Operationalizing the SDRCS Framework with Social Media Data During COVID-19

Liana Bravo-Balsa[1], Vicky Brotherton[1], Alison Gardner[1], Erika Jiménez[1], Todd Landman[1], Benjamin Lucas[1], Elizabeth Such[2], and Nicola Wright[1]

[1] *Rights Lab, University of Nottingham, United Kingdom*
[2] *School of Health and Related Research, University of Sheffield*

November 2021

Abstract

In this working paper, we detail the first steps toward operationalization of the SDRCS Framework ("Social Determinants of Resilience to Contemporary Slavery") from Gardner et al. (2020), via text mining and information retrieval techniques, using a dataset of Tweets made by major anti-slavery organisations during the COVID-19 crisis. This provides insights into the strategic communications of prominent anti-slavery organizations during COVID-19 along the lines of their strategic agendas and tactical priorities in terms of: (a) addressing prominent COVID-19 risks affecting vulnerable populations, victims, and survivors of modern slavery, and, (b) representing and serving these affected populations during this time.

1 Introduction

Gardner et al's (2020) "Social Determinants of Resilience to Contemporary Slavery" (SDRCS) Framework represents a pivotal effort in the development of community-based approaches to anti-slavery resilience building. The authors proposed a set of 32 sub-dimensions, (a) organized in terms of their position along a prevention, discovery, recovery and resilience continuum, and (b) sub-classified by personal, culture and locality, legal and regulatory and structural designations. In doing so, the authors provide a framework for better governance in societies at all levels - geared toward the elimination of modern slavery and fulfilment of United Nations SDG 8.7.

Especially interesting for new empirical research is that this framework, with its proposed mix of intersecting discrete and dynamic constituent dimensions, offers avenues for research employing text mining and information retrieval to quickly define queries to be performed on datasets from sources such as social media, the news media, and reports by organizations engaged in the anti-slavery space. From these queries - text can be quantified into variables for trend analysis, or can be utilized as a sub-dataset for more focused human interpretation.

In this working paper, we explore the initial phases of operationalization of the SDRCS Framework using a dataset of Tweets made by the most Twitter-active anti-slavery organizations during the COVID-19 crisis. The work involved in making this dataset queryable can be thought of, in simple terms, as falling somewhere between the development of qualitative coding schemas, and the development of an information retrieval (e.g. search engine) apparatus. All supporting visualization is performed using the R Package ggplot2 (Wickham et al., 2020).

2 Dataset and Workflow

Figure 1 provides an example query workflow. Here, we take a dataset of 29,977 Tweets and 16,199 Retweets retrieved using the R Package rtweet (Kearney et al., 2020) - showing how these could be filtered based on the five illustrative criteria shown in blue. Specifically, the user has a choice between content mentioning COVID-19 or not, content mentioning victims or survivors of modern slavery or populations vulnerable to modern slavery or not, content mentioning modern slavery prevention (versus discovery, recovery and resilience stages - following Gardner et al., 2020) or not, a sub-query: content also explicitly mentioning risks or not, and finally, the option to filter according to specific expert-defined themes or topics within the modern slavery sphere.

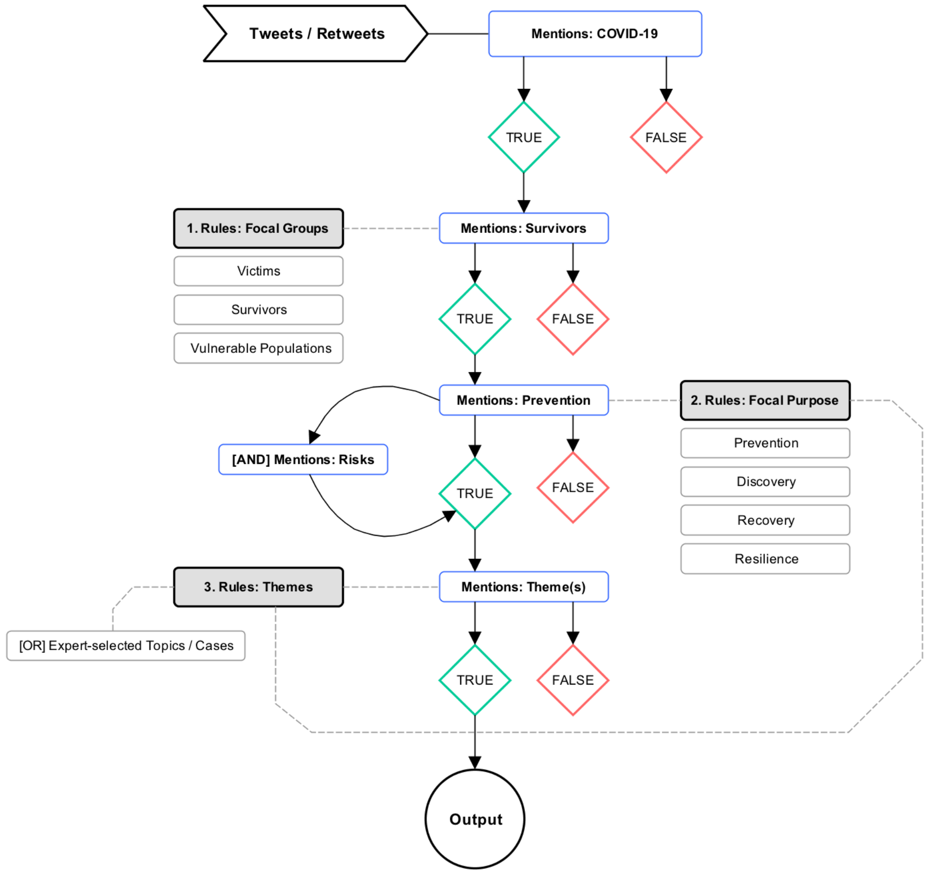

Figure 1: Example Query Workflow

The user can select their own search hierarchy, meaning that fewer or more levels may be employed (e.g. specifically searching for content related to survivors of modern slavery). This proposed workflow also accommodates non-predetermined natural language queries, but advises queries based on a priori designations. This process is discussed next.

Dimension	**Percentage**	**Dimension**	**Percentage**
Advocacy	0.23%	Victims	5.79%
Awareness	2.24%	Immigration	5.77%
Compensation	0.28%	Survivors	4.31%
Consumer	0.39%	Risks	4.20%
Corruption	0.14%	Government	3.63%
Criminalisation	0.08%	Human Rights	3.11%
Data Sharing	0.04%	Vulnerable	3.03%
Domestic	1.55%	Family	2.96%
Education	1.11%	Justice	2.93%
Employment	0.83%	Law	2.39%
Enforcement	0.74%	Awareness	2.24%
Family	2.96%	Supply Chains	2.00%
Gender	1.33%	Domestic	1.55%
Government	3.63%	Gender	1.33%
Healthcare	0.81%	Education	1.11%
Housing	0.32%	Sustainability	0.95%
Human Rights	3.11%	Employment	0.83%
Immigration	5.77%	Healthcare	0.81%
Justice	2.93%	Spot the Signs	0.76%
Labour Rights	0.34%	Enforcement	0.74%
Law	2.39%	Legislation	0.71%
Legislation	0.71%	Poverty	0.49%
Minimum Wage	0.22%	Consumer	0.39%
Multi-agency	0.11%	Labour Rights	0.34%
Poverty	0.49%	Housing	0.32%
Procurement	0.06%	Compensation	0.28%
Risks	4.20%	Advocacy	0.23%
Slavery-free	0.07%	Universal Credit	0.23%
Spot the Signs	0.76%	Minimum Wage	0.22%
Supply Chains	2.00%	Corruption	0.14%
Support Services	0.04%	Multi-agency	0.11%
Survivors	4.31%	Criminalisation	0.08%
Sustainability	0.95%	Slavery-free	0.07%
Universal Credit	0.23%	Procurement	0.06%
Victims	5.79%	Support Services	0.04%
Vulnerable	3.03%	Data Sharing	0.04%
Total:			51.14%

Table 1: Percentage of Tweets Containing Selected Dimensions

3 Query Customization

As introduced, our workflow (Figure 1) allows for customization of queries along the lines of natural language. This gives users the ability to customize query outputs based on any choice of input search terms. This includes queries directly linked to modern slavery, as well as ancillary topics related to other major societal shocks, such as climate change and international elections. To make this process more systematic, and working towards an operationalization approach, here, we focus on a customized rule dictionary, primarily capturing the 32 discrete dimensions within Gardner et al's (2020) framework. We extend this list to 36 dimensions (shown in Table 1), based on a combination of splitting dimensions in the original framework, and adding additional dimensions relevant to the specific context of modern slavery (and its presentation) during COVID-19. "Victims", "Immigration" and "Survivors" are the most prominent dimensions.

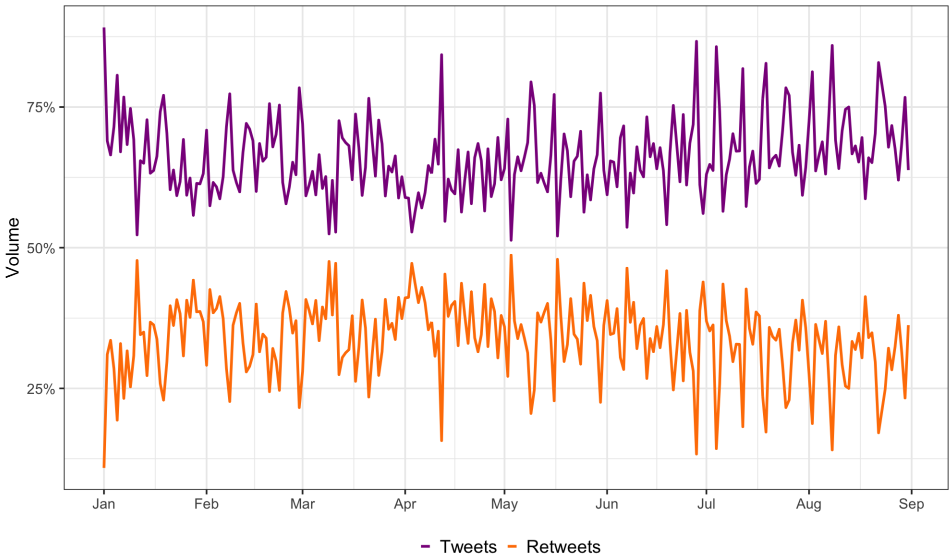

Figure 2: A Comparison of Tweet and Retweet Volume

As also introduced, the purpose of this work is to show how text can be quantified into variables for trend analysis, or can be utilized as a sub-dataset for more focused human interpretation. Here we focus on illustrating the former. The first step in this process is examining other trends present in a given target dataset deriving from usage volume and engagement metrics. As a backdrop example, Figure 2 shows Tweet and Retweet volume in our dataset between the 1st of January 2020 and the 31st of August 2020. Importantly, trend analysis also necessarily involves the introduction of external datastreams. To this end, we demonstrate a re-scaled comparison of such metrics with the John Hopkins University (2020) global COVID-19 death statistics. Figure 3 shows a comparison with any mentions of COVID-19 or Coronavirus, Figure 4 shows a usage comparison of #covid-19 and #modernslavery.

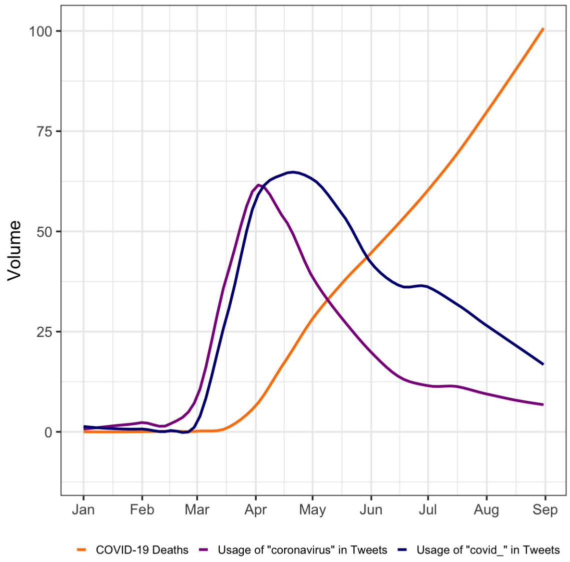

Figure 3: Global COVID-19 Deaths and Dataset Terminology Trends

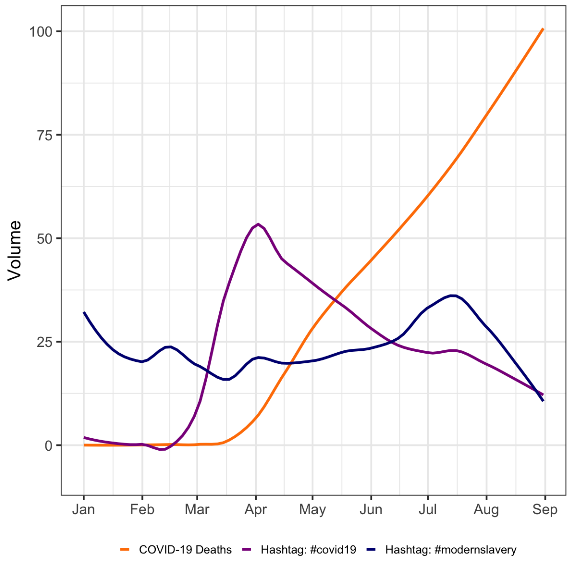

Figure 4: Global COVID-19 Deaths and Dataset Hashtag Trends

The second step returns to Figure 1 - where a query can be based on a hierarchy of search queries, including dimensions falling on the prevention, discovery, recovery and resilience continuum presented in Gardner et al. (2020). That is - once an understanding of the distributions of key metrics in the dataset is established, searches for specific terms can be conducted.

Using our 36 dimensions for illustration, Figure 5 shows trends based on a custom score, using independent (within-dimension) 0-100 normalization. Final calculations are then performed used localized regression (LOESS). These dimensions comprise the results of customized underlying rule dictionaries, and have been assigned simple labels (in UK English) for clarity. A summary of this dictionary development process follows. Figure 6 shows "Awareness", "Risks", "Spot the Signs" and "Victims" as a percentage of total Tweet activity for comparison.

At the most basic level, our system needs to handle singular / plural terms, hyphenation, misspellings and differences in UK / US English. This is achieved by basic customization of our rule dictionary. These are mostly straightforward (e.g. word stems can easily be defined to identify 'family' and 'families' but to also exclude 'familiar'). Some queries require multiple related stems (e.g. for 'immigration' and 'migration' versus 'migrant' and 'migrants'). Other selections are based on expert input and dataset-specific descriptives (e.g. in our dataset, decriminalization usually refers to 'decriminalization' in benefit of survivors of modern slavery, and 'domestic' almost exclusively refers to themes such as 'domestic violence' and 'domestic servitude'). Similarly, a number of the dimensions involve specific individual semantic hierarchies (e.g. "gender" as used as a designation for gender equality as well as themes such as gender-based violence). At a more complex level, decisions were needed to determine parsimonious narrowing of possibly overlapping themes and ambiguous terminology. For example, decisions need to be made around including wider terms such as 'education' or 'employment' versus more specific phrasing such as 'access to education' or 'access to employment opportunities'.

At an even more complex level, some focal dimensions involve even more variation in terminology. One example is that of the overlap between terms such as 'benefits' (or 'unemployment benefits'), 'welfare' (or 'social welfare') and 'social security', which not only vary in their usage by different actors, but also (in the case of the former two) don't have unique explicit meanings outside of context. Related terms such as 'food stamps' are also not captured without deeper customization (NB: in our proof-of-concept here – we focus on 'minimum wage' and ' universal credit' within this sphere, the latter which involves established policy in the UK and by extension, occasionally triggers discussion in other geographic locations).

Other considerations involve the separation of interrelated themes in a manner different to that proposed by Gardner et al. (2020). For example, here we separate legislation and enforcement, and also recognise potential overlap in dimensions originally proposed in this framework (e.g. labor rights and anti-corruption) via other dimensions (e.g. legislation). It is also important to consider the way in which trends vary according to the nature of underlying dimensions (e.g. sustainable business shaped by 'slower' themes such as legislation – which is in itself a specific individual term with a variety of possible contextual connections within the dataset). Another complicating factor is the use of specific wording for campaign purposes (e.g. 'spot the signs' is campaign specific (e.g. #spotthesigns) – but perhaps does not capture the wording used in other similar campaigns outside the UK for example).

A final, and very important complicating factor is the need to consider further sub-themes within hierarchies. For example, we know 'healthcare' is an important higher-level keyword, but we also know that 'mental health' is a very context-specific important keyword (see: Unseen, 2020).

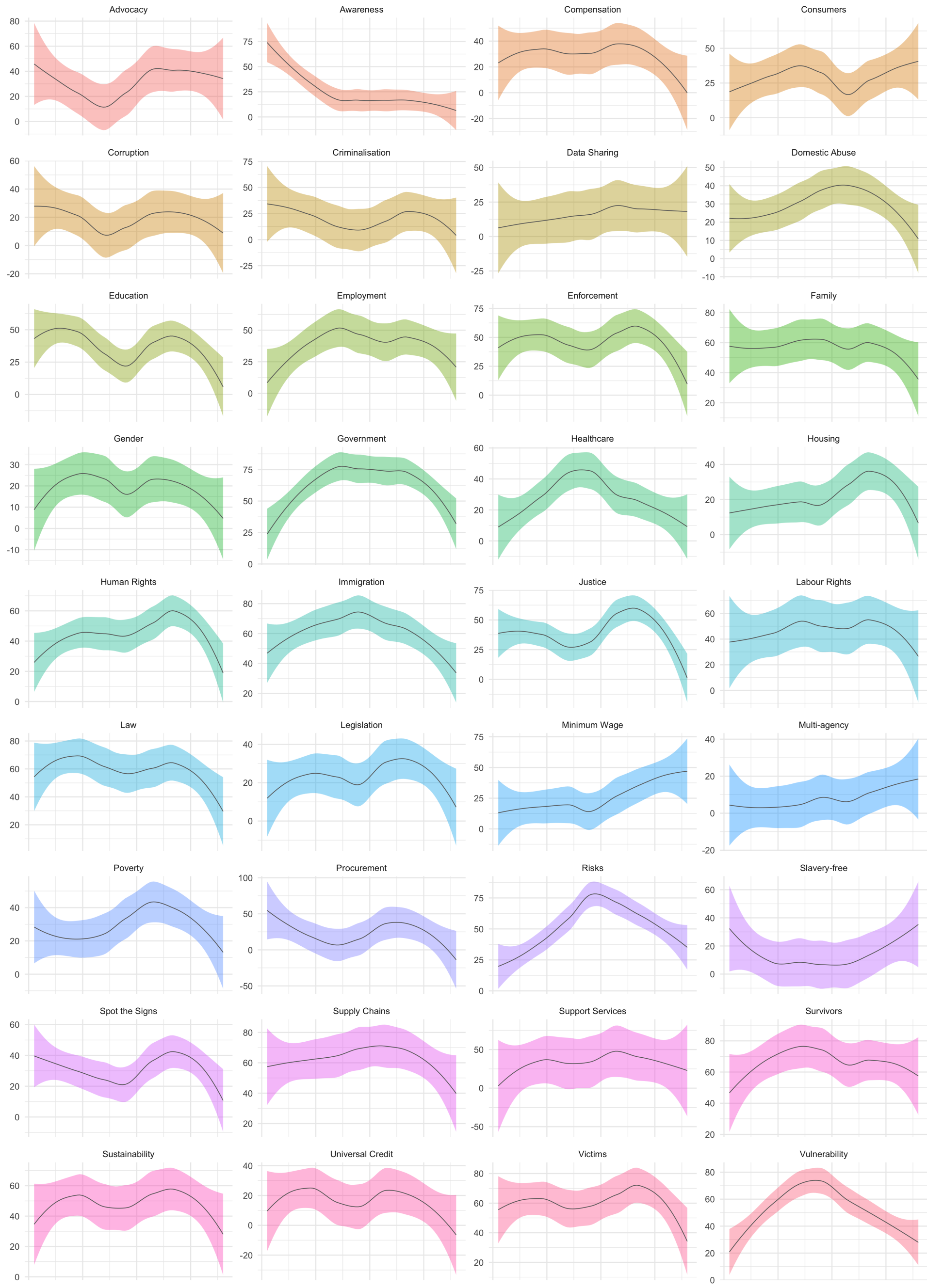

Figure 5: Examples of Emergent Trends (Custom Score)

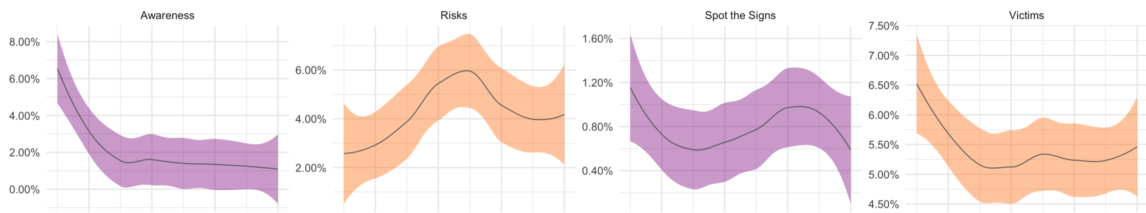

Figure 6: Examples of Emergent Trends (Percentage)

4 Concluding Remarks

Working towards new developments in the development and deployment of novel data analytics approaches within human rights and modern slavery research (see: Landman, 2020), this working paper highlights the opportunities for new research leveraging text mining-based information retrieval approaches to support both trend analysis (see: Lucas and Landman, 2020; Lucas et al. 2020), and narrative-based approaches (see: Choi-Ftitzpatrick, 2017; Nicholson et al., 2018). The latter can also be supported by *data storytelling* (Ojo and Heravi, 2018), such as via simple thematic mapping based on trend similarity measures (such as is briefly illustrated in Figures 7 and 8 - where we use a custom similarity measure to derive a maximal spanning tree of relationships between our extracted dimensions).

Our next steps include, (a) the deeper integration of human expertise into more detailed dataset query customization, (b) detailed time series analysis, and (c) the development of theorizing around key endogenous and exogenous variables.

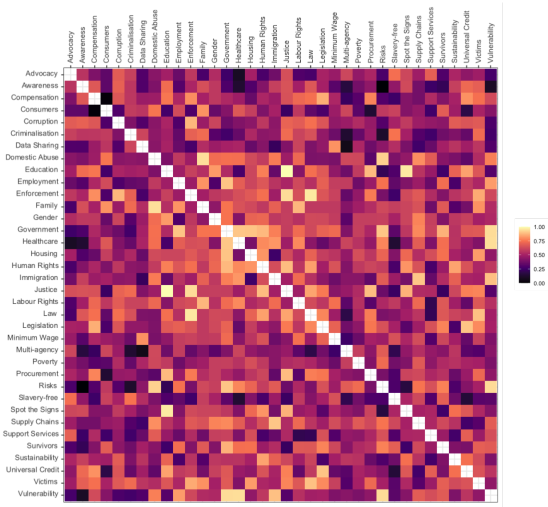

Figure 7: Dimension Similarity Matrix

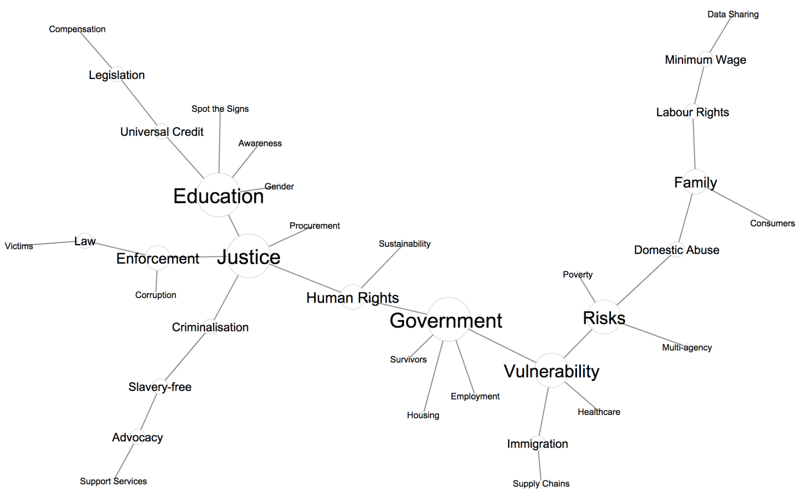

Figure 8: Dimension Theme Map

5 References

Choi-Fitzpatrick, A., 2017. What slaveholders think: How contemporary perpetrators rationalize what they do. Columbia University Press.

Gardner, A., Northall, P. and Brewster, B., 2020. Building Slavery-free Communities: A Resilience Framework. *Journal of Human Trafficking*, pp.1-16.

John Hopkins University. 2020. Coronavirus Resource Center. Retrieved from: https://coronavirus.jhu.edu

Kearney, M.W., Heiss, A. and Briatte, F., 2020. R Package 'rtweet'. [Link]

Landman, T., 2020. Measuring Modern Slavery: Law, Human Rights, and New Forms of Data. *Human Rights Quarterly*, 42(2), pp.303-331.

Lucas, B., Bravo-Balsa, L., Brotherton, V., Wright, N., and Landman, T., 2020. Dashboarding The Online Strategic Communications of Anti-slavery Organizations During COVID-19. *SocArXiv*. Retrieved from: https://doi.org/10.31235/osf.io/e9rbh

Lucas, B. and Landman, T., 2020. Social Listening, Modern Slavery and COVID-19. *Journal of Risk Research*. [Link]

Nicholson, A., Dang, M. and Trodd, Z., 2018. A Full Freedom: Contemporary Survivors' Definitions of Slavery. *Human Rights Law Review*, 18(4), pp.689-704.

Ojo, A. and Heravi, B., 2018. Patterns in award winning data storytelling: Story types, enabling tools and competences. *Digital Journalism*, 6(6), pp.693-718.

Unseen. 2020. The Impact of CV-19 on Modern Slavery Survivors. Retrieved from: https://www.unseenuk.org/news/95

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D. and RStudio. 2020., R package ggplot2. [Link]