA and revenue-growth analyses have 133. This outcome-specific treatment retains a valid ROA observation when a secondary ratio cannot be constructed.

The pooled specification is

$$Y_{i,t+1} = \alpha + \beta S_{it} + \gamma_1 \log(A_{it}) + \gamma_2 D_{it} + \delta_i + \eta_g + \varepsilon_{it}.$$

where $Y_{i,t+1}$ is one of the three next-year outcomes, δ_i is a fiscal-year indicator, and η_g is an industry-group indicator. Coefficient β is an adjusted disclosure-outcome association, not the effect of changing an implementation decision.

The secondary prediction check trains on 2017-2020 records, tunes on 2021, and tests on 2022. It compares a controls-only linear cell with the same cell plus S_{it} . Fixed-effects and tree-based cells supply context because they use different fitted inputs. The score is an observed input, not a proposed learning architecture.

There are 87 training observations, 23 validation observations, and 23 held-out test observations. Settings for the regularized and tree-based comparators are chosen from the validation year alone, using five fixed seeds. The linear comparison is deterministic across those seeds, so it is read as a small held-out accuracy check rather than a stochastic performance contest. Its purpose is to ask whether disclosure adds practical predictive information beyond public size, total liabilities to assets, year, and industry controls.

Robustness analysis estimates full, narrow, and placebo scores for next-year ROA, plus the full score for operating margin and revenue growth. It uses 250 firm-cluster bootstrap resamples with fixed seeds and percentile 95% intervals. Resampling firms, rather than rows, preserves within-firm dependence. Reported inference follows those intervals rather than conventional significance tests.

4. Results

Table 1 describes the analytic records. Disclosure scores differ across firm-years. Operating margin is available for fewer records because some next-year observations lack the required operating-income or revenue tag.

Table 1: Descriptive statistics for the analytic panel

Variable	n	Mean	SD	Min	Max
Disclosure score	133	0.302	0.263	0.040	1.722
Next-year ROA	133	0.093	0.082	-0.144	0.287
Next-year operating margin	97	0.151	0.127	-0.434	0.421
Next-year revenue growth	133	0.059	0.171	-0.484	0.716
Total liabilities-to-assets	133	0.756	0.197	0.378	1.324

The mean disclosure score is 0.302 phrase hits per thousand cleaned word tokens and the maximum is 1.722, so the measure is not driven by one zero-heavy coding rule. Mean next-year ROA is 0.093; its negative minimum shows that the panel includes loss-making next years. These values favour outcome-specific association analysis over a binary successful-transformation label.

Table 2 reports the secondary prediction check. Held-out MAE is 0.05265 for the controls-only linear cell and 0.05320 after the disclosure score is added. In this small time-ordered comparison, the score does not improve predictive accuracy. Fixed-effects and tree-based cells have lower MAE in some cases, but they are contextual comparators with different fitted inputs; their relative accuracy does not identify a disclosure effect.

Table 2: Held-out next-year ROA prediction cells

Comparator	MAE	Out-of-sample R^2
Industry-group mean	0.05179	0.1943
Controls-only linear	0.05265	0.1237
Disclosure-augmented linear	0.05320	0.1236
Fixed-effects linear	0.04170	0.4248
Elastic net	0.05377	0.1166
Random forest	0.04177	0.3682
Gradient boosting	0.04797	0.3129

The primary association estimates appear in Table 3. For the full dictionary, the ROA coefficient is -0.0322 and its firm-cluster bootstrap interval is [-0.1616, 0.3151]. The narrow dictionary also has a negative point

estimate, but its interval crosses zero. The placebo, operating-margin, and revenue-growth intervals do the same. The evidence does not support one directional relation between disclosure and later performance.

Table 3: Association sensitivity estimates. Coefficients are pooled adjusted associations; intervals are percentile 95% firm-cluster bootstrap intervals

Outcome-measure	n	Coefficient	95% interval
ROA-full dictionary	133	-0.0322	[-0.1616, 0.3151]
ROA-narrow dictionary	133	-0.1804	[-0.4377, 0.8462]
ROA-placebo dictionary	133	-0.0102	[-0.0916, 0.0848]
Operating margin-full	97	0.1446	[-0.1296, 0.3088]
Revenue growth-full	133	-0.0520	[-0.1519, 0.3370]

The rows should be read together. The narrow dictionary occurs less often than the full dictionary and yields a wider interval, which is consistent with the limited information in a smaller phrase set. The governance placebo is centred near zero. It checks whether a general corporate vocabulary mechanically recreates the pattern; it is not a causal falsification test. Likewise, the positive operating-margin point estimate is not read as a performance benefit because the resampled interval contains negative and positive associations.

Two limitations delimit the result. Complete filings can describe implementation, intention, or risk, so the score does not verify a completed capability. The retained issuers are established U.S.-listed firms rather than a random population. Source records make those limits auditable, but they cannot remove reverse causality, omitted time-varying differences, or variation in the timing of implementation.

5. Conclusion

Public SEC records make it possible to link issuer identifiers, archived Form 10-K text, a length-normalized digital-resource disclosure score, and next-year Company Facts outcomes. In the 133-observation primary panel, the full-score ROA association and every reported firm-cluster bootstrap interval cross zero. Adding disclosure to the controls-only linear cell also leaves the held-out comparison unimproved.

The inference should remain limited. Filing language records stated digital-resource emphasis; it is neither a verified capability measure nor a causal treatment. The source mapping can support later work with section-restricted dictionaries, implementation dates, complementary organizational measures, or stronger identification designs. Until those data are available, the reported relationships are outcome-specific associations.

The Resource-Based View still explains why a filing-text association may differ across firms. Digital infrastructure and data resources become valuable when routines, governance, and complementary human capabilities accompany them. Those combinations are not observed in this panel. The replication materials instead provide a documented basis for testing them with richer operational evidence in later work on transformation outcomes.

References

- [1] Gregory Vial. 2019. Understanding Digital Transformation: A Review and a Research Agenda. The Journal of Strategic Information Systems 28, 2 (2019), 118-144. <https://doi.org/10.1016/j.jsis.2019.01.003>
- [2] Cheng Gong and Vincent Ribiere. 2021. Developing a Unified Definition of Digital Transformation. Technovation 102 (2021), 102217. <https://doi.org/10.1016/j.technovation.2020.102217>
- [3] Peter C. Verhoef, Thijs Broekhuizen, Yakov Bart, Abhi Bhattacharya, John Qi Dong, Nicolai Fabian and Michael Haenlein. 2021. Digital Transformation: A Multidisciplinary Reflection and Research Agenda. Journal of Business Research 122 (2021), 889-901. <https://doi.org/10.1016/j.jbusres.2019.09.022>
- [4] André Hanelt, René Bohnsack, David Marz and Cláudia Antunes Marante. 2021. A Systematic Review of the Literature on Digital Transformation: Insights and Implications for Strategy and Organizational Change. Journal of Management Studies 58, 5 (2021), 1159-1197. <https://doi.org/10.1111/joms.12639>

- [5] Sascha Kraus, Susanne Durst, João J. Ferreira, Pedro Veiga, Norbert Kailer and Alexandra Weinmann. 2022. Digital Transformation in Business and Management Research: An Overview of the Current Status Quo. *International Journal of Information Management* 63 (2022), 102466. <https://doi.org/10.1016/j.ijinfomgt.2021.102466>
- [6] Jay Barney. 1991. Firm Resources and Sustained Competitive Advantage. *Journal of Management* 17, 1 (1991), 99-120. <https://doi.org/10.1177/014920639101700108>
- [7] David J. Teece, Gary Pisano and Amy Shuen. 1997. Dynamic Capabilities and Strategic Management. *Strategic Management Journal* 18, 7 (1997), 509-533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7<509::AID-SMJ882>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z)
- [8] Constance E. Helfat and Margaret A. Peteraf. 2003. The Dynamic Resource-Based View: Capability Lifecycles. *Strategic Management Journal* 24, 10 (2003), 997-1010. <https://doi.org/10.1002/smj.332>
- [9] Anandhi S. Bharadwaj. 2000. A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. *MIS Quarterly* 24, 1 (2000), 169-196. <https://doi.org/10.2307/3250983>
- [10] Anna Maria Oberländer, Maximilian Röglinger and Michael Rosemann. 2021. Digital Opportunities for Incumbents: A Resource-Centric Perspective. *The Journal of Strategic Information Systems* 30, 3 (2021), 101670. <https://doi.org/10.1016/j.jsis.2021.101670>
- [11] Mobina Zareie, Najah Attig, Sadok El Ghouli and Iraj Fooladi. 2024. Firm Digital Transformation and Corporate Performance: The Moderating Effect of Organizational Capital. *Finance Research Letters* 61 (2024), 105032. <https://doi.org/10.1016/j.frl.2024.105032>
- [12] José-Miguel Gaspar, Sumingyue Wang and Liang Xu. 2024. Digitalization and the Performance of Non-Technological Firms: Evidence from the COVID-19 and Natural Disaster Shocks. *Journal of Corporate Finance* 89 (2024), 102670. <https://doi.org/10.1016/j.jcorpfin.2024.102670>
- [13] Tim Loughran and Bill McDonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance* 66, 1 (2011), 35-65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- [14] Kristina Bochkay, Roman Chychyla, Dushyant Nanda and Peter D. Wysocki. 2023. Textual Analysis in Accounting: What's Next? *Contemporary Accounting Research* 40, 2 (2023), 765-805. <https://doi.org/10.1111/1911-3846.12825>
- [15] Patrick Mikalef and Manjul Gupta. 2021. Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance. *Information & Management* 58, 3 (2021), 103434. <https://doi.org/10.1016/j.im.2021.103434>

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate

credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

