

AI AGAINST MODERN SLAVERY

*Digital insights into modern
slavery reporting – challenges
and opportunities*



ACKNOWLEDGEMENTS

We are grateful to all those involved in drafting and reviewing this paper. We thank Katharine Bryant, and Francisca Sasseti (Walk Free); Nyasha Weingberg, Adriana Bora, Edgar Rootalu, Karyna Bikziantieieva, Yolanda Lannquist, and Nicolas Mialhe (The Future Society); Laureen van Breen (WikiRate); and Patricia Carrier (Business and Human Rights Resource Centre). A special thanks to Tina Mash from Walk Free, who designed the publication.

We also extend our gratitude to the UK Home Office and the Australian Border Force for the fruitful discussions that contributed to the identification of our key asks to governments and businesses on machine readability and its importance for the future of modern slavery reporting.

This paper was presented at the AI for Social Good - AAAI Fall Symposium 2020 Sponsored by the Association for the Advancement of Artificial Intelligence (AAAI), that took place on November 13 – 14, 2020 using a virtual format. We wish to acknowledge and sincerely thank the feedback from the editors and reviewers for the improvement of the manuscript.

Hefei, Anhui Province of China, October 2020.

A technician monitors the parameter index at the laptop assembly workshop of LCFC (Hefei) Electronics Technology Co., Ltd.

Manufacturers of electronics are exposed to the risk of forced labour throughout their supply chains. Forced labour has been reported in the mining of minerals used to build electronics in countries such as the Democratic Republic of the Congo, while forced labour has been reported in electronics factories in Malaysia and China. Photo credit: Zhang Dagang/VCG via Getty Images

From seafood from Thailand and electronics from Malaysia and China, to textiles from India and wood from Brazil, modern slavery exists in all corners of the planet. It is a multi-billion-dollar transnational criminal business that affects us all through trade and consumer choices. In 2016, an estimated 25 million people were forced to work through threats, violence, coercion, deception, or debt bondage. Of these, 16 million were forced to work in the private sector. Given the widespread nature of the problem, governments, corporations, and the general public are increasingly expecting companies to accurately disclose the actions they are taking to tackle modern slavery. Yet, five years on, there are challenges with understanding companies' compliance under the 2015 UK Modern Slavery Act. It is unclear which companies are failing to report under the MSA, while the quality of these statements often remains poor.

Project AIMS (Artificial Intelligence against Modern Slavery) harnesses the power of artificial intelligence (AI) for tackling modern slavery by analyzing modern slavery statements to assess compliance with the UK and Australian Modern Slavery Acts, in order to prompt business action and policy responses. This paper examines the challenges and opportunities for better machine readability of modern slavery statements identified in the initial stages of this project. Machine readability is important to extract data from modern slavery statements to enable analysis using AI techniques. Although extensive technological solutions can be used to extract data from PDFs and HTMLs, establishing transparency and accessibility requirements would reduce the resources required to assess modern slavery reporting and ultimately understand what companies are doing to address modern slavery in their direct operations and supply chains — unlocking this critical 'AI for Social Good' use case.

INTRODUCTION



Amsterdam, Netherlands, November 2020. A Uyghur man gives a speech in front of an Apple store during a demonstration in the centre of Amsterdam. According to the activists, Apple is pitching against human rights legislation, by lobbying against a US Bill aimed at stopping forced labour in China. The activists performed a song against forced labour outside the store. Photo credit: Ana Fernandez/SOPA Images/LightRocket via Getty Images.

From seafood from Thailand and electronics from Malaysia and China, to textiles from India, wood from Brazil, and apparel manufacturing in the United Kingdom,¹ modern slavery exists in all corners of the planet. Modern slavery is a multi-billion-dollar transnational criminal business that affects us all through trade and consumer choices. In 2016, an estimated 25 million people were forced to work through threats, violence, coercion, deception, or debt bondage. Of these, 16 million were forced to work in the private sector (ILO and Walk Free 2017). It is estimated that approximately US\$354 billion worth of products at-risk of being produced by forced labor are imported by G20 countries annually (Walk Free 2018). Given the widespread nature of the problem, governments, corporations, and the general public are increasingly expecting companies to accurately disclose the actions they are taking to tackle modern slavery.² A valuable source of information is corporate reporting resulting from supply chain transparency requirements in domestic legislation.³

The Future Society,⁴ in partnership with Walk Free,⁵ launched Project AIMS (Artificial Intelligence against Modern Slavery) in May 2020. Project AIMS seeks to, firstly, understand how we can harness the power of artificial intelligence (AI) to increase the efficiency of assessing compliance with the UK and Australian Modern Slavery Acts. Secondly, the project will allow us to understand how we can harness the power of AI for policymaking by providing actionable insights for governments, businesses, and civil society organizations. The overarching project will attempt to identify and share best practices in modern slavery reporting, and identify specific sectors where reporting is falling short. It will make recommendations for companies on how to improve compliance with the UK and Australian Modern Slavery Acts and for governments considering developing similar legislation on how to maximize its impact.

Project AIMS builds upon the work of Walk Free, WikiRate,⁶ and Business & Human Rights Resource Centre (BHRRC)⁷ to assess the statements produced under the UK Modern Slavery Act. It draws from the BHRRC Modern Slavery Registry to develop an AI algorithm to ‘read’ and

assess the statements produced by companies under supply chain transparency legislation.⁸ This algorithm will use 18 metrics designed by Walk Free in line with the UK Home Office guidance (UK Government 2017) to assess statements, and will be integrated with the WikiRate platform to enable ongoing human verification of the automated data collection.

There are four phases to Project AIMS. The first phase of Project AIMS is focused on accessing, gathering and structuring the data from existing company statements, building the largest publicly available text corpus of modern slavery statements.⁹ In phase two of the project, we will design an automated labeling function through weak supervision tools to increase the amount of available labeled data. Once sufficient data are correctly labeled, the third phase of Project AIMS begins: using supervised machine learning methods to create a document classifier, which can assess modern slavery statements against the 18 metrics. Lastly, in the fourth phase, the results will be published, and the tool will be made publicly available through an open-source API.

This publication addresses the challenges and opportunities identified during the first phase, namely the process of accessing, gathering and structuring the data from the company statements. Data collection and structuring is a key cornerstone in building any successful AI project and thus this publication puts forward a set of lessons learned and recommendations on good practices to facilitate the application of AI for Social Good. This paper adopts the perspective that although extensive technological solutions can be used to extract data from modern slavery statements in PDF and HTML formats, establishing transparency and accessibility requirements would reduce the resources required to do so. It focuses on changes that would enable resource-constrained technical experts to extract data in a more efficient manner than what is technically feasible today.

BACKGROUND

Following California's 2010 Transparency in Supply Chains Act, the UK developed the first national legal framework for transparency in supply chains: the 2015 Modern Slavery Act. It includes a provision that requires companies supplying goods or services in the UK with an annual turnover of £36 million or more to publish an annual modern slavery statement indicating the steps they are taking to identify and address modern slavery risks.

Yet, five years on, there are several challenges in understanding business compliance with the UK Modern Slavery Act. It is difficult to establish which companies are failing to report, while the variable quality of the statements released makes it difficult to understand the actions companies are taking to address modern slavery. With an estimated 12,000–17,000 UK-based companies having to publish statements per annum, few studies have attempted to assess these reports due to the laborious nature of manually analyzing each statement (Walk Free et al. 2019). For example, even the most comprehensive study to date, conducted by Walk Free and WikiRate (2018), sampled just over 900 reports and took almost two years to complete. As companies continue to report under the UK legislation and start to report under the 2018 Australian Modern Slavery Act, failure to address these obstacles to efficiently and consistently assess modern slavery statements will undermine the potential of this legislation to improve transparency and accountability in business operations and supply chains.

To date, access to modern slavery statements has been through company websites¹⁰ or the compilation efforts by the BHRRC's Modern Slavery Registry,¹¹ TISC,¹² and WikiRate,¹³ who have collected, collated, and analyzed these data. Much of this information has been collated manually, with teams of researchers searching for, and systematically reviewing, available statements. Given this is a costly exercise that requires a lot of man-hours, a more centralized and automated approach is desirable. Promising steps in this regard are the

development of the UK Home Office registry, and the recent launch of Australia's registry, which will centralize the housing of these statements.¹⁴ Technological innovations will also reduce the time taken to extract relevant information from these statements. This enables insights into company disclosure of actions to remove modern slavery from their operations and supply chains, and also facilitates the automation of elements of the assessment of these statements.

This is a technically challenging task. However, the challenges in dealing with large complex structured and unstructured data sets are not new, and neither is the quest to harness AI technologies to tackle them (Pferd 2010).¹⁵ Big data has been widely adopted as a solution to tackle the mammoth task of exploring and extracting meaningful insights from large structured and unstructured datasets (Adnan and Akbar 2019; Rai 2017; Yang et al. 2019). There are also evident gaps in data governance and the need for a more holistic view to guide both practitioners and researchers in this field (Abraham, Schneider, and vom Brocke 2019). Prominent areas of this application include the medical and healthcare sectors, with several studies showing how the use of AI to structure data sets and extract information can contribute to the prevention of infectious diseases and identification of key areas of interventions, but not without its own challenges (Cohen et al. 2017; McCue and McCoy 2017).

This project aims to use AI to support the achievement of the Sustainable Development Goals (SDG), including SDG 8 aimed at:

Promot[ing] sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all¹⁶

AIMS seeks to demonstrate that, beyond optimizing business performance, the use of AI-based solutions can be leveraged to strengthen the rule of law, specifically supply chain transparency legislations that address modern slavery risk, remediation, and prevention.

RECOMMENDATIONS

Based on the creation of the dataset under the first phase of Project AIMS, we set out the following recommendations for policymakers and companies to improve access to modern slavery reporting using technology.

For Policymakers

Recommendation 1:

Governments with modern slavery reporting requirements should publish an up-to-date, comprehensive list of all companies and their subsidiaries subject to reporting.

Recommendation 2:

Governments should keep a single registry where companies must submit their statements. These statements must have consistent formatting to ensure easy retrieval.

- 2a** Ensure that all statements are required to disclose the reporting period, are timestamped, and include relevant metadata,¹⁷ such as the address of company headquarters, that assist interoperability with other data sources.
- 2b** In addition to housing on the company homepage, require companies to submit their statements to the registry.
- 2c** House historical statements in this same registry.

Recommendation 3:

Governments should legislate that companies should publish statements in machine readable formats¹⁸ to improve comparability and support transparency.

For Companies

Recommendation 1:

Companies subject to modern slavery reporting requirements should endeavor to assist governments with keeping an up-to-date, comprehensive list of all companies and their subsidiaries subject to reporting.

Recommendation 2:

Companies should place their modern slavery statements on their homepage, with a URL that includes the reporting year.

- 2a** Ensure that all statements disclose the reporting period, are timestamped, and include relevant metadata that assists interoperability with other data sources, such as the address of company headquarters.
- 2b** In addition to housing on their homepage, submit their statements to the registry, with consistent formatting.
- 2c** Provide records of historical statements on their website and in the registry.

Recommendation 3:

Companies should publish in a machine-readable format, with infographics and images comprehensively explained in text that fully summarizes and references all information contained within.

CHALLENGES

Accessing high-quality, structured, machine readable data from companies' Modern Slavery Act statements is a significant challenge (Rodriguez 2018). This is particularly true when assessing a large number of these statements to identify sector-specific characteristics, or to illustrate change over time. However, detailed reports are not mandatory, nor are these statements standardized or saved in consistent formats. The content included in statements is left to the discretion of companies, resulting in vast differences in substance and quality. This presents several problems for the use and development of AI to facilitate the extraction and analysis of relevant information at scale.

Access Challenges

Access is a significant issue facing anyone who wants to extract data. Data are accessible for AI if it can be identified, extracted, processed, and parsed easily by a computer.

Identification of Relevant Reports

To extract data from relevant reports requires the identification of companies that are subject to mandatory reporting requirements. In the UK, this process currently requires visiting company websites and drawing from existing datasets (such as WikiRate or the Modern Slavery Registry), as there is no centralized government registry yet. This raises a number of issues, including:

Finding companies that are subject to a reporting duty

To date, there is no publicly available list of companies which are in scope of the UK Modern Slavery Act. This makes it incredibly difficult, if not impossible, to identify which companies are in scope of the Act, and pinpoint which should have reported, but have not yet done so. The AIMS project compares metadata variables (e.g. 'name' or 'URL') across two separate data sets of modern slavery statements from the WikiRate platform and the Modern Slavery Registry. While these databases have collected statements published by companies, they are inevitably incomplete due to the inherent difficulties of collecting all statements in scope. The goal of this comparison was to conduct a gap analysis and assist with the identification of additional statements. This analysis has revealed the difficulty of analyzing companies with complex structures, often with multiple subsidiaries, inconsistent industry classifications, and companies that span multiple industries, which creates challenges for generating a comprehensive streamlined dataset of companies.

Scraping reports from company websites

Based on the analysis by Project AIMS, from the approximately 17,000 unique statement URLs stored in the Modern Slavery Registry, just 12,005 could be accessed. In approximately 4,913 cases, errors blocked the scraping process, of which 328 errors were related to HTML stored formats. The remaining 4,585 errors affected those stored in PDF format. More precisely, of the approximately 10,700 URLs containing the statements in PDF format, only 6,212 statements could be accessed.

These errors were caused by a number of issues that make website scraping complicated, such as shifting webpage structures, redirects and CAPTCHAS,¹⁹ unclear navigation, and unstructured HTML.²⁰ These issues include:

- Statement missing from homepage. Not all companies follow government requirements to publish modern slavery statements in a prominent place on their homepage (Home Office 2019).
- Shifting webpage structures. Website redesign means that sections can become more complicated to access.
- Connection issues caused by URL structures. Complicated URL structures, including multiple query strings and hashes create significant connection issues.
- Connection issues caused by server connection errors. To scrape the statements, the computer sends a request for processing to the web server that hosts the statement, and the server then sends a response to the computer running the code. If the server is not connected, it is not possible to scrape data.
- Unclear links. Some links direct to a page which hosts a number of links to modern slavery statements instead of the most recent statement itself. This leads to the text being extracted from that website instead of the text from the actual reports.²¹ While it is helpful for companies to have a webpage that links to all of the previous modern slavery statements in one place, it is essential for automation purposes to have the most up-to-date statement in an easily traceable location.
- Blocked scraping. Some websites block instances when text is scraped multiple times. While this may be useful in some contexts, when applied to a page that houses modern slavery statements it hampers transparency.

Format of Reporting

The format of modern slavery reporting can also add hurdles for the extraction of data.

Lack of Digital Formats

A new EU regulation requires all financial statements to be published in a digital format (Laermann 2018). The UK Modern Slavery Act, on the other hand, does not mandate companies to publish modern slavery statements in a single electronic format. This means that the statements are inconsistent and different approaches need to be taken to extract data from each format.

Extraction of Data from PDFs

Data published in PDF format, which is the form that modern slavery statements often take, is not easily machine readable (Pollock 2016), which makes it more difficult to identify, read, extract and analyze information automatically. Specific issues include:

- **Scanned PDFs.** Scanned PDFs are often not machine readable as they are captured as a solid image. Optical Character Recognition (OCR) can help conversion into machine-encoded text that can then be analyzed, but this is more arduous than using a PDF saved directly from a computer.
- **Formatting.** The use of formatting, including borders, multiple columns, inconsistent column widths, pop-out boxes, headers, and footers, adds to the complexity of data extraction.
- **Data embedded in images and graphics.** A further challenge is extracting data that are embedded in images and graphics., which present challenges to the structuring and automatic processing of data. Without developing specialized methods for extracting data from complex tables, figures, and graphs, these can scramble the information contained within. Based on the analysis so far, out of the 5 903 extracted statements in HTML format, 96 statements have data embedded in meaningful images, while out of 6,092 statements in PDF format, 237 contained meaningful images.²²

Figure 1. Example risk assessment heatmap.
Source: authors

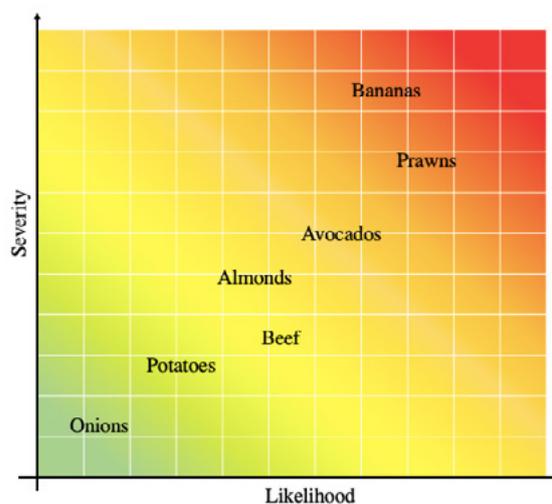


Figure 1 demonstrates some of these challenges. If the information contained within the heatmap was captured within a paragraph text, the tool could easily extract the information “Bananas and prawns are the products most at risk.” It is possible to use computer vision to read the text, but without additional code to read the colors as risk indicators we would not be able to rank the information contained within the figure.

Lepe, Spain, May 2020. A self-made home constructed with a surfboard, plastic sheet, and blankets in a shanty town. Migrants living in this shanty town in the Huelva province of Spain are labourers picking berries. Berry production has dramatically increased in the last three decades. Mobility restrictions as a result of the Covid-19 pandemic made access to work more difficult for those who do not own a car or a bicycle. Public buses offered limited places and people must sit separated. In some cases, owners of plantations organised to pick up workers one-by-one. Marked by labour abuses that taint entire supply chains in the past, farming and agriculture continue to be high risk sectors for migrant workers, now more aggravated by the pandemic. Photo credit: Niccolo Guasti/Getty Images.





- **Sub-formats of PDF.** Each specific format of PDF requires a separate OCR solution for extracting the data.

Figure 2. Example of diagram. Images and diagrams can make text extraction difficult. Source: authors.



Figure 2 also provides important information on a company's modern slavery strategy, but the use of a diagram creates additional difficulties in the process of reading, extracting and structuring data. These diagrams are, however, essential for a number of stakeholder groups to help them understand company modern slavery strategies, which is why we do not recommend removing them, but rather supplementing them with a text-based description.

Structure of Reports

- **Section Titles.** Without clearly demarcated section headings that mirror the government's sections for reporting, it can be difficult to find relevant information for specific metrics.
- **Alignment with reporting standards.** Use of conventional terminology enables easier extraction, and further analysis would be enhanced if this aligned with globally specific reporting standards or frameworks (e.g. SASB, EU frameworks).
- **Tagging.** Labels by companies to assist machine readability would be very helpful; this is particularly important in areas where we see inconsistent typologies used by companies to describe similar phenomena. This could follow suggested or mandated criteria from governments.²³

OPPORTUNITIES

Using Structured, Machine Readable Formats across Corporate Reports

Machine-readable formats would make information contained in modern slavery statements more easily accessible, which would facilitate data retrieval and allow for better comparisons between companies within and across sectors and countries, and show change over time. It would also allow for adjustments that enable access for people with disabilities.

The methodology developed to extract relevant information through Project AIMS could also be applied to other reporting frameworks to develop a comprehensive picture of corporate disclosure and activity. In particular, there is an opportunity to apply this extraction technology to Environmental, Social and Governance (ESG) reporting frameworks. There are currently more than 230 sustainability reporting frameworks, which ultimately impairs rather than aids the extraction, comparability, and analysis of the wealth of information contained within these reports (XBRL 2018).

There is also an opportunity to extend financial reporting requirements to modern slavery reporting and ESG data to assist efforts to source and efficiently integrate data into cross-asset investment decisions and implementation. Companies' annual financial reports are made machine-readable under new European Securities and Markets Authority rules. Doing the same with modern slavery statements and ESG data would improve comparability, support

transparency and contribute to increased investor protection (Rust 2017). At a minimum, structured reporting, even if not in XHTML format, would align with the EU's 2013 Transparency Directive Recital 26 which states that:

a harmonised electronic reporting format would be very beneficial for issuers, investors and competent authorities, since it would make reporting easier and facilitate accessibility, analysis and comparability of annual financial reports (European Securities and Markets Authority n.d.).

Given the challenges and opportunities for machine readability of modern slavery reporting explored in this paper, we believe that establishing transparency and accessibility requirements would reduce the resources required to assess modern slavery reporting, increase understanding of the actions companies are taking to address modern slavery, and ultimately hold companies accountable for the exploitation that occurs in their direct operations and in their supply chains.

Brazil, April 2018. The European Commission announces a meat embargo on meat and meat products manufactured by 20 Brazilian plants due to public health reasons. In recent years, Reporter Brasil, an independent research group, has uncovered dozens of cases of extreme worker abuses among the farms of Brazil's top beef producers. Photo credit: Cris Faga/NurPhoto via Getty Images.



Serwinda Zuccarelli

Servindo Qualidade Há Mais de 100 Anos!!!



Novo Boi

COSTEL
GAÚCHA

14,9



TOLEDO



Peso Kg

Preço/kg R\$

Preço Total R\$

0
15,90

ENDNOTES

- 1 A recent undercover investigation brought to light the slavery-like exploitative conditions in a factory in Leicester producing clothes for fashion giant Boohoo, where workers received significantly less than minimum wage and worked without protective equipment (Duncan 2020; Matety 2020).
- 2 Business and Human Rights Resource Centre, modernslaveryregistry.org/. Accessed October 1, 2020.
- 3 See UK Modern Slavery Act 2015, Australian Modern Slavery Act 2018, California Supply Chain Transparency Act 2010, French Duty of Vigilance Law 2017.
- 4 The Future Society is an independent 501(c)(3) nonprofit think and do tank working on advancing the responsible adoption of AI and other emerging technologies for the benefit of humanity. The Future Society, thefuturesociety.org/. Accessed October 1, 2020.
- 5 Tackling one of the world's largest and most complex human rights issues requires serious strategic thinking. Walk Free approaches this challenge by integrating world class research with direct engagement with some of the world's most influential government, business, and religious leaders. We invest our time and resources in a collaborative manner to drive behavior and legislative change to impact the lives of the estimated 40 million people living in modern slavery today. Walk Free, walkfree.org/. Accessed October 1, 2020.
- 6 WikiRate is a nonprofit that hosts an open data platform which allows anyone to systematically gather, analyze and report publicly available information on corporate Environmental, Social and Governance (ESG) practices. By bringing this information together in one place, and making it accessible, comparable and free for all, the organization provides society with the tools and evidence it needs to spur companies to respond to the world's social and environmental challenges. To date, WikiRate.org is the largest open source registry of ESG data in the world, with currently almost 900,000 data points for over 55 000 companies. WikiRate, wikirate.org/. Accessed October 1, 2020.
- 7 The BHRRC is an international, non-profit organization that works to advance human rights in business and eradicate abuse. Its website tracks the activities of more than 10 000 companies around the world. Business and Human Rights Resource Centre, businesshumanrights.org/en/. Accessed October 1, 2020.
- 8 Business and Human Rights Resource Centre, modernslaveryregistry.org. Accessed October 1, 2020.
- 9 This corpus combines the small amount of "labeled statements" (the modern slavery statements manually benchmarked against the 18 metrics by volunteers from WikiRate and Walk Free) with the large amount of "unlabeled statements" (the statements from the Modern Slavery Registry that have not yet been benchmarked).
- 10 Examples of best practice under the UK Modern Slavery Act (Business & Human Rights Resource Centre 2018); Examples of publications under the Duty of Vigilance Law: (Carrefour 2018).
- 11 Business and Human Rights Resource Centre, modernslaveryregistry.org. Accessed October 1, 2020.
- 12 TISC Report, tiscreport.org/. Accessed October 1, 2020.
- 13 WikiRate, wikirate.org/. Accessed October 1, 2020
- 14 Australian Modern Slavery Registry, modernslaveryregister.gov.au/. Accessed October 1, 2020.
- 15 It is important to note that many important data management and analytics tasks cannot be full done by automated processes, therefore crowdsourcing is used to harness human cognitive abilities to process some computer tasks, such as sentiment analysis and image recognition. This area of work has been extensively studied in recent years as Li et al. (2017) suggest.
- 16 United Nations, sdgs.un.org/goals/goal8. Accessed October 1, 2020.
- 17 Based on our research to date, a good metadata for this purpose would be the address of company headquarters.
- 18 A machine-readable format is a type of structured format that can be read and processed by a computer. Examples suitable for modern slavery statements include Extensible Markup Language (XML). A machine-readable format does not include PDF, although different PDF formats facilitate readability.
- 19 A CAPTCHA is a type of test to determine whether or not a particular user is human
- 20 Unstructured HTML is where the HTML has not been tagged in a consistent pattern that allows for analysis. Sometimes, unstructured HTML is simply a consequence of bad programming (Kansal 2019).
- 21 Victoria and Albert Museum, vam.ac.uk/info/reportsstrategic-plans-and-policies. Accessed October 1, 2020.
- 22 A meaningful image is any kind of infographics containing information that is important for the benchmarking of a metric (e.g. report in image format, supply chain embedded in the image, description of the company etc. This does not include images containing signatures).
- 23 For example, guidance by the Australian Government Department of Home Affairs (2018) identifies seven mandatory criteria for reporting which can be clearly tagged in headings across the statements.

REFERENCES

- Abraham, Rene, Johannes Schneider, and Jan vom Brocke. 2019. "Data Governance: A Conceptual Framework, Structured Review, and Research Agenda." *International Journal of Information Management* 49:424-38. doi: 10.1016/j.ijinfomgt.2019.07.008.
- Adnan, Kiran, and Rehan Akbar. 2019. "An Analytical Study of Information Extraction from Unstructured and Multidimensional Big Data." *Journal of Big Data* 6(1):91. doi: 10.1186/s40537-019-0254-8.
- Australian Government Department of Home Affairs. 2018. *Commonwealth Modern Slavery Act 2018 - Guidance for Reporting Entities*.
- Business & Human Rights Resource Centre. 2018. "Unilever Discloses Entire Palm Oil Supply Chain; Explains Decision as Vital in Addressing Deforestation and Human Rights Abuses." *Business & Human Rights Resource Centre*, February 20.
- Carrefour. 2018. "2.6.2 Le Plan de Vigilance Du Groupe Carrefour."
- Cohen, Trevor, Kirk Roberts, Anupama E. Gururaj, Xiaoling Chen, Saeid Pournajati, George Alter, William R. Hersh, Dina Demner-Fushman, Lucila Ohno-Machado, and Hua Xu. 2017. "A Publicly Available Benchmark for Biomedical Dataset Retrieval: The Reference Standard for the 2016 BioCADDIE Dataset Retrieval Challenge." *Database* 2017. doi: 10.1093/database/bax061.
- Duncan, Conrad. 2020. "Boohoo 'Facing Modern Slavery Investigation' after Report Finds Leicester Workers Paid as Little as £3.50 an Hour." *The Independent*, July 5.
- European Securities and Markets and Authority. n.d. "European Single Electronic Format."
- Home Office. 2019. "Publish an Annual Modern Slavery Statement." *GOV.UK*. Retrieved October 19, 2020 (<https://www.gov.uk/guidance/publish-an-annual-modern-slavery-statement>).
- ILO, and Walk Free. 2017. *Global Estimates of Modern Slavery: Forced Labour and Forced Marriage*. Geneva, Switzerland: International Labour Office (ILO).
- Kansal, Satwik. 2019. "Advanced Python Web Scraping: Best Practices & Workarounds." *Codementor Blog*. Retrieved October 19, 2020 (<https://www.codementor.io/blog/python-web-scraping-63l2v9sf2q>).
- Laermann, Michael. 2018. "Machine-Readable Disclosure and ESEF 2020 - A New Era of Corporate Digital Reporting?" *Sustainable Brands*. Retrieved October 19, 2020 (<https://sustainablebrands.com/read/marketing-and-comms/machine-readable-disclosure-and-esef-2020-a-new-era-of-corporate-digital-reporting>).
- Li, Guoliang, Yudian Zheng, Ju Fan, Jiannan Wang, and Reynold Cheng. 2017. "Crowdsourced Data Management: Overview and Challenges." Pp. 1711-1716 in *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD '17*. New York, NY, USA: Association for Computing Machinery.
- Matety, Vidhathri. 2020. "Boohoo's Sweatshop Suppliers: 'They Only Exploit Us. They Make Huge Profits and Pay Us Peanuts.'" *The Sunday Times*, July 5.
- McCue, Molly E., and Annette M. McCoy. 2017. "The Scope of Big Data in One Medicine: Unprecedented Opportunities and Challenges." *Frontiers in Veterinary Science* 4. doi: 10.3389/fvets.2017.00194.
- Pferd, Jeffrey W. 2010. "The Challenges of Integrating Structured and Unstructured Data." *Technical Paper*. Landmark.
- Pollock, Rufus. 2016. "Tools for Extracting Data and Text from PDFs - A Review." *Open Knowledge Labs*. Retrieved October 19, 2020 (<http://okfnlabs.org>).
- Rai, Rajeshwari. 2017. "Intricacies of Unstructured Data." *ICST Transactions on Scalable Information Systems* 4:153151. doi: 10.4108/eai.25-9-2017.153151.
- Rodriguez, Arturo. 2018. "It Takes Two to Make Data Coverage Go Right." *SASB*. Retrieved October 19, 2020 (<https://www.sasb.org/blog/it-takes-two-to-make-data-coverage-go-right/>).
- Rust, Susanna. 2017. "Company Financial Reports to Be Made Machine-Readable." *IPE*. Retrieved October 19, 2020 (<https://www.ipe.com/company-financial-reports-to-be-made-machine-readable/10022424.article>).
- UK Government. 2017. *Transparency in Supply Chains. A Practical Guide*.
- Walk Free. 2018. *Global Slavery Index*.
- Walk Free, WikiRate, Business & Human Rights Resource Centre, and Australian National University. 2019. *Beyond Compliance in the Hotel Sector: A Review of UK Modern Slavery Act Statements*.
- WikiRate, and Walk Free Foundation. 2018. *Beyond Compliance: The Modern Slavery Act Project*.
- XBRL. 2018. "CRD Launch Project to Standardise ESG." *XBRL*. Retrieved October 19, 2020 (<https://www.xbrl.org/news/crd-launch-project-to-standardise-esg/>).
- Yang, Longzhi, Jie Li, Noe Elisa, Tom Prickett, and Fei Chao. 2019. "Towards Big Data Governance in Cybersecurity." *Data-Enabled Discovery and Applications* 3(1):10. doi: 10.1007/s41688-019-0034-9.

WALKFREE.ORG