Plan Overview

A Data Management Plan created using DMPonline

Title: Digital Reporting Quality: Determinants and Consequences

Creator: Duy Tan Do

Principal Investigator: Theodosia Konstantinidi

Data Manager: Duy Tan Do

Affiliation: City St George's, University of London

Funder: Economic and Social Research Council (ESRC)

Template: ESRC Template

ORCID iD: 0000-0002-4531-7728

Project abstract:

Machine-readable financial reporting is reshaping modern capital markets, enabling investors, analysts and regulators to access and analyse corporate information more quickly and systematically. At the heart of this transformation lies something fundamental: the reliability of digital information. Flaws in XBRL filings can undermine automated financial analysis and may have consequences beyond the filings themselves – affecting investors, financial intermediaries and regulatory bodies.

This project examines digital reporting quality in the United States, drawing on more than fifteen years of filings prepared using eXtensible Business Reporting Language (XBRL). We use filing-level validation results produced by XBRL US, which identify errors and warnings in public company filings, alongside accounting, market, analyst and regulatory data. Combining these data allows us to examine XBRL filing quality in the context of firms' financial reporting practices, market activity and regulatory oversight.

We address three research questions. First, we investigate what explains variation in XBRL filing quality, considering the roles of firm and filing complexity, financial reporting quality and auditor involvement. Second, we examine how XBRL errors and warnings are reflected in market prices and analysts' expectations, focusing on stock-price reactions, subsequent returns and volatility, as well as analyst forecast dispersion, forecast errors and revisions. Third, we assess whether XBRL errors and warnings predict subsequent regulatory scrutiny and corrective action by examining whether filings with more validation errors or warnings are more likely to be followed by SEC comment letters, financial statement restatements or filing amendments.

Beyond the US analysis, the project takes an international perspective; it builds on our existing paper on European digital reporting and compares filing quality across the US and Europe. Because the two regions use different taxonomies and validation systems, the comparison focuses on dimensions that can be meaningfully compared. These include the prevalence and composition of validation problems, changes in filing quality over time, and the capital market consequences of reporting deficiencies across the two regions.

Our central contribution is to establish what XBRL validation errors and warnings reveal and what they do not. This distinction is important for understanding how much weight should be placed on validation results when assessing the reliability of machine-readable financial information. The findings will help investors and analysts interpret XBRL validation errors and warnings in the context of financial analysis and investment decisions, inform regulators' assessment of digital reporting, and provide firms, auditors and software providers with evidence on the factors associated with higher-quality filings.

ID: 213362

Start date: 01-09-2027

End date: 31-08-2029

Last modified: 21-10-2026

Grant number / URL: https://www.ukri.org/opportunity/esrc-responsive-mode-secondary-data-analysis-round-two/

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their

own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Digital Reporting Quality: Determinants and Consequences

Assessment of existing data

Provide an explanation of the existing data sources that will be used by the research project, with references

The project will use existing secondary data from established financial reporting databases and publicly accessible repositories. The primary data source is XBRL US, which provides errors and warnings relating to the quality and technical validity of corporate XBRL filings. The analysis will use validation results generated by XBRL US SEC Filer Member Rules and the Data Quality Committee, which assess aspects of financial reporting quality and data consistency. Firm-level information on internal controls, auditors, audit fees, SEC comment letters and restatements will be obtained from Audit Analytics, with filing amendment data from XBRL US. These data will be combined with financial information from Compustat and CRSP and analyst forecasts and reported actuals from I/B/E/S. For supplementary international comparisons, the project will use inline XBRL (iXBRL) filings and associated validation information from the XBRL International filings repository, covering 26 European jurisdictions, including the UK. These sources provide the existing data needed to examine the determinants and consequences of XBRL filing quality in the US and to conduct supplementary comparisons across European reporting jurisdictions. The project will use these existing datasets rather than collect new primary data. European repository coverage varies across jurisdictions, which will be taken into account when interpreting the comparative findings.

Provide an analysis of the gaps identified between the currently available and required data for the research

We have confirmed with XBRL US that filing-level validation results first became publicly available in mid-2019. This allows us to compare findings before and after public access began, and use pre-2019 observations as placebo tests for apparent market responses around validation processing dates.

Separately, some filings have no recorded processing date because they contain no validation errors or warnings. To retain these filings in the analysis, we will calculate the average time between SEC submission and validation processing for filings with recorded dates and use this interval to estimate the missing dates.

Lastly, there is incomplete coverage of European XBRL filings, which varies across jurisdictions. Some national repositories do not provide filings that can be readily accessed for systematic analysis. The availability and consistency of validation data will be taken into account in cross-country comparisons.

Information on new data

Provide information on the data that will be produced or accessed by the research project

No new primary data will be created. The project will construct new analytical datasets by integrating information from existing secondary data sources.

Quality assurance of data

Describe the procedures for quality assurance that will be carried out on the data collected at the time of data collection, data entry, digitisation and data checking.

As the project uses existing secondary data, quality assurance will focus on checking data completeness, accuracy and consistency during data extraction, merging and analysis. Data will be checked for missing values, duplicate records, inconsistent formats and implausible observations. Data from different sources will be matched using appropriate identifiers, and the accuracy of data merges will be verified. Any discrepancies or limitations will be documented, and checks will be repeated after data cleaning to ensure the final dataset is suitable for analysis. No primary data collection or manual data entry is planned.

Backup and security of data

Describe the data security and backup procedures you will adopt to ensure the data and metadata are securely

stored during the lifetime of the project.

All data will be stored securely in City St George’s approved institutional storage systems and cloud-based storage environment (Microsoft OneDrive).

Management and curation of data

Outline your plans for preparing, organising and documenting data.

Data will be organised into clearly labelled files and folders by source and research purpose. Consistent file-naming conventions will be used, and data preparation steps, variable definitions, and cleaning procedures will be documented to ensure transparency and reproducibility. A data dictionary will be maintained to describe key variables, their sources and any transformations applied. Original data will be retained separately from cleaned and analysis-ready datasets.

Difficulties in data sharing and measures to overcome these

Identify any potential obstacles to sharing your data, explain which and the possible measures you can apply to overcome these.

The data sources that we will use have different access and sharing conditions. Audit Analytics, CRSP, Compustat, I/B/E/S and XBRL US are licensed resources and their underlying data will not be redistributed. By contrast, the XBRL International repository is publicly accessible and currently permits use of its data, including the validation results used in our European research. We will therefore distinguish throughout the project between licensed source data that cannot be shared and project materials that can be made openly available.

Consent, anonymisation and strategies to enable further re-use of data

Make explicit mention of the planned procedures to handle consent for data sharing for data obtained from human participants, and/or how to anonymise data, to make sure that data can be made available and accessible for future scientific research.

No data obtained from human participants are used as the project will use secondary data on publicly listed companies.

Copyright and intellectual property ownership

State who will own the copyright and IPR of any new data that you will generate.

No new primary data will be collected. The project will use existing secondary data, for which copyright and intellectual property rights remain with the respective data providers, subject to their applicable terms and licences. Any intellectual property rights arising from newly constructed analytical datasets will be governed by City St George's intellectual property policy and the applicable data licences.

Responsibilities

Outline responsibilities for data management within research teams at all partner institutions

Both investigators, Duy Tan Do and Theodosia Konstantinidi, will share responsibility for data management, including data organisation, quality assurance, documentation and secure storage.

Preparation of data for sharing and archiving

Are the plans for preparing and documenting data for sharing and archiving with the UK Data Service appropriate?

XBRL International validation data will be prepared for deposit with the UK Data Service or another appropriate research repository, subject to the terms in force at the time of deposit.

Is there evidence that data will be well documented during research to provide highquality contextual information and/or structured metadata for secondary users?

Data sources, variable definitions, data preparation and cleaning procedures, and any transformations will be documented throughout the research. A data dictionary and clear file-naming conventions will be used to provide context and support transparency, reproducibility and future reuse of the data, subject to the relevant data providers' terms and licences.

To support reproducibility, we will share all code developed for the project, allowing researchers with access to the licensed databases to reconstruct the analytical datasets.