Institution: Lancaster University		
Unit of Assessment: 17, Business and Management Studies		
Title of case study: Improving UK financial market outcomes through the analysis of qualitative disclosures		
Period when the underpinning research was undertaken: 2012 - 2020		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s): Steven Young Paul Rayson Mahmoud El-Haj	**Role(s) (e.g. job title):** Professor (Accounting) Reader (Computing) Lecturer (Computing) & Research Fellow (Linguistics)	**Period(s) employed by submitting HEI:** 1995 – present 1990 – present 2019 – present 2016 – 2017
Period when the claimed impact occurred: 2016 - 2020		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact

Unstructured text represents 90% of the data produced since 2005 (Bank of England 2015). Quantitative financial models traditionally ignore these data because consistent, large-scale text retrieval and processing is technically difficult and very costly. Text retrieval software and analysis produced by a Lancaster University Management School (LUMS)-led interdisciplinary team harnesses text resources for financial market stakeholders. [text removed for publication]. A new scoring model for annual reports has expanded the Investor Relations (IR) Society quality review capacity from 30 to 300 reports. [text removed for publication], CFA UK, and Pensions and Lifetime Savings Association (PLSA) has been used to lobby government and influenced best practice. [text removed for publication]. The research is being used by RPMI Railpen to inform their development work and is also the basis of training for asset managers and quantitative analysts in UK and Europe.

2. Underpinning research

Ground-breaking interdisciplinary research led by LUMS accounting faculty (Young) and uniting colleagues from computer science (Rayson), corpus linguistics (El-Haj), and financial reporting (Walker at Manchester University and Athanasakou at Saint Mary's University Halifax, Canada) has analysed qualitative (text-based) corporate disclosures using natural language processing (NLP) and artificial intelligence (AI).

The volume of information disclosed by large companies has grown dramatically over the past two decades. For example, the length of an annual report published by a typical London Stock Exchange-listed company increased from 47 pages (14,954 words) in 2003 to 74 pages (33,193 words) in 2016 [R1]. These disclosures provide commentary on strategy, risk, corporate governance arrangements, and broader social welfare matters including emissions, gender pay, employee welfare, and human rights in supply chains. The majority of these disclosures are unstructured text, meaning that the content, position and format of the commentary in the document vary dramatically across firms and over time.

With tens of thousands of complex corporate disclosures published annually, demand for automated methods to help filter content and improve decision-making is high. Several structural factors inhibit progress. First, inconsistent document structure and content makes automatic retrieval and classification of textual data extremely challenging. Second, many reports are published as PDF files, which restricts the ability to extract and manipulate content because most of the metadata in the original source file is lost on conversion to PDF. Third, progress necessitates interdisciplinary teams with expertise in computing, linguistics, and financial markets. However, technical barriers to entry reinforced by discipline silos thwarts collaborations between researchers in these diverse fields. Academic progress in the field has therefore been limited, which in turn undermines regulatory efforts to improve corporate transparency.

The research began in December 2012 with a project to develop methods for text retrieval and classification [G1]. The initial focus was UK PDF annual reports because these

documents are the centrepiece of the annual corporate reporting cycle. The research team's contribution was the CFIE (Corporate Financial Information Environment) app. The app extracts annual report text and classifies content into generic categories (e.g., Chair's Statement, Governance Statement, Remuneration Report) to support time-series and cross-sectional analysis. Articles published in accounting and computer science [R2, R3] describe the method and resources to support academic research (e.g., report scores and financial corpora for >25,000 UK annual reports). Accuracy rates for retrieval exceed 90% and for classification performance are >95% for key report sections. In addition to academic colleagues, various financial market stakeholders have leveraged these resources including the CFA UK [G4], [text removed for publication] [G5], [text removed for publication] [G6]. The research team has also developed methods for analysing other document types including annual earnings announcements [R3].

In subsequent work, the CFIE app was used to measure corporate reporting quality in terms of what management say (content) and how they say it (language) [G2-G6]. The team use AI to isolate the properties of award-winning annual reports and then incorporate these features into a model that scores report quality out-of-sample. Findings reveal that language accessibility (more relevancy markers, greater connectivity, more exclusive language forms, and fewer grammatical words) combined with more information on value creation distinguish award-winning reports. A model incorporating these features predicts future winners with 75% accuracy and now informs IR Society rankings [R4]. The team also develop and test AI models for scoring optimism and causal reasoning at the sentence level. The AI models are able to replicate reading performed by a human with accuracy rates of up to 80% [R3]. An output of the research is the provision of resources to assist automatic text scoring.

In related work, Young et al study corporate policy and transparency on workforce rewards, skills, and engagement using a combination of manual and automatic methods to evaluate reporting practices. Part of this work reflects requests from industry bodies including PLSA [G3] and CFA UK [G4] to evaluate and promote best practice [R5, R6]. Key findings include evidence that executive compensation contracts for FTSE 350 firms rely on measures of performance that promote short-term decision-making, and that a high fraction of FTSE 100 firms fall short of best practice guidance in many areas of workforce reporting.

3. References to the research

[R1] Lewis, C., **Young, S.** (2019). Fad or future? Automated analysis of financial text and its implications for corporate reporting. *Accounting & Business Research,* 49(5): 587-615. https://doi.org/10.1080/00014788.2019.1611730 (25 citations Google Scholar, Altmetric score: 1)

[R2] **El-Haj, M.,** Alves, P., **Rayson, P.,** Walker, M., **Young, S.** (2020). Retrieving, classifying and analysing narrative commentary in unstructured (glossy) annual reports published as pdf files. *Accounting and Business Research,* 50(1): 6-34. https://doi.org/10.1080/00014788.2019.1609346 (17 citations Google Scholar, Altmetric score: 16)

[R3] **El-Haj, M., Rayson, P.,** Walker, M., **Young, S**., Moore, A., Athanasakou, V., Schleicher, T. (2016). Learning tone and attribution for financial text mining. In: N. Calzolari, K. Choukri, T. Declerck, M. Grobelnik, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.) *Proceedings of LREC 2016, Tenth International Conference on Language Resources and Evaluation. European Language Resources Association*, 1820-1825. [The LREC Conference is one of the leading international natural language processing (NLP) conferences]

[R4] **Young, S**., Munro, J., Alves, A. (2019). Replacing the judges? Using AI to measure annual report quality. *INFORMED* (IR Society), 104 (Autumn): 10.

[R5] **Young, S.,** Li, W. (2016). *An analysis of CEO pay arrangements and value creation for FTSE-350 companies*. CFA UK.

[R6] **Young, S.,** Rawsthorne, S., Hildyard, L. (2017*). Hidden talent: What do companies' annual reports tell us about their workers?* Pensions and Life Savings Association.

3.2 Research funding

[G1] [Understanding the Influences of Financial Reporting, Corporate Disclosures and Financial Media on the Corporate Financial Information Environment](#) (Walker PI; Young Co-I), GBP281,674, ESRC (ES/J012394/1) plus GBP53,000 from Institute of Chartered Accountants in England and Wales (Dec 12 - Jan 15).

[G2] Understanding Corporate Communications (Young Co-I) on subproject in "[ESRC Centre for Corpus Approaches to Social Science (CASS)](#)" (McEnery PI), GBP3 million ESRC (ES/K002155/1) (Mar 13 - Mar 18).

[G3] PLSA Workforce Reporting Toolkit: Impact Assessment (Young PI), GBP4,000, PLSA, (June 17-Dec 17).

[G4] An Analysis of CEO Pay Arrangements and Value Creation for *FTSE 100 & FTSE 350 Companies* (Young PI), GBP2730 Mar-Jan 2015 & GBP15,000 Dec 16- April 17, CFA UK.

[G5] [Analysing Narrative Aspects of UK Preliminary Earnings Announcements and Annual Reports: Tools and Insights for Researchers and Regulators](#) (Young PI) GBP424,390, ESRC (ES/R003904/1) plus GBP63,000 [text removed for publication] (Dec 17- Jul 20).

[G6] Detecting financial misreporting (Young PI) GBP 26,000, [text removed for publication] (Mar 18-to date).

[G7] Construction of a Governance/Long-Termism Natural Language Processing Signal (Young PI), RPMI, GBP22,800 (Oct 20 – Nov 21).

4. Details of the impact

Process from the Research to the Impact
Initial work developed the CFIE app plus other text retrieval and classification procedures. Subsequent research applied NLP and AI methods to score corporate reporting quality on a range of dimensions. The following examples of impact involve retrieval and classification tools, NLP and AI-derived quality scores, or a combination of both.

Description of the Impact
Detecting false and misleading financial statements [text removed for publication]

Fostering high-quality financial reporting and governance ([text removed for publication], IR Society, CFA UK & PLSA) (period of impact: between 2016 and 2020) Transparent reporting and progressive governance underpin the integrity of UK business and financial markets. The IR Society plays an important influencer role in the development of UK corporate reporting best practice by raising the standard of investor communications and recognising best-in-class performance (the society has over 850 members from the UK, Europe and beyond, including representatives from most of the FTSE 350). The work is helping the IR Society, which is the UK leader in investor relations, to improve the quality of UK annual reporting [R4]. The IR Society uses the AI reporting quality model outlined in [R4] to inform its annual Best Practice Awards in Annual Reporting. As the IR Society CEO explains, "*The annual report quality model developed by the research team is transforming how we approach the task of identifying superior annual reporting. In particular, using scoring methods based on artificial intelligence is enabling us to complement our traditional self-nomination procedure with a broader proactive screening approach. The result is a 900% increase in reports (from approximately 30 to 300) that we are able to consider at no incremental cost to our organisation*" [S2].

LUMS methods and data also feature in influential reports aimed at improving corporate governance practices. For example, CFA UK-commissioned work on the link between CEO pay outcomes and the transparency of executive pay reporting [G4 & R5] was featured in press coverage with an estimated reach of 9,016,541 [S3] and used by CFA UK in seminars (13 January 2015, 18 January 2017, & 4 May 2017) with institutional shareholders and stakeholder representatives to promote the importance of linking pay to measures of long-term value creation. Following the seminars, the research formed the centrepiece of CFA

UK's response (February 2017) to the UK Government's Corporate Governance Reform green paper (November 2016) [S4].

Similarly, the PLSA, representing 1,300 pension schemes with GBP1 trillion in assets under management, commissioned an evaluation of FTSE 100 companies' workforce policies and reporting [G3]. The final report [R6] highlights deficiencies in current reporting practices. The report was the centrepiece of a PLSA stakeholder engagement event (28 September 2018) and was cited widely in the business press with an estimated reach of 2,879,565 [S5]. Conclusions formed a central pillar of the PLSA response (February 2018) [S6] to the [text removed for publication] Consultation on a revised UK Corporate Governance Code (December 2017). The [text removed for publication] cites the research in its follow-up January 2020 project report [S7] and [text removed for publication] [S8].

The research team have also worked closely with the [text removed for publication] to promote an evidence-based approach to financial reporting regulation. The research [G5 & R2] has helped [text removed for publication] evaluate compliance with its reporting guidance. Analysis of compliance with [text removed for publication] guidelines on the use of alternative performance measures highlighted areas where compliance is weak, as well as specific companies with opaque reporting practices that the [text removed for publication] [S9]. Analysis using methods and data from [R2] and [R3] also formed the basis for a report on preliminary earnings announcements discussed at 2 [text removed for publication] meetings about regulation in this area [S10(a)]. [text removed for publication] [S10(b)].

Promoting text processing in the asset management industry (period of impact: between 2018 and 2020)
Fidelity International is one of the world's largest asset management firms, serving >2.4 million clients globally and managing assets exceeding GBP300 billion. Reporting complexity is forcing companies like Fidelity to seek automated solutions to support their investment strategies. Fidelity used datasets from the CFIE app [R1] as a benchmark for developing and testing its own tools for analysing global corporate disclosures. As their Senior Data Scientist states, "*We searched globally for commercial and academic solutions to help process large volumes of unstructured annual report automatically… The approach described in El-Haj et al. (2019) is pioneering. Based on our search, we concluded that no other option came remotely close to the functionality of CFIE app*" [S11].
[text removed for publication] [S11].

The quantitative analysis team at RPMI Railpen is also working with Young (G7) following his presentation on semantic tagging and topic modelling (R4) at Wolfe Research's 2nd Annual QES European Quant and Macro Investing Conference (17 Jun 2019). RPMI is responsible for investing circa GBP30 billion on behalf of the Railways Pension Scheme's 350,000 members (one of the UK's largest and longest established pension funds). The work, which is subject to a confidentiality agreement, involves modelling topics and analysing management commentary [R3, R4] to create text-based investment signals. RPMI's Investment Director (Alternative Risk Premia) confirms that, "*NLP insights and resources from LUMS research are informing development work in a material way*" [S12].

Finally, Young's work is also influencing the investment community through training. INQUIRE UK & Europe invited Young to present an overview of [R1, R2, R3] at their 2019 Spring Residential Seminar involving approx. 60 asset managers and quantitative analysts. The demand for further information on applying NLP and AI methods led INQUIRE UK to collaborate with Young to deliver a training programme (five sessions; 15 hours) on textual analysis in financial markets to investment professionals from leading financial institutions (BlackRock, UBS, BofA Securities, Fidelity, Jupiter Asset Management, FTSE and MSCI). INQUIRE UK's Chairman views the training as an important step towards a wider appreciation of NLP methods among investment professionals in the UK. He also confirms that the novel training collaboration has influenced INQUIRE's broader strategy for delivering value to its members in the future [S13].

5. Sources to corroborate the impact

[S1] Testimonial from [text removed for publication].

[S2] Testimonial, CEO, IR Society, 2021.

[S3] Media coverage on the CFA UK-commissioned research on executive pay arrangements.

[S4] Results cited in CFA UK responses to the UK Government's Corporate Governance Reform Green Paper, available here (Nov 2016) and here (Mar 2017).

[S5] Media coverage of the Hidden Talent report.

[S6] PLSA, Response to the FRC Consultation on a revised UK Corporate Governance Code. Available here, 2018.

[S7] [text removed for publication].

[S8] Email from [text removed for publication].

[S9] Report on [text removed for publication].

[S10] a) Report on [text removed for publication]. b) [text removed for publication].

[S11] Testimonial from Senior Data Scientist, Fidelity International, 2020.

[S12] Testimonial from Investment Director (Alternative Risk Premia), RPMI Railpen, 2021.

[S13] Testimonial from Chairmain, INQUIRE UK, 2021.