



BIROn - Birkbeck Institutional Research Online

Ballatore, Andrea and Katerinchuk, Val and Poulouvassilis, Alex and Wood, Peter (2023) Tracking museums' online responses to the Covid-19 pandemic: a study in museum analytics. *ACM Journal on Computing and Cultural Heritage* , ISSN 1556-4673.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/51338/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

Tracking museums' online responses to the Covid-19 pandemic: a study in museum analytics

Author version • To appear in
ACM J. Comput. Cult. Herit. in 2023

ANDREA BALLATORE, King's College London

VALERI KATERINCHUK, ALEXANDRA POULOVASSILIS, and PETER T. WOOD, Birkbeck, University of London

The COVID-19 pandemic led to the temporary closure of all museums in the UK, closing buildings and suspending all on-site activities. Museum agencies aim at mitigating and managing these impacts on the sector, in a context of chronic data scarcity. "Museums in the Pandemic" is an interdisciplinary project that utilises content scraped from museums' websites and social media posts in order to understand how the UK museum sector, currently comprising over 3,300 museums, has responded and is currently responding to the pandemic. A major part of the project has been the design of computational techniques to provide the project's museum studies experts with appropriate data and tools for undertaking this research, leveraging web analytics, natural language processing, and machine learning. In this methodological contribution, firstly, we developed techniques to retrieve and identify museum official websites and social media accounts (Facebook and Twitter). This supported the automated capture of large-scale online data about the entire UK museum sector. Secondly, we harnessed convolutional neural networks to extract activity indicators from unstructured text in order to detect museum behaviours, including openings, closures, fundraising, and staffing. This dynamic dataset is enabling the museum studies experts in the team to study patterns in the online presence of museums before, during, and after the pandemic, according to museum size, governance, accreditation, and location¹.

CCS Concepts: • **Applied computing** → **Arts and humanities**; • **Information systems** → **Digital libraries and archives**; • **Computing methodologies** → **Knowledge representation and reasoning**.

Additional Key Words and Phrases: museum studies, social media, web analytics, web science, Covid-19 pandemic

ACM Reference Format:

Andrea Ballatore, Valeri Katerinchuk, Alexandra Poulouvassilis, and Peter T. Wood. 2022. Tracking museums' online responses to the Covid-19 pandemic: a study in museum analytics. *ACM J. Comput. Cult. Herit.* 1, 1, Article 1 (January 2022), 30 pages. <https://doi.org/10.1145/3343871>

1 INTRODUCTION

Museums are a vital part of the UK's cultural and economic landscape. There are currently over 3,300 open museums, ranging from one-room displays focussed on a narrow topic to the UK's large national-level museums [11]. In 2019, UK museum activities generated an estimated revenue of £810 million². However, there is deep concern

¹For the purposes of open access, the authors have applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission.

²<https://www.statista.com/statistics/433813/museum-activities-enterprises-turnover-uk-united-kingdom/>

Authors' addresses: Andrea Ballatore, andrea.ballatore@kcl.ac.uk, King's College London, The Strand, London, United Kingdom, WC2R 2LS; Valeri Katerinchuk, v.katerinchuk@bbk.ac.uk; Alexandra Poulouvassilis, a.poulouvassilis@bbk.ac.uk; Peter T. Wood, p.wood@bbk.ac.uk, Birkbeck, University of London, Malet Street, London, United Kingdom, WC1E 7HX.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

XXXX-XXXX/2022/1-ART1 \$15.00

<https://doi.org/10.1145/3343871>

that many museums will not survive the impact of the COVID-19 pandemic, with a correlative loss to the UK’s cultural and economic landscape.

Although UK museum agencies are providing funding and developing policies to manage the impact of COVID-19, they do not have established mechanisms for gathering comprehensive data on the UK museum sector, for tracking the ways in which museums have responded to the pandemic. The “Museums in the Pandemic” (MIP) project³ — seeks to provide timely data on which museums close, which remain resilient, and how the profile of the UK museum sector changes as a result of COVID-19. The project’s research draws on multi-disciplinary expertise from museum studies, computer science, data science and geographical information science, and combines both quantitative and qualitative methods, including natural language processing, machine learning, data visualisation, interview-based research, and primary data collection.

In the earlier “Mapping Museums” project, conducted by the authors and their collaborators⁴, we gathered authoritative longitudinal data on the entire UK museum sector, from 1960 to date, including the dates of museum permanent closures. That data is publicly available⁵ through a browsable database, search app and visualisations. However, it is missing data relating to museums’ responses to the COVID-19 pandemic and hence cannot be used to investigate how the profile of the UK museum sector may be changing as a result of COVID-19. This need has motivated our research into extracting and analysing data from museums’ public websites and social media posts in order to investigate museums’ responses to the COVID-19 pandemic. In this context, museum closures are particularly significant events, and several types of closure were observed: temporary, indefinite, or permanent.

This paper focuses on the design, implementation and evaluation of computational techniques we have developed in the MIP project, utilising data scraped from museums’ public websites and social media posts to detect a set of high-level activity indicators for museums that are relevant to the COVID-19 pandemic. As social media platforms, we considered Facebook and Twitter, which are used respectively by 51 and 19 million people in the UK⁶. We have developed visualisations to track occurrences of these indicators over the duration of the MIP project, across all museums and split according to one or more key museum attributes such as Governance, Size, Subject matter, Accreditation status and Location⁷, in order to allow the project’s museum studies experts to assess whether the pandemic is having a disproportionate impact on particular types of museums.

The contributions of this paper are our methodology and methods for extracting meaningful data from web-based resources relating to a large number of institutions. Our research is novel in aiming to analyse the behaviour of the entire UK museum sector at an institutional level, in contrast to works that focus on aspects such as cultural artefacts, cultural production, museum services, visitors, and visitor-generated content (see Related Work). These methods are transferable to large-scale analysis of other aspects of museums’ online content, beyond our specific focus here on the response of the UK’s museums to the COVID-19 pandemic. While this article has a methodological focus, detailed analyses of online content production and engagement are conducted in forthcoming studies by the MIP project team [25, 26]. The data and materials used in this article are available online as open data (see Data Availability Statement).

1.1 Related Work

We view our research as falling into the broad field of Digital Humanities [15], lying at the intersection of traditionally qualitative humanistic disciplines, such as museum studies, and the use of digital methods and data science. More specifically, our research relates to and contributes to:

³Funded by the UKRI-AHRC Rapid Recovery Scheme, Grant No. AH/V015028/1, “UK Museums during the COVID-19 crisis: Assessing risk, closure, and resilience”, January 2021-December 2022

⁴Funded by AHRC grant Ref AH/N007042/1 October 2016 – September 2021

⁵<https://www.mappingmuseums.org>, accessed in Feb 2023.

⁶<https://napoleoncat.com/stats>, accessed in Feb 2023

⁷We refer readers to the Glossary on the Mapping Museums website for definitions of these terms, at <https://museweb.dcs.bbk.ac.uk/glossary>.

- *web science*, which studies the structure and evolution of online resources, linkages and content [18] – in our case, museums' websites and social media posts as relating to their responses to the COVID-19 pandemic;
- *web analytics*, which seeks to understand users' online behaviour at a large scale [39] – in our case, the “user” entities being the museums in the UK; and
- *social media analytics*, which seeks to gain insights into users' behaviours, sentiments and preferences on social media platforms – in our case, how such platforms are being used by the UK's museums to communicate their responses to the COVID-19 pandemic – through the application of web scraping, data cleansing, and data analysis and visualisation techniques [6, 47].

A related area is that of *cultural analytics* which uses data science techniques and “big data” to study cultural artefacts and cultural production at large scale [31]. However, in contrast to this, our research falls into a new application area that we term *museum analytics*, characterised by the large-scale application of data science methods to analyse museums' online presence via their institutional websites and social media at the scale of an entire museum sector (the UK museum sector) rather than to study cultural artefacts or production.

A number of papers perform analyses of museum collections or other data held by museums. For example, the collection of papers in [7] investigates the classification of cultural assets, the analysis and retrieval of visually linked paintings, image reconstruction methods, enhancing the end-user experience, the analysis and restoration of historical manuscripts, and the use of named entity recognition methods for the analysis of historical text. The chapter on “Museum Big Data” [37] reviews methods and techniques to identify new and uncover hidden information, patterns, clusters and relationships within museum data. Such data comprises museum artefacts and services, data related to museum visits, and visitor-generated data on the web and social media. Other papers perform analyses specifically on museum visitors. For example, a classification of online visitors into six categories using both web analytics and traditional surveys is undertaken in [49]. One goal of the work is to suggest to museums that the online experience should be differentiated for the various categories of visitors. By contrast, physical visitors to museums are studied in [50]. Here a multilevel logistic model is used to show that the places where individuals reside impact on museum participation.

Several studies apply machine learning to explore museum data about collections and visitors [30]. Shao et al. [46] use topic modelling on online reviews to understand visitor experiences in a London museum. Sentiment analysis can produce insights about online opinions about tourist destinations, including museums, at a large scale [8]. Machine learning is also deployed to make museum collection data more accessible, inter-linking entities as Linked Open Data [13]. Social media analytics can also provide creative input to museum curators, extracting concepts and ideas from visitor comments and responses [14]. Compared to these studies, our research is concerned with neither museum assets nor visitors. Instead, our focus is on analysing information posted by museums themselves, whether on their websites or through their social media channels, as it relates to the COVID-19 pandemic.

There have been numerous studies investigating museums' responses to the pandemic (summarised in the following paragraphs). In contrast to our work, these are relatively small-scale in relation to the size of the museum sector concerned, sampling a selection of a country's or countries' museums, whereas we aim to extract and analyse data from museums' online presence covering the entire UK museum sector.

Some studies have considered museums' responses through their web-based activities:

Using a geographically representative sample of UK museums, King et al. [21] identified 88 temporary exhibitions that would have opened during the first lockdown in the UK and analysed the 21 online exhibitions that were put in place as a substitute. They identified themes of access, embodiment and human connection emerging from the exhibition content, and raised questions around the conceptualisation, presentation and value of digital collections. Samaroudi et al. [45] analysed the web-based content produced during the first lockdown period

(April-July 2020) by a sample of 48 UK and 35 USA “memory institutions”, such as museums. They identified trends in how institutions restructured their digital content in terms of different types of content, museums and audiences, and made recommendations on how institutions could enhance the value of their digital content. Burke et al. [10] discussed three types of digital content offered by a small sample of Norwegian and international museums during the pandemic: virtual tours, online exhibitions, and crowdsourced art creation. Gutowski and Kłos-Adamkiewicz [17] evaluated a sample of 136 virtual tours of Polish museums and monuments taken in April 2020, finding no significant increase in digital content compared with a pre-pandemic sample taken in August 2019. Jin and Min [19] examined how over 1,300 Chinese museums provided new online exhibitions, educational programmes and livestreaming services to connect with their audiences during the first pandemic lockdown, analysing these offerings and their effects on audience engagement and making recommendations for museums’ ongoing development of their digital communication strategy. Raimo et al. [42] studied the effect of the pandemic on the digitisation processes of three Italian museums, finding more frequent website updates, increased use of social media, and more virtual exhibitions.

Several studies have considered museums’ responses through their use of social media:

Agostino et al. [3] analysed data from Italy’s Ministry for Cultural Heritage and Tourism about the hundred most-visited Italian state museums, finding a doubling of these museums’ use of social media during lockdown and creation of new types on online content. Magliacani and Sorrentino [29] surveyed 34 Italian University museums to investigate how they maintained audience experience during lockdown, finding that the time they spent on social media management did not change significantly but that over half offered video narrations of their collections and over a quarter offered live online events. Kyprianos and Kontou [24] undertook a questionnaire-based survey of 101 museums in the Attica region of Greece, finding that most of the 52 museums that responded had increased the time they spent on social media management during the pandemic and 50% had seen a moderate or significant rise in user traffic on their social media accounts. Ryder et al. [44] surveyed the types of digital content 66 cultural institutions in the USA produced during lockdown and the effect on audiences’ engagement with their social media accounts; they found an increase in both live and serialised digital content and almost all institutions reporting an increase in social media engagement.

Other studies include [28], who conducted interviews with ten UK museum operations professionals to investigate how museums dealt with the initial stages of the pandemic, identifying themes of emotional impact on employees, importance of staff adaptability and flexibility, disappointment with the crisis response of the UK governments, and mutual support of professionals in the sector; and [32], who surveyed approximately 1500 Italian museums to investigate whether they activated new digital channels during lockdown, finding only a minority of museums were working towards the provision of new digital activities.

Internationally, major bodies such as UNESCO, ICOM and NEMO have surveyed samples of museums from many countries to investigate trends in museums’ behaviours during the pandemic. UNESCO [48] and ICOM [36] reported on the increased online presence of museums across a variety of digital activities, the economic impact of the pandemic, and concerns about laying-off of staff and permanent museum closures. NEMO [35] similarly reported an increase in digital services in a majority of museums, increased online visits, and significant loss of income caused by closure during lockdown. Summaries of these and other studies are presented by Noerher et al. [34] and Raved and Yahel [43].

1.2 Outline of the paper

The remainder of the paper is structured as follows. Section 2 highlights the key research challenges faced in undertaking the work presented here. Section 3 presents our research methodology and computational methods for identifying museums’ official websites and social media accounts, collecting data from these websites/accounts, and detecting occurrences of a set of indicators relevant to the COVID-19 pandemic in the data. Section 4 describes

our implementation of these computational methods. Section 5 presents statistics and visualisations allowing the MIP project's museum studies experts to explore trends in the indicator occurrences. Section 6 describes an additional interactive tool co-designed with these experts allowing them to search the data for specific phrases and providing a range of statistics to support further analysis. Section 7 gives our concluding remarks and identifies directions of further research.

2 RESEARCH CHALLENGES

In order to track museums' responses to the COVID-19 pandemic, the MIP project faced a number of significant challenges:

- The combination of research objectives, the sheer scale of data collection, and time constraints made the research design challenging. Based on our experience with the previous Mapping Museums sector-wide survey [11, 40], qualitative methods would have been hard to deploy in the required time frame. Furthermore, even mixed research methods would not have been suitable to produce a sector-wide overview, beyond individual case studies. For this reason, we adopted a big data approach, relying on machine learning, web scraping and social media analytics to build the dataset. This approach produced consistent, granular, and longitudinal data about the entire UK museum sector.
- Secondly, the correct website for each museum needed to be identified so that museums' online posts relating to the pandemic could be discovered. With over 3,300 currently open museums, this could not be achieved manually. An automated solution has to take into account that several different websites might refer to the same museum (including, for example, tourism-related websites) and also that some museums do not have their own websites but are instead included on the websites of other organisations (e.g., the National Trust). Our machine learning-based solution to automatically discover museums' websites is described in Section 3.2.
- Similarly, to be able to monitor museums' social media posts relating to the pandemic, we needed to identify museums' Facebook and Twitter accounts. Again, due to the number of museums involved, an automated approach was required. Our solution utilises a combination of the ML techniques we developed for identifying museums' websites along with the extraction of embedded links to Facebook and Twitter accounts from these websites, as described in Section 3.3.
- After collecting museums' website and social media content (described in Section 3.4), we needed to be able to detect museums' posts relating to the pandemic within this content. We started by extracting manually from a small set of websites all linguistic phrases indicative of museums' activities relating to the pandemic, with the aim of classifying the diversity of such linguistic phrases into a smaller set of 'activity indicators' that would capture the full semantic scope of the language used but that would be more clearly indicative of the distinct semantic themes (e.g. museum open, museum closed, online engagement) and their sub-divisions. To determine whether or not a web page or social media post includes text that matches one of these linguistic phrases required investigation of a number of ML classifiers, described in Section 3.5, with convolutional neural networks emerging as the top performing model.
- Interdisciplinarity: The MIP project needed to design analyses and visualisations that are useful to the project's museum studies experts in investigating the project's research questions. We employed again the participatory iterative approach pioneered in the earlier Mapping Museums project [40] in which all research stakeholders are involved at all stages, so as to ensure that all stakeholder viewpoints and requirements are reflected in the development of the research. Similarly to the earlier MM project, the MIP project exhibits the characteristics of Broad, Cooperative, Integrated, Methodological, Bridge-Building, Instrumental and Exogenous, from Klein's taxonomy of interdisciplinary research [23].

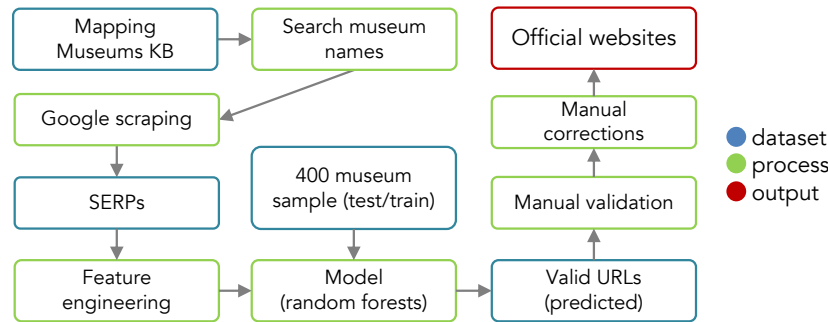


Fig. 1. Workflow of the identification of official museum websites (where SERPs denotes “search engine result pages”)

3 METHODOLOGY

3.1 Research process overview

The MIP project team comprised specialists from computer science, data science and geographical information science (whom we term the “technical team” below) and from museum studies (whom we term the “museum studies experts”). The team began by identifying a sample of 62 museums that had made public statements on their websites and social media about the effects the COVID-19 pandemic was having on them. All 62 museums were present in the database created by the earlier Mapping Museums (MM) project and we identified their website and social media URLs through a manual web search. To increase its representativeness, this sample of 62 museums was stratified according to the museums’ size and location attributes (as recorded in the MM database). Henceforth we call this the *initial sample*.

This initial sample was subsequently used to (i) gain an understanding of the format of the URLs of museums’ official websites so as to enable their automatic discovery (see Section 3.2), and (ii) identify key linguistic phrases used by museums to refer to the effects the COVID-19 pandemic is having on them (see Section 3.5). After the identification of the initial sample, we proceeded to automatically identify the URLs of the websites of open museums (3,344 open museums), exploring both knowledge-based and machine learning (ML) methods, as described in Section 3.2. We repeated this effort to identify the museums’ Facebook and Twitter accounts, using a combination of direct web page scraping and predictive modelling, as described in Section 3.3. Having identified the target websites and social media accounts of the museums, our next step was to periodically extract data from this corpus, and to apply natural language processing and ML techniques to detect the presence of activity indicators within the data, as described in Sections 3.4 and 3.5, respectively.

3.2 Identification of official museum websites

3.2.1 Museum data. The MM database provided the ideal starting point to ensure high coverage of UK museums in our research [40]. This resource took 4 years to compile, and in doing so all likely sources of museum data were tracked down. As of January 2021 – which is when the MIP project began – the MM database included information about 4,166 museums, including closed ones. Museums that had closed before January 2020 (i.e. before the start of the COVID-19 pandemic in the UK) were not considered in our research, leaving a total number of 3,344 open museums to be considered in our analyses. While the museums’ websites were recorded for a small number of museums in the MM database, the vast majority of museum websites had not been recorded by the earlier MM project and therefore we needed to devise a systematic method for discovering them. Given the large

number of open museums, a manual web search process was deemed impractical and we proceeded to devise an automated method, whose workflow is summarised in Figure 1 and described below.

3.2.2 Discovering official museum websites. The challenge faced in this step is the definition of the best website that represents a museum. For example, <http://www.britishmuseum.org> is the official website of the British Museum, while <https://www.gov.uk/government/organisations/british-museum> is relevant and maintained by the British Government, but does not capture salient information for our analysis. Hence, we defined “official websites” as websites directly maintained by museums, as opposed to any other websites containing information about them. Ideally, an official website should be exclusively about a single museum to help the data extraction process. However, some large organisations manage multiple museums and host their official pages on a single website, as is the case for Tate (tate.org.uk) and the local authority museums in Birmingham (birminghammuseums.org.uk).

3.2.3 Searching for museum websites. To identify official websites, we used Google Search. We extracted all museum names from our initial sample, generating a set of alternative Google searches for each, including: (i) the museum name, (ii) if the museum name does not contain the word “museum” then the museum name followed by “museum”, and (iii) the museum name followed by the location. For example, for the Michelham Priory, we generated three searches: “Michelham Priory”, “Michelham Priory museum,” and “Michelham Priory, Hailsham.”

Many of the resulting top-ranked results appeared to contain the official websites, so we proceeded to generate searches for all 3,344 open museums⁸. The search queries for websites and the two social media platforms were executed using Google Search from the UK in February 2021, retrieving a maximum of 50 results per query. Each search engine result page (SERP) contains URLs ranked from 1 to n , where 1 is the most relevant result. This process produced a total of 198,134 URLs, with an average of 20 URLs per museum⁹.

These search results had to be classified as valid or invalid. An inspection of the resulting URLs revealed three recurring situations: (i) the official website is ranked as 1; (ii) the official website is present, but is not the top result; (iii) the official website is not present and all results are therefore invalid. To handle these cases, we performed feature engineering to characterise each candidate URL with a set of variables that predicted their validity as the official website. Table 1 outlines the selected variables, calculated for all 198,134 URLs and their corresponding museums.

Given the high variability in museum names, frequent usage of abbreviations and alternative names, we conducted the similarity matching not only between the URL and the primary museum name but using a pool of possible name variants constructed using various grammatical rules. After developing several iterations of such algorithms, the variety of cases emerged as too wide to be handled with a knowledge-based approach, leading to results of insufficient accuracy. A machine learning (ML) approach was therefore explored in order to improve the quality of results.

3.2.4 ML model to identify official museum websites. As a first step to developing an ML model to identify the correct website for each museum, we produced a training and test dataset. In this process, we did not distinguish between museums having dedicated websites or pages on larger websites, opting for a binary approach (i.e., web presence or not). A key decision concerned the size of the dataset necessary to obtain sufficiently reliable estimates of the results' accuracy¹⁰. We therefore calculated the margin of error as $MOE_{\gamma} = z_{\gamma} \sqrt{\frac{\sigma^2}{n}}$, where γ is the

⁸Due to Google's Terms of Service, the results of these searches cannot be shared and were used exclusively for this not-for-profit project.

⁹Among the top-ranked websites, we observed large museum information aggregators such as whichmuseum.com that are well-optimised for Google, but not official museum websites according to our definition. About 80 highly-ranked tourism websites were also removed, mostly having the word “visit” in their URL, e.g., visitbelfast.com and visitscotland.com. Additionally, we excluded URLs including the keywords “tripadvisor”, “wikipedia”, “facebook”, “instagram”, “twitter”, “google”, and “expedia” that appeared frequently in the top 10 results.

¹⁰We adopt the standard definitions of *accuracy* as $(TP+TN)/(TP+FP+FN+TN)$, where TP=true positives, TN=true negatives, FP=false positives, FN=false negatives.

Table 1. Variables characterising URLs from Google results to classify them as official websites or not. These variables were calculated for 198,134 URLs, and then used as input in a Random Forest model.

Variable	Description	Type
<i>URL features</i>		
Google rank	Google ranking from 1 to n .	Numeric
URL size	Number of characters in URL.	Numeric
URL no. of /	Number of ‘/’ symbols in URL. Higher numbers indicate a deep folder structure.	Numeric
Has visit	The URL contains the keyword ‘visit’. True for URLs such as visitengland.com .	Boolean
Has museum	The URL contains the keyword ‘museum’. True for URLs such as yorkarmymuseum.co.uk .	Boolean
Museum location	The URL contains the location name (village town or city) of the museum.	Boolean
<i>Similarity between URL and museum name</i>		
URL similarity	Levenshtein distance similarity ratio between ($url, name$). Score in range [0, 100], where 100 means identical strings. The calculation returns the highest scores from a pool of name variants.	Numeric
Inverse URL similarity	Same score calculated on ($name, url$).	Numeric
Domain similarity	Same score calculated on ($domain, name$).	Numeric
Inverse domain similarity	Same score calculated on ($name, domain$).	Numeric
Fuzzy URL similarity	Partial similarity ratio, calculated as the maximum Levenshtein distance similarity ratio between a shorter string and every substring of length m of the longer string. Score in range [0, 100]. The calculation returns the highest scores from a pool of name variants.	Numeric
Inverse fuzzy URL similarity	Same score calculated on ($name, url$).	Numeric
Fuzzy domain similarity	Same score calculated on ($domain, name$).	Numeric
Inverse fuzzy domain similarity	Same score calculated on ($name, domain$).	Numeric
<i>Museum features</i>		
Museum size	Size of the museum from ‘small’ to ‘huge’, based on the estimated number of yearly visits. Data from Mapping Museums.	Categorical
Museum governance	Type of museum governance, including ‘government’, ‘university’, and ‘independent’. Data from Mapping Museums.	Categorical

confidence level at 90%, z at 1.96 (95%), and n is the sample size. For example, when $n = 100$, the corresponding MOE is 5.6%.

In the context of this study, we set a maximum MOE of 3%, requiring $n > 384$. This MOE implies that, if the estimated accuracy is 90%, the real value falls between 87% and 93%, which was deemed satisfactory by the project’s museum studies experts for the project’s research purposes. We therefore produced a stratified sample

of 400 open museums (corresponding to 12% of the total number of 3,344 open museums). The stratification was conducted by museum size, governance, accreditation and region (i.e. Scotland, Wales, Northern Ireland, and the nine English Regions, as recorded in the MM database) ensuring that the sample was representative across these key attributes. For each museum in the stratified sample, we generated the top 10 results from Google Search for websites, Facebook, and Twitter. Two of the project's museum studies experts annotated the resulting dataset, classifying each Google Search result as 'valid' or 'invalid'. When a museum did not have an official website or Facebook or Twitter account, it was flagged as 'no_resource'.

3.2.5 Learning valid museum websites. Harnessing the 400 museum URL dataset for model training and validation, we investigated different approaches to identifying the museum websites. Among a number of traditional models, including linear regression, logistic regression and neural networks, random forests emerged as the most appropriate [9]. This method relies on bootstrap re-sampling and training decision trees on the samples in an iterative manner. With our relatively small training dataset and the presence of co-linear variables (particularly between the string metrics), this method reduces over-fitting and handles the co-linearity between features effectively. The variables described in Table 1 were used as predictive attributes in the random forests.

Table 2. Test results for the identification of official museum websites, calculated over different train-test splits on the 400 museum URL dataset. Each result is the mean of 100 trials.

Training/test split	50/50	60/40	70/30	80/20	90/10
Mean precision	0.91	0.94	0.95	0.94	0.95
Mean accuracy	0.99	0.98	0.99	0.99	0.99
Mean sensitivity	0.92	0.92	0.93	0.96	0.97
Mean specificity	0.99	0.99	0.99	0.99	0.99

In order to evaluate the performance of random forests on our data, we tested five different train-test splits on the 400 museum URL dataset, calculating precision, accuracy, sensitivity and specificity, as displayed in Table 2¹¹. Given the stochastic nature of this method, we ran each model 100 times, keeping the mean value. These results showed how effective the random forests were, reaching results higher than 0.9 in all performance indicators, with minor differences between the splits. This level of accuracy was considered satisfactory for the project's context, and we trained the classification model on the 80/20 split to provide balanced data. The model was then applied to the set of 198,134 URLs, obtaining a final dataset of 3,295 museum websites. These websites represent 98.5% of the 3,344 museums, indicating that almost all museums maintain a web presence in the form of either a dedicated website or a set of pages on a larger website.

3.3 Identification of social media accounts

3.3.1 Platform selection. To study museums' presence and behaviour on social media, we selected Facebook and Twitter as data sources. This choice was based on the fact that these platforms reach a significant portion of the British population and have been dominant for over a decade. In particular, Twitter users tend to have attained a higher level of formal education than the general population [51] and Facebook is widely adopted among older Internet users [4]. Both of these groups are well over-represented in museum audiences [1]. Among other relevant platforms, Instagram and TikTok exhibit faster growth rates and attract a younger population, but their data feeds are harder to access and more video-oriented, making them difficult to analyse at a large scale. Hence, we confined the study to Facebook and Twitter text data, leaving other platforms and media for future studies.

¹¹We define *precision* as $TP/(TP+FP)$, *accuracy* as $(TP+TN)/(TP+FP+FN+TN)$, *sensitivity* as $TP/TP+FN$, and *specificity* as $TN/TN+FP$, where TP=true positives, TN=true negatives, FP=false positives, FN=false negatives.

```

for all museumid do
  if MLU[museumid] == "" then                                     ▶ There is no machine-learnt URL
    URL[museumid] = SU[museumid]
  else                                                           ▶ There is a machine-learnt URL
    if SU[museumid] ==  $\emptyset$  then
      URL[museumid] = { MLU[museumid] }
    else
      if MLU[museumid]  $\in$  SU[museumid] then
        URL[museumid] = { MLU[museumid] }
      else
        URL[museumid] = SU[museumid]  $\cup$  { MLU[museumid] }
      end if
    end if
  end if
end for

```

Fig. 2. Algorithm for combining scraped URLs with machine-learnt URLs.

3.3.2 *Account detection.* While the above ML model appeared very effective for websites, the results were less encouraging for museums' Facebook and Twitter accounts, with accuracy falling into the mid to high 80% range, depending on the test-train split. Thus it was decided to utilise this predictive model to identify the museums' websites and to scrape these websites in order to obtain Twitter and Facebook links to each museum's social media pages. We then utilised the links scraped from the museums' web pages in combination with the predictive model results for museums' Facebook or Twitter accounts, as detailed below. This method, while developed for Facebook and Twitter, can be applied to other platforms.

For each of Facebook and Twitter, the overall process for associating a set of URLs with each museum was as follows. For the stratified sample of 400 museums, a single, manually validated URL was used as the correct URL. For the remaining museums, two arrays, each indexed by museum id, were created: SU, representing the set of scraped URLs associated with each museum, and MLU, representing the single URL discovered by the ML model. An entry in SU might be the empty set, while an entry in MLU might be the empty string. We noted that certain scraped URLs appeared multiple times in SU. It turned out that these URLs corresponded to generic social media websites, rather than ones specific to a museum. As a result, we decided to remove from the sets of URLs in SU all URLs that occurred three or more times overall. After this, the SU and MLU arrays were "merged" using the process described in Figure 2 to produce an array named 'URL' which associates with each museum a (possibly empty) set of URLs. We allow multiple URLs to be associated with a museum because we noted when validating results that a number of museums had multiple valid social media accounts for both Twitter and Facebook.

The logic depicted in Figure 2 is as follows. If there was no machine-learnt URL for a museum, then the final set of URLs for the museum was the set of scraped URLs (even if this set was empty). If there was a machine-learnt URL, then that was used as the final set if there were no scraped URLs. When there was both a machine-learnt URL and a set of scraped URLs, the machine-learnt URL was used as the final set if it was among the scraped URLs; otherwise, the final set was taken as the union of the scraped and machine-learnt URLs. The reasoning for this final step was as follows: if the machine-learnt URL was among the scraped URLs, this was a strong indication that the URL was correct; if it was not, there was no way to be sure which of the URLs were valid and so all were retained. The final accuracy achieved with this method was 96.2% for Facebook and 98.5% for Twitter, producing high-quality input for data collection from these platforms.

3.4 Data collection

3.4.1 Scraping official museum websites. Having identified the target 3,295 museum websites with high confidence, we implemented a web scraping tool to periodically and automatically collect this corpus¹². Starting from the website URLs as seeds, the scraper visits the landing page, stores its content, and then proceeds to visit all links found on the page within the same domain. This strategy significantly reduced the collection of non-relevant content, as most pages had a number of links to non-official museum resources and to social media accounts, which were collected separately. Each scraped page was subsequently processed to extract relevant information for our analysis, i.e. text visible in the page body, excluding JavaScript code and other non-visible elements.

The first scraping session occurred on 4 March 2021 and was then repeated on a bi-weekly basis up to May 2022 (14 months). The database storing the corpus contains the complete source code of each page, with a separate table for each snapshot, enabling a variety of future text and content analyses. The median of scraped pages in each snapshot is approximately 102,000, for a total of about 5,700,000 pages. As detailed in the remainder of this paper, this museum text corpus was used to extract activity indicators, as well as to perform lexical searches using regular expressions to study museum behaviour at a large scale during the COVID pandemic.

3.4.2 Social media accounts. Collecting data from Twitter and Facebook required a different approach. Starting with the 2,223 Twitter accounts and 2,547 Facebook pages we identified as being managed directly by museums, we developed software to collect historical tweets and Facebook posts, complying with the platforms' terms of use. The temporal scope of the data collection was set between 1 January 2019 and 31 May 2022, capturing a full year of activity before the beginning of the pandemic in the UK, for a total of approximately 3.5 years. The scope of this data enables a range of analyses, including a before/after comparison to study the effects of the pandemic on online behaviour.

The data collection included all Twitter museum accounts, representing 66% of open museums, retrieving all tweets, re-tweets, and responses for the study period¹³. Additionally, the data included engagement statistics at the account and tweet level, such as likes, retweet counts, and followers. As a result, 3,921,780 tweets were collected from 343,053 unique accounts, of which 72% from museum accounts and the rest from non-museum accounts that interacted with the museum accounts. Replies to tweets amount to 41% of the data. Of 2,223 museum accounts, 1,937 produced at least one tweet and were therefore included in the tweet dataset analysed by the research team.

The 2,547 Facebook pages we identified represent 76% of museums and are primarily used to share posts about the organisation's activities. The posts were collected using CrowdTangle,¹⁴ a social media analytics tool by Meta that allows retrieving Facebook content programmatically. The dataset included 1,466,315 posts from 2,169 unique museum pages, showing that about 85% of the target accounts had some content to be retrieved. These posts are categorised by Facebook as photographs (70%), links (14%), Facebook videos (8%), status updates (4%), and YouTube videos (1%).

These two datasets were the cornerstone of the analyses of social media that the research team performed on museum online behaviour. It is important to note that considerably fewer museums are present on Twitter and Facebook compared to websites.

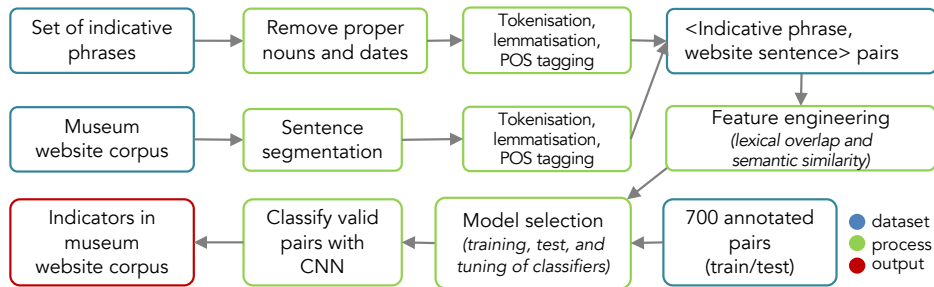


Fig. 3. Workflow of the detection of activity indicators in the museum website corpus, starting from the set of manually identified indicative phrases.

3.5 Extraction of activity indicators

3.5.1 Indicative phrases. The process identifying the occurrence of activity indicators in the museum website corpus is outlined in Figure 3. The first step of identifying the occurrence of activity indicators required examining the text appearing on museum websites to look for references to the pandemic. In February 2021, two museum studies experts from the research team scanned the text occurring on the websites from the initial sample of 62 museums, looking for all linguistic expressions relating to the pandemic. The resulting 278 phrases covered a variety of themes, including closure (e.g., “currently we are closed due to Covid restrictions”), fundraising (e.g., “support us in this pandemic”), and online initiatives (e.g., “we are still telling our stories online”). Given their conceptual importance, we devoted particular attention to distinguishing between different types of closures. Linguistic expressions referred to museums being currently closed, including seasonal closures; closed indefinitely (“until further notice”); and permanently closed. Partial closures were identified indirectly through other expressions (e.g., “only the cafe is open”).

In an iterative process, the whole research team collaborated to derive a set of activity indicators capturing the full semantic scope of this set of 278 phrases. During this process, 36 of the 278 phrases were eliminated as being semantically ambiguous or irrelevant, leaving 242 valid phrases in the dataset. After several iterations, a consensus was reached on 22 indicators which are summarised in Table 3. For each indicator, the table includes its semantic theme, an example of the set of indicative phrases for that indicator, and the number of indicative phrases (No. Phr.) As can be observed, the indicative phrases are unevenly distributed between the indicators (ranging from a maximum of 45 phrases indicative of *fundraising* to just one phrase indicative of *cafe open*, *restructuring staff* and *made Covid safe*). As discussed next, these indicative phrases were used to detect the occurrence of activity indicators in the corpus of museum websites and social media data.

3.5.2 Feature engineering for indicator matching. Given the large size of the text corpus collected (about 5.7 million web pages), it was crucial to identify the presence of activity indicators automatically in a reliable way, harnessing the indicative phrases as training data. The objective of this step consists of classifying input phrases from a website’s unstructured text as containing a certain indicator or not, e.g. ascertaining that the text “we look forward to welcoming you after COVID” expresses *reopen intent* and not *fundraising*.

¹²The web scraper was built with the Python package Scrapy (v2.4), storing the data in a PostgreSQL (v10) database.

¹³The Twitter API v2 endpoint provided the historical tweets through an account with non-commercial Academic Research access, <https://developer.twitter.com/en/docs/twitter-api>, accessed in June 2022.

¹⁴We used the CrowdTangle API with an academic account that provides access to Facebook data, <https://www.crowdtangle.com>, accessed in April 2022.

Table 3. Museums' activity indicators during the COVID-19 pandemic, with examples of indicative phrases and the number of indicative phrases (total phrases 242).

Theme	Indicator	Example indicative phrase	No. phr.
closure	closed currently	<i>we are currently closed to visitors</i>	21
	closed indefinitely	<i>closed until further notice</i>	4
	closed permanently	<i>we have to inform you of the decision to close</i>	4
finances	finance health	<i>we are keeping our survival appeal open</i>	6
	did not get funding	<i>we have been overlooked for grant funding</i>	7
	fundraising	<i>a successful crowdfunding campaign was launched</i>	45
	government emergency funding	<i>receives lifeline grant from Government Culture Recovery Fund</i>	11
	other emergency funding	<i>external financial support has been essential and extremely welcome</i>	4
online activity	online engagement	<i>blog pages, which are regularly updated</i>	12
	online event	<i>we have Virtual Talks</i>	8
	online exhibition	<i>virtual tour now available</i>	7
open/reopen	cafe open	<i>our cafe is open</i>	1
	open currently	<i>Now open for careful visitors</i>	3
	online shop open	<i>Visit our online shop</i>	8
	reopen intent	<i>look forward to welcoming you again soon</i>	35
	reopen plan	<i>We currently plan to re-open</i>	4
staffing	hiring staff	<i>We are hiring new volunteers</i>	8
	restructuring staff	<i>major review and restructuring of its operations</i>	1
	staff working	<i>help our reduced team keep our site safe and clean</i>	11
misc.	language of difficulty	<i>a very challenging time</i>	37
	made Covid safe	<i>Find out how we've made your visit safe</i>	1
	project postponed	<i>Our guided tours have been postponed until later in the year</i>	3

We first pre-processed the indicative phrases to maximise their semantic relevance by removing proper nouns (e.g., museum and city names) and dates. Secondly, we needed to construct a suitable classification model. Current NLP toolkits provide a range of options to classify text, using both supervised and unsupervised methods. Multi-class text classification is an active research area, in which recent approaches based on deep learning generally outperform more traditional ones [33]. Hence, we experimented with the Bidirectional Encoder Representations from Transformers (BERT) model, which achieves state-of-the-art results in many NLP tasks [12].

Adopting BERT with a pre-trained English language model, it is possible to perform fine-tuning on relatively small sets of examples. However, when considering our context, even pre-trained models such as roBERTa [27] were not applicable because of our relatively high number of indicator classes and low number of indicative phrases. Expanding the set of indicative phrases might have been a possibility to enable the use of such methods, but an inspection revealed that the linguistic variation in the indicative phrases was relatively low. Furthermore, on classification tasks, simpler methods based on bag-of-words (BOW) models can be as effective as deep learning [20]. For this reason, simpler classification methods appeared more suitable and we proceeded to devise a BOW-based model, combining lexical matching, semantic similarity, and supervised machine learning classification.

The text from museum websites was parsed, tokenised, lemmatised, part of speech (POS)-tagged, and segmented into sentences¹⁵. For each pair containing an indicative phrase and a museum sentence, the binary classifier must predict whether it constitutes a correct match for the relevant indicator. In order to perform feature engineering on the data, we extracted a sample of pairs. When considering the Cartesian product of the 242 indicative phrases and all museum website sentences, we expect the vast majority of pairs to be non-matches, making the class distribution between matches and non-matches unbalanced. Hence, to re-balance the sample, we had to identify a method to over-represent the proportion of matches against the extremely common non-matches. To rapidly identify the part of the dataset with a high number of matches, we considered several approaches, selecting word overlap as the most effective, calculated as a BOW (i.e., ignoring the order). We calculated the number of shared lemmas, counting how many lemmatised words co-occur in both indicative phrase (P) and website sentence (W). We then obtained the word overlap as $\hat{P} = |P \cap W|/|P|$, ranging from 0 (no overlap) to 1 (identical text).

Using word overlap as a crude proxy for the likelihood of a match, we stratified a sample of 1,000 pairs, balancing it between low and high overlap. Changing thresholds for \hat{P} over two iterations, upon manual inspection, the sample contained approximately 35% true and 65% false matches, over-representing true matches to allow us to extract effective features. By observing the matches in the sample, we selected a number of features that capture the likelihood of a positive match, ranging from simple word counts to more complex ones. The complete set of features that we produced for the activity indicator matching model is displayed in Table 4. The core variables are derived either from the overlap between exact tokens (e.g., “exhibition” and “exhibitions” would not match) or that between lemmas (so “exhibition” and “exhibitions” would match). The reason to keep both, even though highly correlated, is that an exact match at the token level can be considered semantically slightly stronger than an overlapping lemma.

To add a more gradual, nuanced dimension to the comparison, we included a semantic similarity measure ranging from 0 (no similarity) to 1 (identical content) [5]. Unlike simple lexical matching, this approach takes into account semantically similar words (e.g., “donation”, “grant”, and “endowment”), capturing their presence in the pairs. This measure relies on a pre-trained linguistic model and is calculated between the full input phrase and website sentence as cosine distance between word embeddings generated through the GloVe tool [38]. Moreover, after several iterations, to reduce the matching space, we observed that matches without overlap between semantically important content words were extremely unlikely. Hence, we annotated all indicative phrases with a set of *critical words*, defined as the most important content words. For example, the phrase “in line with government restrictions remain closed for now” has “government restrictions” and “closed” as critical words.

The set of features in Table 4 was calculated for all pairs. After a manual inspection of the results, we noted that COVID-related communications appeared to be located on the websites’ home pages and not on secondary pages. For this reason, we only included home pages in the analysis, as opposed to all scraped pages that were initially selected, giving an approximate total of 1.7 billion pairs (242 phrases \times \sim 80 sentences \times 3,295 museum web pages for each temporal snapshot). All variables were scaled using a min-max scaler, making them comparable on a 0-1 scale. Given the significant processing power required for such a calculation, we reduced the combinatorial space by excluding pairs with a semantic similarity lower than a threshold of 0.45, which consistently corresponded with a non-match. The resulting dataset comprises an average of about 5.2 million pairs for all websites per temporal snapshot, without any information loss.

3.5.3 Model comparison and selection. To support the selection of the most appropriate model for indicator matching, we built a dataset for the purpose of training and testing the ML models. Given the diversity of pairs and the project’s resource constraints, we considered 700 pairs as a reasonable sample size, stratified following the strategy described above to over-represent positive cases to provide the models with enough examples of actual matches. Two members of the project team manually independently annotated this dataset, resolving

¹⁵These steps were performed with the *spaCy 2.3* package with language model *en_core_web_lg*.

Table 4. Features in the activity indicator matching model. This set of variables captures the matching process between the indicative phrases and the sentences from museum websites. High values of overlap and similarity suggest a possible positive match. Some variables are integers in the range $[0, n]$, while others are normalised between 0 and 1. The categorical variable is represented as a set of dummy variables. Highly informative variables are shown in boldface.

Variable	Type
Indicative phrase length	n
Indicative phrase indicator (e.g., “closed currently”)	Categorical
Museum sentence length	n
Overlapping lemmas	n
Overlapping lemmas (with duplicates)	n
Overlapping tokens	n
Overlapping tokens (with duplicates)	n
Overlapping critical words	n
Overlapping critical words (with duplicates)	n
Indicative phrase overlap, lemma (overlap / length of phrase)	$[0, 1]$
Indicative phrase overlap, token (overlap / length of phrase)	$[0, 1]$
Museum sentence overlap, lemma (overlap / length of sentence)	$[0, 1]$
Museum sentence overlap, token (overlap / length of sentence)	$[0, 1]$
Semantic similarity (cosine distance between word embeddings)	$[0, 1]$

ambiguous cases through subsequent discussions involving a third team member. The stratified random sample turned out to contain 33% positive and 67% negative matches, an appropriate proportion to reduce the expected class unbalance in this kind of task.

These 700 cases provided the platform for the exploration and comparison of different models. The performance of these models was measured with metrics including accuracy, precision, recall and F-score, and with the confusion matrix when appropriate. As a baseline, we tried to systematically apply simple thresholds on the features, for example identifying a rule that all pairs with token overlap greater than .8 are valid. As predictable, this approach obtained limited success, with an average F-score of 0.56 and an accuracy of 0.62 (coupled with a very low precision of 0.45). This confirmed the need for more sophisticated approaches.

The subsequent selection process hinged on the systematic comparison of a number of alternate models. The models were selected among well-known methods that are suitable for a binary classification of multi-variate data¹⁶. Ordering the classifiers from worst to best performance, we considered K-nearest neighbours, decision trees, support vector machines, random forests, multi-layer perceptrons, logistic regression, Gaussian process regression, and convolutional neural networks [2]. The only classifier that did not outperform the baseline method was the Naive Bayes classifier. Table 5 provides a summary of the performance of the different models.

The results of this binary classification problem can be evaluated through a confusion matrix, comparing predicted matches and actual matches as identified by human annotators. We calculated true positives (TP, correct matches between indicative phrases and website text), true negatives (TN, non-matches), false positives (FP, matches that were found, but are incorrect), and false negatives (FN, correct matches that were not found). From these values, it is possible to derive for each model *accuracy* as $(TP+TN)/(TP+FP+FN+TN)$, *precision* as $TP/(TP+FP)$, *recall* as $TP/(TP+FN)$, and F-score as a combination of precision and recall. These three metrics capture complementary aspects of a model's performance, and we aimed at maximising precision and F-score. Given the nature of our data, we favoured models that produced a relatively low proportion of FP compared to FN.

¹⁶The machine learning models were generated and tested with Python package *scikit-learn* 0.24. The deep learning models were built with packages *keras* and *tensorflow* 2.6.

Table 5. Model comparison and selection for the matching process between indicative phrases and website text. The 700 manually annotated cases were used as training and test set. Each stochastic model was run 10 times with a different random seed. The rows are sorted by mean accuracy in descending order. All metrics (precision, recall, F-score, accuracy, false positive, and false negatives) are averaged over four different training/test splits (20%, 30%, 40%, and 50%) and 10 trials (40 cases). Classifiers with lower performance are omitted. (*) indicates the top-performing models in terms of average F-score and accuracy.

Classifier	Mean Prec.	Mean Rec.	Mn±SD F-score	Mn±SD Accur.	False pos. %	False neg. %
Convolutional Neural Network	.76	.59	.66 ± .03*	.80 ± .02*	7.1	14.1
Gaussian Process Regression	.72	.57	.64 ± .02	.78 ± .01	7.5	14.2
Logistic Regression	.75	.51	.61 ± .02	.78 ± .01	5.5	16.2
Multi-layer Perceptron	.74	.50	.59 ± .03	.77 ± .01	6.0	16.7
Random Forests	.72	.52	.60 ± .02	.77 ± .01	7.1	15.9
Support Vector Machine	.68	.58	.62 ± .01	.77 ± .01	9.1	14.0
Decision Trees	.61	.54	.57 ± .04	.73 ± .01	11.7	15.5
K-Nearest Neighbours	.61	.58	.59 ± .03	.73 ± .02	12.4	13.7
Manual thresholds (<i>baseline</i>)	.45	.71	.56 ± .07	.62 ± .03	28.1	11.6

We considered the simple threshold model as a baseline (the last row in the Table 5), discarding all models that did not outperform it in terms of F-score and accuracy. To observe the stability of results, all models were tested over four different training/test random splits (20%, 30%, 40%, and 50%), obtaining a realistic indicator of the models' ability to classify matches correctly and stably. Each model was run ten times, for a total of 40 different trials. For the sake of brevity, models exhibiting a worse performance are omitted from the table. To provide a comprehensive comparison, we tuned the hyperparameters for each model, selecting a range of values and observing their effects. Each model was tested first with a minimal set of four variables that appeared highly informative, and then with all available variables, excluding highly-correlated ones (Spearman's $\rho > .7$) (see Table 4). Empirically, all variables provided a contribution—even if minor—to the model's performance, and therefore were included in all models.

For example, random forests are sensitive to the number of trees, the maximum depth of the trees, and whether bootstrap is used for the training or not [41]. In our analysis, we varied these hyperparameters, observing that the maximum depth had a visible and yet modest impact on the results: deeper trees (max depth 20) obtained a slightly higher F-score (0.61), but the results were overall very stable, and the best ones were obtained with 100 trees and a maximum depth of 5. The random forest models with a lower performance were discarded and not included in the table.

3.5.4 Selection of CNN. Convolutional neural networks (CNN) emerged as the top-performing model. This result is unsurprising, given the wide range of problems in which deep learning outperforms other methods [16]. In order to tune the network, we defined a structure comprising a first layer with a Rectified Linear Unit activation function (ReLU), followed by a varying number ReLU layers, closing the network with a sigmoid activation function. As shown in Table 6, we varied the number of layers and their neurons, increasing the depth and complexity of the network.

As a diagnostic tool, we used binary cross-entropy as the loss function, which is suitable for binary classification problems — the lower the loss value, the better the model fit. For the training process, we adopted the Adam optimisation algorithm, a well-known method to learn network weights iteratively from the training data, which outperforms traditional stochastic gradient descent [22]. Other optimisation methods, such as gradient descent

Table 6. Comparison and selection of convolutional neural networks (CNN) for the matching process between indicative phrases and website text, using the manually annotated 700 cases as training and test set. For example, for layer neurons 8, 16, the network includes a first ReLU layer, followed by a layer with 8 neurons, another layer with 16 neurons and the final layer with a sigmoid activation function. The loss function is binary cross-entropy. (*) indicates the top-performing models in terms of average F-score and accuracy.

Layer neurons	Batch size	Mean prec.	Mean rec.	Mean F-sc.	Mean acc.	False pos.	False neg.	Loss
8*	8	.75	.57	.64	.79	.06	.14	.49
	32*	.75	.58	.66*	.8*	.07	.14	.47*
8, 16	8	.71	.59	.64	.79	.08	.13	.58
	32	.7	.61	.65	.79	.08	.12	.48
8, 16, 32	8	.68	.61	.64	.78	.09	.13	.83
	32	.71	.61	.65	.79	.08	.13	.59
32, 32	8	.67	.61	.63	.77	.1	.13	.92
	32	.72	.6	.65	.79	.08	.13	.65
256, 512	8	.64	.6	.62	.75	.11	.13	2.32
	32	.63	.6	.61	.76	.11	.13	1.64

with momentum (SGD) and Adamax, did not provide significantly better results. We varied the batch size (i.e. the number of training examples in one forward-backward pass) between 8 and 32, avoiding higher values that typically are not recommended with a small training set such as ours. We adopted binary accuracy as our performance metric, calculating how often predictions match manual annotations.

Interestingly, the best performance was obtained with a single layer with 8 neurons, with a batch size of 32 (Table 6). Smaller batch sizes consistently performed less well, indicating that a larger set of training examples was beneficial. Increasing the number and size of layers did not yield improvements to the performance, suggesting that the scope and size of the training data were better handled by a single, relatively small layer. This CNN was selected as our matching model, using an early stopping strategy over 100 epochs to select the model with the best accuracy. As shown in Table 5, the mean accuracy of this model, across the different training/test splits, is $.8 \pm .02$, with an F-score of $.66 \pm .03$, and a relatively low rate of false positives.

It is important to note that this CNN obtained a better performance compared to Gaussian process regression and logistic regression, but only by a small margin. This suggests that the complex nature of the task results in a limited information problem, where even state-of-the-art methods cannot obtain higher accuracy, precision, and recall. In the context of our study, we consider these results satisfactory, considering the intrinsic limitations of the BOW model, the large number of categories to identify, and the limited training data available. Using this CNN, we proceeded to classify all matches between indicative phrases and the data from 3,295 museum websites, considering ten temporal snapshots (see workflow in Figure 3). In each snapshot, between 81,000 and 87,000 matches were classified as valid and analysed in the next step.

3.5.5 Validation and merge of indicator categories. The research team's museum studies experts analysed the results of the indicator-matching process. A relatively high inter-indicator variability emerged, with some indicators obtaining very few matches, because of the limited training data available for some indicators (see Table 3). To explore the heterogeneity of the results in order to validate them, we re-calculated the accuracy, precision and recall by indicator, complementing Table 5.

The detailed results in Table 7 confirmed the high variability between the different indicators, some of which had extremely high accuracy on a high number of cases, while others were almost absent from the dataset. In

Table 7. Validation of indicators found in the 3,295 museum websites, including the percentage of matches in the whole museum website corpus and training examples. Precision, recall, F-score, and accuracy were calculated for each indicator, showing how many training cases were available. For two indicators (*language of difficulty* and *staff working*), an extra validation with a new sample with 100 cases was carried out.

Theme	Indicator	Matches	Train.	Prec.	Rec.	F-sc.	Acc.	Action
		%	exam.					
closure	closed currently	13.1	83	0.89	0.71	0.79	0.93	Valid
	closed indefinitely	4	22	0.89	0.89	0.89	0.29	Merge w/ closed currently
	closed permanently	1.3	4	0	0	0	–	Remove
finances	finance health	0.9	7	1	1	1	–	Remove
	did not get funding	2.9	21	0	0	0	–	Remove
	fundraising	13.9	74	0.91	0.36	0.51	0.67	Valid
	gov. emerg. funding	5.5	19	1	0.5	0.67	0.71	Merge w/ fundraising
	other emerg. funding	0.9	2	0	0	0	–	Remove
online activity	online engagement	3.1	28	0.74	0.94	0.83	0.89	Valid
	online event	4.2	62	0.5	0.11	0.18	1	Merge w/ online engagement
	online exhibition	3.9	64	1	0.08	0.15	1	Merge w/ online engagement
open/reopen	open cafe	0.7	10	0	0	0	–	Remove
	open currently	4.5	17	0.72	0.71	0.71	0.8	Valid
	online shop open	5.5	32	0.86	0.75	0.8	0	Remove
	reopen intent	17.2	106	0.8	0.77	0.78	0.75	Valid
	reopen plan	1.6	16	0.88	0.78	0.82	0.69	Merge w/ reopen intent
staffing	hiring staff	1.8	15	0	0	0	–	Remove
	restructuring staff	0	–	–	–	–	–	Remove
	staff working	5.3	46	0.57	0.72	0.63	0.67	–
	[extra validation]	0	[100]	0.73	0.32	0.45	0.75	Valid
misc.	language of difficulty	7.8	67	1	0.13	0.22	–	–
	[extra validation]	0	[100]	1	0.1	0.19	0.67	Remove
	made Covid safe	1.3	5	0	0	0	–	Remove
	project postponed	0.7	–	–	–	–	0	Remove
<i>Total</i>	100	700	–	–	–	–	–	

extensive discussions involving the project’s technical team and its museum studies experts, the whole team discussed each indicator’s reliability and semantic interpretability, opting for one of three actions: keep the indicator as valid, drop it, or merge it with another. As can be observed in Table 7, some indicators were removed because they were only marginally present in the training dataset and in the websites (e.g. closed permanently, made COVID safe, project postponed). Other indicators were insufficiently distinguishable semantically from one another and were therefore merged into a single indicator. Two indicators (*language of difficulty* and *staff working*) had insufficient training and test data but appeared semantically crisp. The team carried out an extra validation, collecting a sample of 100 cases for each of these two indicators, which provided conclusive evidence of whether to exclude or retain them. Favouring precision over recall, this process finally produced six indicators

with overall high accuracy and F-score: *closed currently, funding, online engagement, open currently, reopen intent, and staff working*.

3.5.6 Matching on social media corpus. As the identification of indicators using the set of indicative phrases produced satisfactory results for museum websites, we moved on to the detection of indicator occurrences within the social media data. The advantages of this dataset, compared to the website corpus, lie in the finer-grained temporal demarcation of messages, unlike website indicators that have a less clear timeline. Two datasets, containing respectively 3.4 million tweets and 1.2 million Facebook posts (see Section 3.4), were harmonised into a single corpus containing social media messages. Based on the workflow for websites shown in Figure 3, we pre-processed this dataset by tokenising, lemmatising, and POS-tagging the 4.6 million messages.

Adopting the same strategy to filter out obvious non-matches that we had applied for website data, we obtained about 28.9 million pairs between indicative phrases and social media messages. Subsequently, we applied our method based on a BOW model and a CNN for pair matching to the corpus. Of 28.9 million pairs, our method classified 1,380,340 of them as valid matches for 1,176 museums. 52% of these matches were found in Facebook posts and 48% in the Twitter data.

We note that our focus on text constitutes an important limitation in this indicator-matching process, as this social media content includes a high proportion of visual content, such as captioned photographs and videos. This content may be semantically relevant for our project's research aims, but was left unexamined in our analysis and presents an opportunity for further analysis in the future.

We present visualisations of the results arising from the indicator matching process in Section 5 below. Extending the work of the MiP project, the techniques described in Section 3.2 could form the basis for discovering the official websites of other groups of institutions, for example museums beyond the UK, or other types of institutions. The techniques of Section 3.3 can be extended to discover also museums' TikTok, Instagram and other social media accounts. The techniques of Section 3.5 can be used to gather linguistic data and analyse other aspects of museums' online content.

4 IMPLEMENTATION

The analysis detailed in the previous section required the design of a complex software architecture to identify, collect, organise and process museum data at scale. Drawing on our prior experience on the Mapping Museums project in developing tools to support the research of humanities scholars [40], we designed a modular architecture to organise and rapidly evolve the inter-connected modules. From an infrastructural perspective, we used a Linux multi-core virtual machine managed by the Computer Science department at Birkbeck, University of London. We developed all code within an Anaconda environment with Python 3.9 on a GitHub repository¹⁷. For data storage, we deposited small datasets in GitHub as CSV, TSV, or pickle format, whilst large and dynamic datasets were stored in a PostgreSQL database in a centralised DBMS, accessed from the Python code. Additionally, the research team used an MS Teams folder to share files between the technical team and the museum studies experts without having to rely on GitHub. We structured our Python code into scripts for the scraping module and data pre-processing, and into Jupyter notebooks for the analyses, importing the same modules in both scripts and notebooks to reduce redundancy and facilitate code maintenance.

Our Python code follows open-source good practice, and relies on a set of packages, currently all available on Anaconda and versioned in our Conda environment for reproducibility. The main packages are *pandas* for data processing and analytics, *scrapy* for Web scraping, *requests* to retrieve data from social media APIs, *beautifulsoup4* for parsing and information extraction from HTML, *multiprocessing* for parallel computing, *fuzzywuzzy* for string comparison and similarity metrics, *sklearn* for ML models, excluding deep learning CNN that were implemented with *keras*, and *seaborn* for data visualisation.

¹⁷<https://github.com/Birkbeck/museums-in-the-pandemic>, accessed in August 2022.

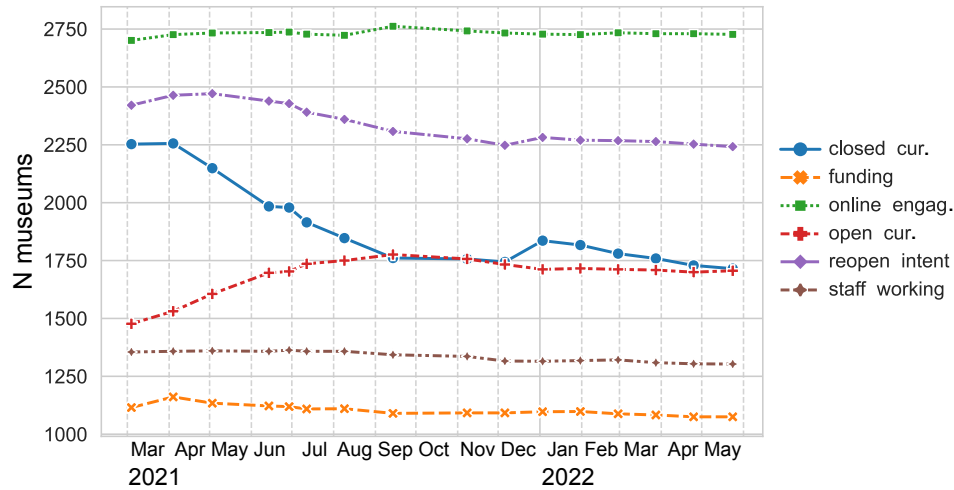


Fig. 4. Indicators in museum websites from March 2021 to May 2022 detected through our BOW and CNN method. The museum website corpus includes 3,295 websites. The y axis shows the number of museums whose websites contain a reference to an indicator. The slow rate of change reflects the relatively static nature of the museum website content.

5 RESULTS

Based on the six activity indicators identified in Section 3.5.5, a set of statistics and visualisations have been produced for use by the museum studies experts, collaboratively refined through several iterations. Figures 4-7 illustrate a sample of these visualisations.

Figure 4 presents the trends detected in the occurrence of activity indicators within the full set of 3,295 museum websites during the period March 2021-May 2022. We can see the fall and then stabilisation of the ‘closed currently’ indicator (falling from approximately 75% to 60% of websites), the commensurate rise and then stabilisation of the ‘open currently’ indicator (rising from approximately 50% to 60%), and the mild rise and fall of the ‘reopen intent’ indicator (varying between 80% to 75% of websites). The other three indicators appear relatively static over the period: ‘online engagement’ occurs in about 90% of websites while ‘staff working’ and ‘funding’ occur at a lower prevalence of about 45% and 40%, respectively.

A range of similar but more detailed statistics and visualisations have been produced, breaking down the number of indicator occurrences according to attributes such as museum governance, size, subject matter, accreditation status and region. For example, Figure 5 shows the prevalence of indicator occurrences for the three largest categories of museums according to their governance status, namely the Local Authority museums (a subcategory of Government museums) and the Not-for-profit and Private museums (two subcategories of Independent museums), thereby allowing the museum studies experts to compare the trends side by side. We can observe again the fall in ‘closed currently’ and rise in ‘open currently’, for all three categories; the highest levels of ‘open currently’, ‘reopen intent’, ‘online engagement’ and ‘staff working’ in the local authority museums; the highest levels of ‘funding’ in the not-for-profit museums; and the lowest levels of ‘online engagement’, ‘staff working’ and ‘funding’ in the private museums.

Figure 6 illustrates a further breakdown of indicator occurrences for the above three governance types according also to three categories of museum size: Large, Medium, or Small (resulting in 9 groups of museums in total). From the topmost three groups we can observe the highest levels of ‘open currently’, ‘online engagement’ and ‘funding’

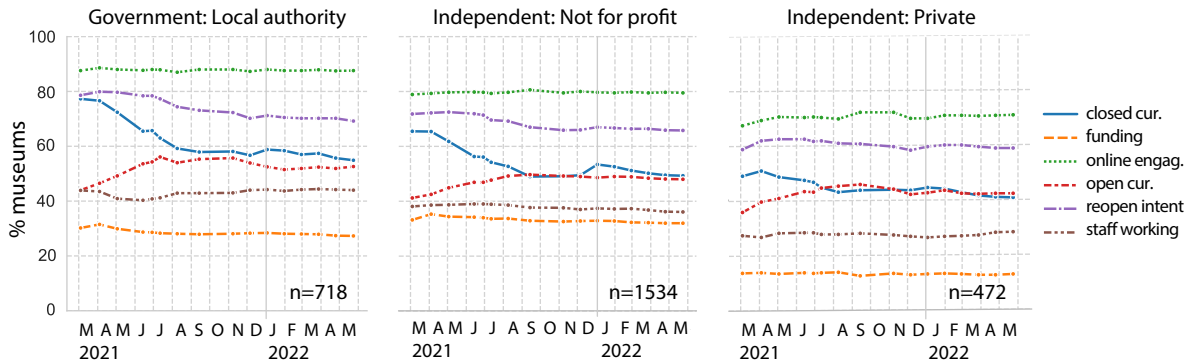


Fig. 5. Indicators occurring in three different categories of governance. The y axis shows the percentage of museums whose websites contain a reference to an indicator. The total number of museums in each category is also displayed.

in the large local authority museums, with similar trends in the middle three groups (large, medium, small not-for-profit museums) and bottom three groups (large, medium, small private museums). From the leftmost three groups, we can observe highest levels of 'online engagement' and 'staff working' in the large local authority museums; highest levels of 'funding' in the large not-for-profit museums; and highest levels of 'reopen intent' in the large private museums.

Analogous statistics and visualisations have been produced for occurrences of the six indicators within the social media data, over the longer period of data collection of the Twitter and Facebook data from January 2019 to May 2022. It is important to note that, while our data include engagement metrics, this analysis and the associated visualisations only focus on content production, rather than user engagement, which will be addressed in a future dedicated study. For example, Figure 7 shows the total number of indicators over the period occurring in the Twitter data (upper graph) and Facebook data (lower graph). Because of the finer-grained collection of the social media data, allowing each occurrence to be timestamped, it is possible to temporally aggregate the indicator occurrences within this data at multiple levels, e.g. daily, weekly, monthly. In the line graphs shown, the temporal aggregation is at the weekly level. We can observe the more extreme fluctuations in these graphs compared to Figure 4 – because museums' websites were only scraped every two weeks and also because museums' social media posts are much more frequent compared to updates of their websites. We can see the levels of indicator activity before the start of the pandemic (the first UK lockdown occurred in March 2020), the peaks in 'closed currently' and 'reopen intent' in response to government announcements of increased/decreased restrictions, and the higher levels of 'online engagement' since March 2020. We also see in the Facebook data (lower graph) rises in references to 'staff working' and 'funding', and fluctuations in references to 'open currently'. Similar but less marked trends are seen in the Twitter data (upper graph).

Our activity indicators extracted from museums' online content resonate with the themes of museums' increased online activities, impact on staff, economic impact, and concerns about permanent closures emerging from the studies reviewed in Section 1.1. In addition to the indicator occurrences in museums' websites, we have generated more detailed statistics and visualisations, breaking down the number of indicator occurrences in the social media data according to attributes such as museum governance, size, subject matter, accreditation status and region to support the research of the museum studies experts. Our methodology has successfully produced valuable data and usable visualisations to analyse the rapid changes that occurred during the COVID-19 pandemic. The project's museum studies experts have analysed the online content production of UK museums, showing how organisations have engaged in online communications in the different phases of the pandemic [25] and how

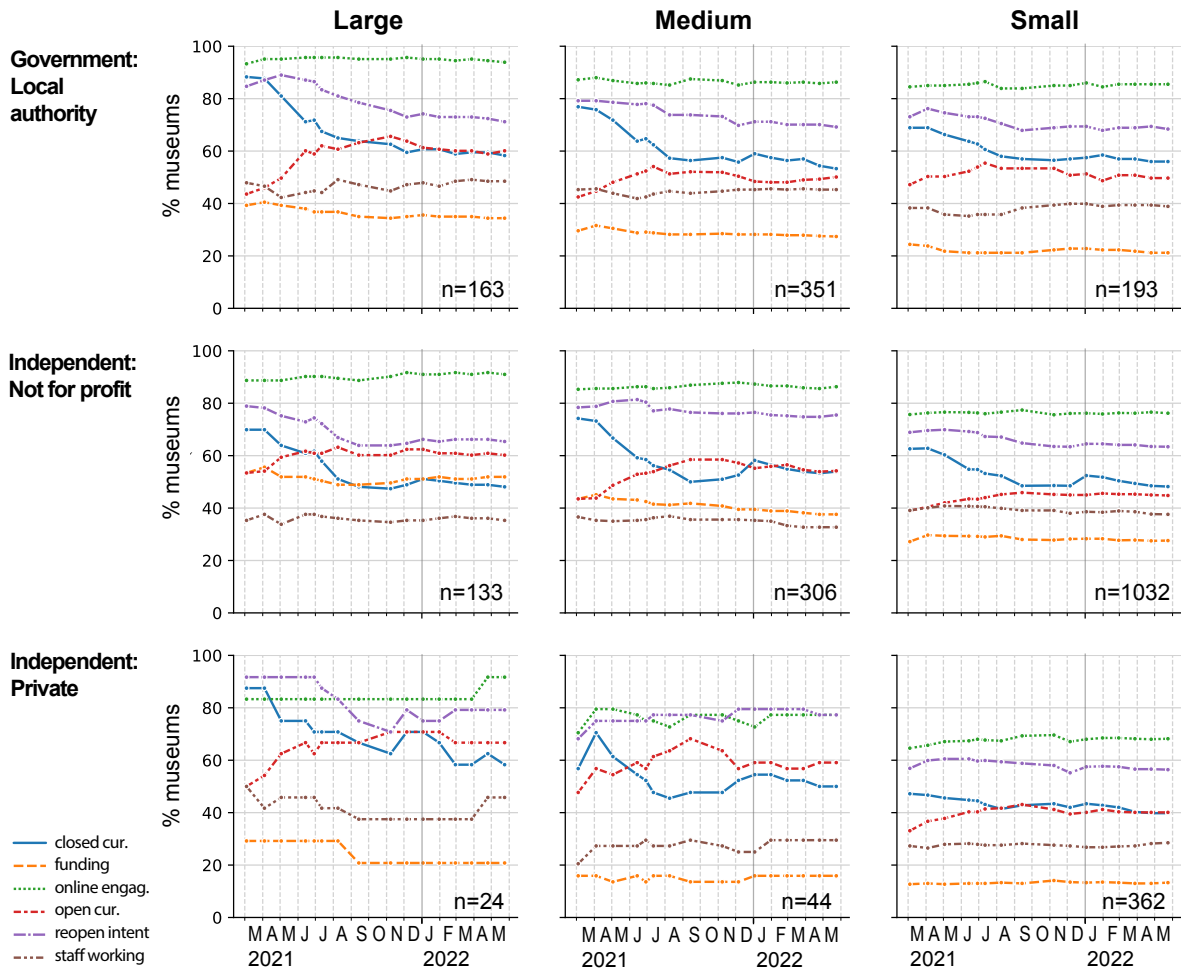


Fig. 6. Indicators in three different categories of governance (y axis) and size (x axis). The y axis in each chart shows the percentage of museums whose websites contain a reference to an indicator. The total number of museums in each group, n, is also displayed.

restrictions have shaped closure patterns [26], countering received narratives about a ‘digital pivot’ and about museums in distress.

Beyond this project, the visualisations described here could be extended to visualise activity indicators detected in additional social media outlets (e.g., TikTok, Instagram). More generally they could be repurposed to visualise other sets of activity indicators, arising from linguistic analysis of other aspects of museums’ online content.

6 THE MIP SEARCH APP

The indicators detected in the museums’ websites and social media messages are critical to quantifying museums’ online behaviour in response to the COVID-19 pandemic, according to the project’s objectives. However, while

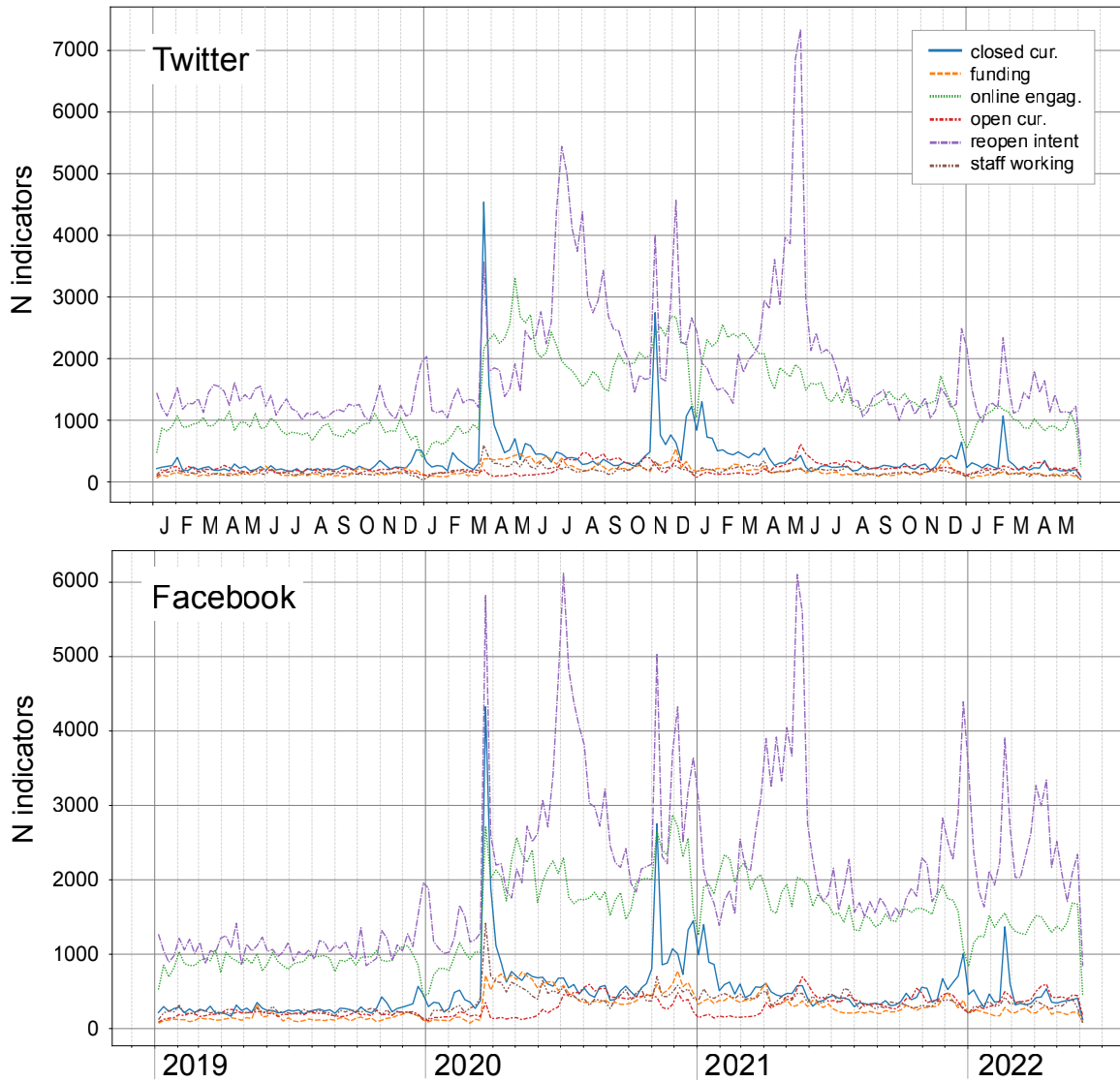


Fig. 7. Indicator occurrences in the UK museum social media data from January 2019 to May 2022. The y axis in each chart shows the number of indicator references found. The number of occurrences is aggregated temporally at the weekly level. The datasets include 3.9 million tweets by and to museum accounts, and 1.5 million posts on Facebook pages of museums.

visualising and analysing the results, as described in Section 5, the research team identified the need also for a tool for the museum studies experts to perform *ad hoc* lexical searches within the website and social media corpora. In particular, being able to search for lexical phrases within these corpora using regular expression

The screenshot shows a Jupyter notebook interface for the MIP Search App. The notebook title is 'mip_search_app-v1.ipynb'. The search form contains the following parameters:

- `search_string`: "online exhibition"
- `search_string_not`: "Insert text here"
- `case_sensitive`:
- `search_website_sentences`:
- `search_facebook`:
- `search_twitter`:
- `search_twitter_include_replies`:
- `museum_name`: "Insert text here"
- `museum_governance`: "Insert text here"
- `museum_size`: "large"
- `begin_date`: 2019 / 1 / 1
- `end_date`: 2021 / 12 / 31
- `max_results`: 100
- `context_window`: 20

Fig. 8. Search interface of the MIP Search App, searching for “online exhibition” in large museums. This tool supports search for lexical patterns in the museum websites and social media corpora, with a variety of parameters. The app is hosted on Google Colab and developed as a single Jupyter notebook.

patterns would enable them to investigate “narrow” case studies, complementing and corroborating the “broad” results arising from the indicator analysis.

Hence, we developed also the *MIP Search App*, using Google Colab as the hosting infrastructure¹⁸. The tool is structured as an interactive Jupyter notebook comprising a search cell, a cell for presenting the search results, and a cell presenting a variety of analyses of the results. The search and analysis results can be exported as CSV files for further processing.

Part of the search interface is shown in Figure 8. The search input parameters comprise a search string, a negated string (i.e. a search pattern to exclude from the results), museum name, museum governance, museum size, start and end dates of the data to be searched, and the size of context window around the matched pattern (i.e. the number of words before and after the match to be included in the analysis). A results cell (not shown in Figure 8) displays the matches found in the corpora. For example, Table 8 lists 9 of the results arising from the search of Figure 8, showing a sample of results from Twitter, Facebook, and museum websites.

¹⁸<https://colab.research.google.com>, accessed in July 2022.

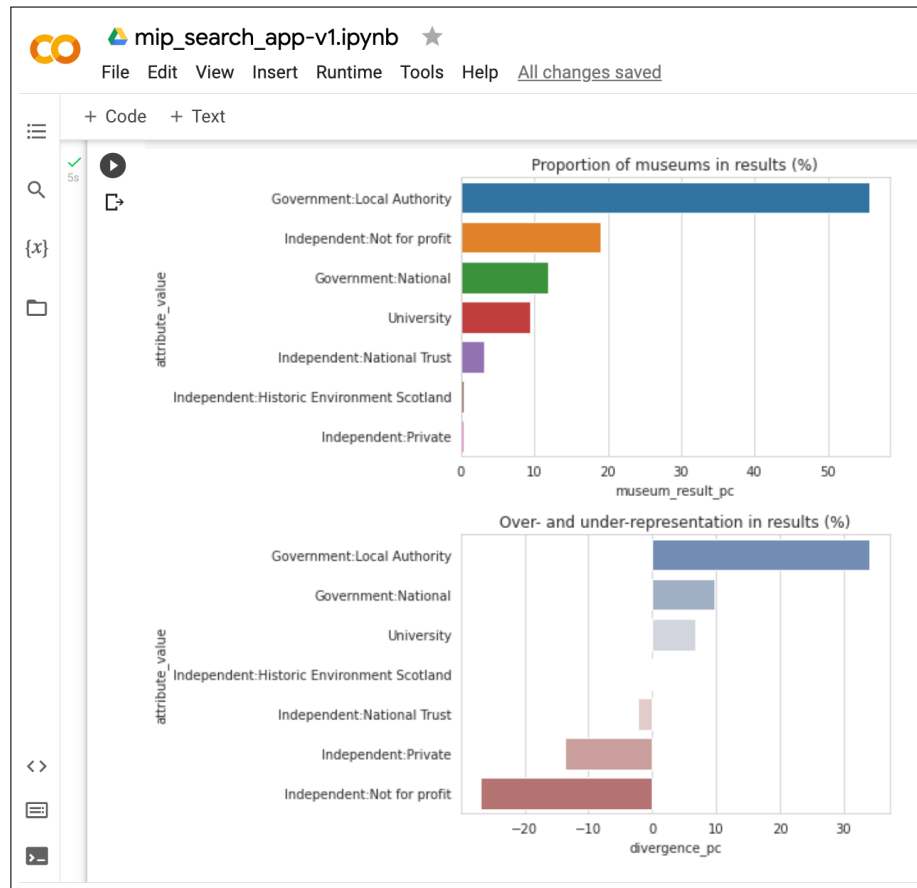


Fig. 9. Analytics interface of the MIP Search App (fragment), showing search results for “online exhibition” in large museums, grouped by museum governance. Among other statistics, the tool shows the composition of the search results by attributes, comparing the absolute counts of results to their relative over- and under-representation to other museum groups.

Working collaboratively with the museum studies experts, we also co-designed functionality to analyse the search results and facilitate insights. In particular, an analytics cell reports on a number of statistics, including number of results in each of Twitter, Facebook and websites, and the number of distinct museums over which they occur; analysis of co-occurring words within the context window before and after the search string (distinct words and number of occurrences of each); temporal breakdown of the number of results, in each of Twitter, Facebook and websites; breakdown of results according to key museum attributes such as governance, size, subject matter and region. Additionally, the interface displays the over- and under-representation of different museum groups in the results. For example, 19% of large museums using the term “online exhibition” are independent not-for-profit organisations. As 26% of museums belong to this category of governance, the tool shows it as under-represented (see Figure 9).

Museum ID	Lexical match of “online exhibition”
<i>Twitter</i>	
mm.domus.SC244	Scotland’s best-known locations from the comfort of your sofa? Then check out our new online exhibition ‘Old Ways New Roads’! Following in the footsteps of 18th-century travellers ...
mm.ace.685	The Hepworth Wakefield is closed on Mondays and Tuesdays however you can visit our online exhibition with Google Arts & Culture. Zoom into Hepworth’s sculptures - photographed in ultra-high resolution using the latest Google technology.
mm.domus.WM013	To mark the end of South Asian Heritage Month we have curated a new online exhibition , featuring all the items we have showcased over recent weeks. To view the show ...
<i>Facebook</i>	
mm.aim.0057	Happy International Women’s day! Why not peruse our online exhibition about all the fantastic women that have featured and worked on designing banknotes over the years #IWD202
mm.domus.WM038	Discover an online exhibition for #BlackHistoryMonth - ‘Pilots of the Caribbean: Volunteers of African Heritage in the Royal Air Force’
mm.domus.SE393	Join us this Friday to launch Museum of Colour with its first online exhibition ‘People of Letters’. Ten leading writers & composer/musician @randolphinfo offer their personal take on a chosen object ...
<i>Websites</i>	
mm.aim.0057	Explore our recent online exhibition about the new £50 note featuring Alan Turing ...
mm.domus.SE379	Visit the online exhibition “Lockdown creativity” ...
mm.domus.NE038	visit us at Kirkleatham Museum, we have worked with local artists to create our first online exhibition ‘Living Through Lockdown’ which captures the stillness that came from lockdown and the rediscovery of the importance of nature ...

Table 8. A sample of results for the search “online exhibition” on the MiP Search Tool, showing the match and the text coming before and after for the three corpora (tweets, Facebook posts, and websites).

Beyond the work of the MiP project, the MiP Search App could easily be extended to handle text arising from additional museum-related data sources and corpora. This could include automatically-generated text transcripts of audio files or textual descriptions of images and videos.

7 CONCLUSIONS

The MiP project is seeking to provide timely data on how the UK’s museum sector is responding to the COVID-19 pandemic in order to allow research into which museums close, which remain resilient, and how the profile of the UK museum sector may be changing. This paper has motivated, presented and evaluated methods for extracting data from UK museums’ websites and social media posts in order to support this research. The technical contributions of the paper include:

- development of a machine learning model for discovering museums’ websites;
- development of a hybrid ML and information retrieval method for discovering museums’ Twitter and Facebook accounts;
- creation of a text corpus comprising content from museums websites covering the period March 2021-May 2022, and Twitter and Facebook posts covering the period January 2019-May 2022;
- identification of a set of activity indicators relating to museums’ responses to the COVID-19 pandemic;
- development of a bag-of-words-based linguistic model for discovering occurrences of these indicators within the text corpus;

- design of visualisations to present trends in the number of occurrences of the activity indicators over time, including splitting these numbers according to key attributes of museums such as governance, size, subject matter, accreditation status and location;
- design of an interactive search tool over the text corpus allowing the project's museum studies experts to search for occurrences of specific lexical phrases, with the possibility of focussing their search on a particular time period or on a subset of museums according to museums' key attributes, and presenting a range of statistics to support their analysis of the data.

The MIP project's museum studies experts have used these technical resources and tools to undertake several analyses of the data, making a number of key observations, further articulated in [25, 26]¹⁹:

- Some 550 museums remained completely closed four months after the relaxation of restrictions in the UK for reasons connected to the pandemic. There were also references to the closure of parts of museums' sites (e.g. cafes, galleries), so in addition to museums that were entirely closed, many more were experiencing some degree of closure.
- The analysis of closures and reopenings by museum governance, size, and location showed significant differences across the sector. For example, the data suggests that private museums, unaccredited museums, and smaller museums were more likely to remain closed for longer.
- Accredited museums were more likely to have reopened than unaccredited ones, and to be making fewer references to closure. Larger museums were more likely to be reopening than smaller museums, and again made fewer references to closure. Among the three largest groups of museums according to governance (local authority, private and not-for-profit), local authority museums were more likely to have reopened.
- Fewer museums closed during 2020 and 2021 than in previous years, most likely because of the furlough scheme and other sources of funding²⁰. However, fewer museums opened during 2020 and 2021 than in previous years, although with some upturn in 2021.
- There was an initial surge in social media activity by museums at the start of the pandemic, followed by a relative decline in activity. The contraction in Twitter and Facebook use is contrary to expectations set by other surveys that suggest sustained social media activity during the pandemic [36] – Instagram and TikTok trends might diverge significantly from this.
- Larger museums were better able to maintain social media activity throughout the pandemic than smaller ones, suggesting digital infrastructure and capacity issues across the sector and digital inequalities.
- The use of the UK government's Coronavirus Jobs Retention Scheme, furloughing of staff, and continual closures of some museums may have contributed to decreases in social media activity. Digital fatigue may have also played a role in the decline in social media activity.

This is the first time that this kind of research has been enabled and conducted for the entire museum sector at the national level. The interdisciplinary methodology and computational methods described here could be applied to analysing other aspects of museums' online content, beyond their response to the COVID-19 pandemic. It would also be possible to extend the data collection and research to other geographical contexts, beyond the UK, and other sources of relevant online content (e.g. Instagram, TikTok), and also to cross-reference our data with other museum-related data, e.g. on museum visitors from the Audience Agency²¹.

Our methodology has enabled a range of studies led by the MIP project's museum studies experts, who are performing analyses of the activity indicators and also undertaking an in-depth analysis of the longitudinal social media data for the entire period January 2019-May 2022, comparing pre-, during, and post-pandemic

¹⁹For details, see also the blog posts at <http://blogs.bbk.ac.uk/mapping-museums/2022/05/19/museum-governance-and-reopening-after-lockdown/> and <http://blogs.bbk.ac.uk/mapping-museums/2022/06/14/accreditation-size-and-museum-reopening-after-lockdown/>, accessed in Feb 2023.

²⁰<http://blogs.bbk.ac.uk/mapping-museums/2021/07/01/funding-for-uk-museums-during-the-pandemic/>, accessed in Feb 2023.

²¹<https://www.theaudienceagency.org/>, accessed in Feb 2023.

behaviours [25]. This research effort enlightens how different types of museums responded to the COVID-19 pandemic according to attributes such as governance, size, accreditation status, subject matter and location. As future research avenues, we consider extending our work to include Instagram and TikTok, analysing video and image content. Recently developed NLP methods could be adopted to study the linguistic patterns in our corpora of museum websites, improving our understanding of how museums communicate online and present themselves. Finally, cross-referencing this online data with visitor data would unlock new insights into museum operations in their geographical and cultural context.

DATA AVAILABILITY STATEMENT

The data, software, and materials used to produce this article are available online under a Creative Commons license at <https://github.com/Birkbeck/museums-in-the-pandemic>. Some data, including scraped museum websites and social media messages, cannot be republished as open data for copyright reasons. The input dataset about UK museums used to retrieve online content is available at <https://github.com/Birkbeck/mapping-museums>.

ACKNOWLEDGMENTS

We gratefully thank UKRI-AHRC for funding the project, all members of the MIP project team and the MIP Advisory Board, and all participants in the project's design and validation activities. We also thank the developers of the open-source packages used in the project.

REFERENCES

- [1] The Audience Agency. 2018. *Museums Audience Report*. Technical Report. The Audience Agency, London. 1–16 pages. <https://www.theaudienceagency.org/asset/1995>
- [2] Charu C. Aggarwal. 2018. *Machine Learning for Text*. Springer, Berlin.
- [3] Deborah Agostino, Michela Arnaboldi, and Antonio Lampis. 2020. Italian state museums during the COVID-19 crisis: from onsite closure to online openness. *Museum Management and Curatorship* 35, 4 (2020), 362–372.
- [4] Brooke Auxier and Monica Anderson. 2021. *Social media use in 2021*. Technical Report. Pew Research Center, Washington, DC. 1–18 pages. <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>
- [5] Andrea Ballatore, David C. Wilson, and Michela Bertolotto. 2013. Computing the semantic similarity of geographic terms using volunteered lexical definitions. *International Journal of Geographical Information Science* 27, 10 (2013), 2099–2118.
- [6] Bogdan Batrinca and Philip C Treleaven. 2015. Social media analytics: A survey of techniques, tools and platforms. *AI & Society* 30, 1 (2015), 89–116.
- [7] Abdelhak Belhi, Abdelaziz Bouras, Abdulaziz Khalid Al-Ali, and Abdul Hamid Sadka (Eds.). 2021. *Data Analytics for Cultural Heritage: Current Trends and Concepts*. Springer, Berlin.
- [8] Fernando Borraro-Millán, María-del-Mar Alonso-Almeida, María Escat-Cortes, and Liu Yi. 2021. Sentiment Analysis to Measure Quality and Build Sustainability in Tourism Destinations. *Sustainability* 13, 11 (2021), 6015.
- [9] Leo Breiman. 2001. Random forests. *Machine Learning* 45, 1 (2001), 5–32.
- [10] Verity Burke, Dolly Jørgensen, and Finn Arne Jørgensen. 2020. Museums at home: Digital initiatives in response to COVID-19. *Norsk museumstidsskrift* 6, 2 (2020), 117–123.
- [11] Fiona Candlin, Jamie Larkin, Andrea Ballatore, and Alexandra Poulouvassilis. 2020. *Mapping Museums 1960-2020: A report on the data*. Technical Report. Birkbeck, University of London. <https://eprints.bbk.ac.uk/id/eprint/31702/>.
- [12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [13] Kalyan Dutia and John Stack. 2021. Heritage connector: A machine learning framework for building linked open data from museum collections. *Applied AI Letters* 2, 2 (2021), e23.
- [14] David Gerrard, Martin Sykora, and Thomas Jackson. 2017. Social media analytics in museums: extracting expressions of inspiration. *Museum Management and Curatorship* 32, 3 (2017), 232–250.
- [15] Matthew K Gold (Ed.). 2012. *Debates in the Digital Humanities*. University of Minnesota Press, Minneapolis, MN.
- [16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT press, Cambridge, MA.
- [17] Piotr Gutowski and Zuzanna Klos-Adamkiewicz. 2020. Development of e-service virtual museum tours in Poland during the SARS-CoV-2 pandemic. *Procedia computer science* 176 (2020), 2375–2383.

- [18] James Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, and Daniel Weitzner. 2008. Web science: an interdisciplinary approach to understanding the web. *Commun. ACM* 51, 7 (2008), 60–69.
- [19] Yang Jin and Liang Min. 2021. Public benefits or commercial gains: chinese Museums' online Activities in the Covid-19 age. *Museum International* 73, 3-4 (2021), 32–43.
- [20] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Association for Computational Linguistics, Valencia, Spain, 427–431. <https://aclanthology.org/E17-2068>
- [21] Ellie King, M Paul Smith, Paul F Wilson, and Mark A Williams. 2021. Digital Responses of UK Museum Exhibitions to the COVID-19 Crisis, March–June 2020. *Curator: The Museum Journal* 64, 3 (2021), 487–504.
- [22] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [23] Julie Thompson Klein. 2010. A taxonomy of interdisciplinarity. *The Oxford Handbook of Interdisciplinarity* 15 (2010), 15–30.
- [24] Konstantinos Kyprianos and Panagiota Kontou. 2022. The use of social media in Greek museums during the COVID-19 pandemic. *Museum Management and Curatorship* (2022), 1–22.
- [25] Jamie Larkin, Andrea Ballatore, and Katerina Mityurova. 2023. Museums and social media during the Covid-19 Pandemic: Was there a pivot to digital? *Curator: The Museum Journal* (2023). Forthcoming.
- [26] Mark Liebenrood, Andrea Ballatore, Fiona Candlin, Jamie Larkin, Alexandra Poulouvassilis, and Peter Wood. 2023. Closing Time? Investigating Museum Closure in the UK during the Covid-19 Pandemic. *in preparation for Museum Management and Curatorship* (2023).
- [27] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).
- [28] Rachel Mackay. 2021. An Abrupt and Brutal Audit: An Analysis of the Crisis Management Response of UK Museums and Heritage Attractions to the Covid-19 Pandemic. *Museum International* 73, 3-4 (2021), 8–19.
- [29] Michela Magliacani and Daniela Sorrentino. 2022. Reinterpreting museums' intended experience during the COVID-19 pandemic: insights from Italian University Museums. *Museum Management and Curatorship* 37, 4 (2022), 353–367.
- [30] Mahshid Majd and Reza Safabakhsh. 2017. Impact of machine learning on improvement of user experience in museums. In *2017 Artificial Intelligence and Signal Processing Conference (AISP)*. IEEE, 195–200.
- [31] Lev Manovich. 2020. *Cultural Analytics*. MIT Press, Cambridge, MA.
- [32] Marianna Marzano and Monia Castellini. 2022. Museums and Crisis Management Due to Covid-19. Effects of Pandemic and the Role of Digital and Social Networks Communication. *European Scientific Journal* 18, 19 (2022), 37–57.
- [33] Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. 2021. Deep learning–based text classification: A comprehensive review. *ACM Computing Surveys (CSUR)* 54, 3 (2021), 1–40.
- [34] Lukas Noehrer, Abigail Gilmore, Caroline Jay, and Yo Yehudi. 2021. The impact of COVID-19 on digital data practices in museums and art galleries in the UK and the US. *Humanities and Social Sciences Communications* 8, 1 (2021).
- [35] Network of European Museum Organisations (NEMO). 2020. Survey on the Impact of the COVID-19 Situation on Museums in Europe Findings and Recommendations. https://www.ne-mo.org/fileadmin/Dateien/public/NEMO_documents/NEMO_Corona_Survey_Results_6_4_20.pdf.
- [36] International Council of Museums (ICOM). 2020. Museums, museum professionals and COVID-19. <https://icom.museum/wp-content/uploads/2020/05/Report-Museums-and-COVID-19.pdf>.
- [37] Georgios Papaioannou. 2021. Museum Big Data: Perceptions and Practices. In *Big Data in Education: Pedagogy and Research*. Springer, 201–215.
- [38] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543.
- [39] Andrew Phippen, L Sheppard, and Steven Furnell. 2004. A practical evaluation of Web analytics. *Internet Research* (2004).
- [40] Alexandra Poulouvassilis, Nick Larsson, Fiona Candlin, Jamie Larkin, and Andrea Ballatore. 2020. Creating a Knowledge Base to research the history of UK Museums through Rapid Application Development. *Journal on Computing and Cultural Heritage (JOCCH)* 12, 4 (2020), 1–27.
- [41] Philipp Probst, Marvin N Wright, and Anne-Laure Boulesteix. 2019. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9, 3 (2019), e1301.
- [42] Nicola Raimo, Ivano De Turi, Alessandra Ricciardelli, and Filippo Vitolla. 2021. Digitalization in the cultural industry: evidence from Italian museums. *International Journal of Entrepreneurial Behavior & Research* 28, 8 (2021), 1962–74.
- [43] Noga Raved and Havatzelet Yahel. 2022. Changing Times—A Time for Change: Museums in the COVID-19 Era. *Museum Worlds* 10, 1 (2022), 145–158.
- [44] Brittany Ryder, Tingting Zhang, and Nan Hua. 2021. The social media “magic”: Virtually engaging visitors during COVID-19 temporary closures. *Administrative Sciences* 11, 2 (2021), 53.
- [45] Myrsini Samaroudi, Karina Rodriguez Echavarría, and Lara Perry. 2020. Heritage in lockdown: digital provision of memory institutions in the UK and US of America during the COVID-19 pandemic. *Museum Management and Curatorship* 35, 4 (2020), 337–361.

- [46] Jun Shao, Qinlin Ying, Shujin Shu, Alastair M Morrison, and Elizabeth Booth. 2019. Museum tourism 2.0: experiences and satisfaction with shopping at the National Gallery in London. *Sustainability* 11, 24 (2019), 7108.
- [47] Stefan Stieglitz, Milad Mirbabaie, Björn Ross, and Christoph Neuberger. 2018. Social media analytics: Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management* 39 (2018), 156–168.
- [48] Scientific United Nations Educational and Cultural Organisations (UNESCO). 2020. Museums around the world in the face of COVID-19. <https://unesdoc.unesco.org/ark:/48223/pf0000373530>.
- [49] Elena Villaespesa. 2019. Museum collections and online users: Development of a segmentation model for the metropolitan museum of art. *Visitor Studies* 22, 2 (2019), 233–252.
- [50] Paul Widdop and David Cutts. 2012. Impact of place on museum participation. *Cultural Trends* 21, 1 (2012), 47–66.
- [51] Stefan Wojcik and Adam Hughes. 2019. *Sizing up Twitter users*. Technical Report. Pew Research Center, Washington, DC. 1–23 pages.