

Online User-Generated Contents for Performance Management of Cultural Institutions

Ph.D. Candidate: Paola Riva

Supervisors: Deborah Agostino, Associate Professor; Michela Arnaboldi, Full Professor

Affiliation: Politecnico di Milano, School of Management

Text-type: 4a Research Proposal

Topic area: Public Management and Non-Profit Management

Advices on the following themes would be particularly appreciated during the EURAM 2021 Doctoral

Colloquium:

- Feedback on relevance of the chosen topic;
- Feedback and suggestions on the overall structure of the research proposal;
- Feedback on the coherence of research questions with targeted research problems and on research question formulation;
- Suggestions on research papers, in terms of research questions, methodology, relevance for communities, particularly on papers 2 and 4.

Abstract

Though academics and scholars recognize the potentialities of online User-Generated Contents (UGCs) to support management decisions, a gap exists in literature connecting the topics of performance management (PM), text modelling, online UGCs, especially with reference to cultural institutions. Targeting specifically these organizations, this research explores the role that online UGCs could assume in PM, investigating ways to account for measures of thoughts and needs expressed by online users through online User-Generated Texts (UGTs).

This is pursued through three main steps. First, a critical literature review of mathematical models for machine-readable representations of online text supports the development of a methodological framework to accurately select the most adequate online text embedding for the specific PM task of interest. The result of this analysis demonstrates to the PM community of academics and experts the potential and limitations of adopting specific methodologies for the analysis of online UGTs and specific guidelines for selecting the most adequate criteria for the specific purposes of interest. Second, guided by the aforementioned methodological framework, this research evidences the usability of online UGTs as tools for PM of cultural institutions identifying performance dimensions by means of quantitative statistical analysis of the own words of online users of cultural institutions. This also allows to underline the additional value brought to decision makers automatically exploiting online UGTs, since elaborating these data could enrich the information that the decision maker already acquires through more traditional data and methods, like interviews or surveys.

Finally, to complement the theoretical and quantitative results, an exploratory qualitative analysis allows to understand the cultural managers' perspectives towards the design, use, reporting and trust of online UGCs as PM tools. This step aims at supporting the development of a framework for PM of cultural institutions based also on online UGTs.

From an academic perspective, this work aims at contributing to PM literature, exploring the possibility of including textual expressions of online users' perceptions, grasped through online UGTs, into a system for the management of performances of institutions. The implementation of this PM system to the specific case of cultural institutions is expected to offer an innovative methodological perspective to PM of cultural institutions, accounting for the perspective of large amounts of actual cultural users on the line of a more participatory approach to Culture.

Keywords: online user-generated content; cultural institutions; text modelling; performance management; decision making

1. Introduction

This research aims to understand the role that online UGCs¹ assume for PM of cultural institutions², investigating ways to account for measures of thoughts, perceptions and needs expressed online by cultural users.

Among studies connected to online UGCs in social sciences (), recent contributions in PM literature evidence the need to pursue a new line of research, aimed at rethinking the role and characteristics of performance management within popular culture, online media platforms and online contents diffusion (Begkos and Antonopoulou, 2020; Fisher, 2019; Lassila et al., 2019; Agostino and Sidorova, 2017, 2016; Arnaboldi et al., 2017; Jeacle, 2017b, 2012; Quattrone, 2016; Jeacle and Carter, 2014). Examples in this direction include the analysis of the role of PM practices in specific online platforms, like online marketplaces (e.g., Leoni and Parker, 2019; McDaid et al., 2019; Jeacle, 2017a; Kornberger et al., 2017), movie streaming platforms (e.g., Bialecki et al., 2017), social media (e.g., Begkos and Antonopoulou, 2020; Agostino and Arnaboldi, 2019, 2017, 2016; Agostino, 2018) and online travel platforms (e.g., Jeacle and Carter, 2011; Scott and Orlikowski, 2012).

¹ In this paper, the definition provided in 2007 by the Organization for Economic Co-operation and Development is used, therefore online UGC are data being publicly available online data developed outside the professional practice and reflecting a certain amount of creativity (OECD, 2007).

² The term "cultural institution" refers here to the definition of the European Commission "architecture, archives, libraries and museums, artistic crafts, audiovisual (including film, television, video games and multimedia), tangible and intangible cultural heritage, design (including fashion design), festivals, music, literature, performing arts, books and publishing, radio, and visual arts" (EU Proposal Regulation, 2018, Chapter 1, Article 2.2).

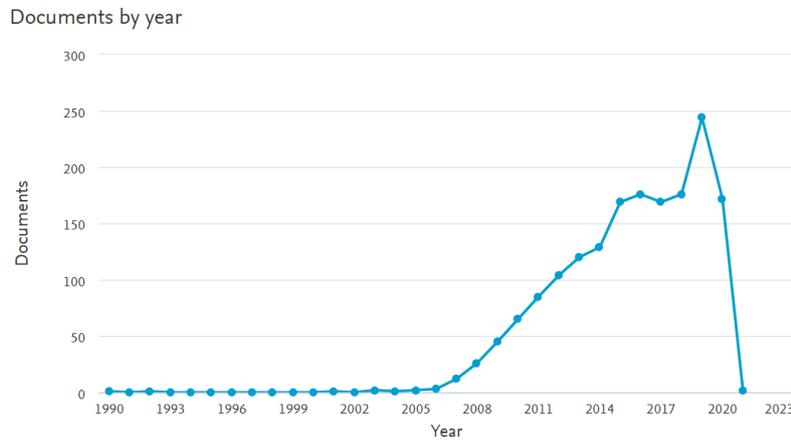


Figure 1 Trend in time of 1705 documents retrieved querying Scopus database for TITLE-ABS-KEY ("user generated" AND online) AND SUBJAREA (arts OR busi OR deci OR econ OR psyc OR soci). Source: Scopus. Last accessed: October 2020.

Although prior researches in PM literature recognize the potential of exploiting information deriving from contents shared online by users (e.g., Lassila et al., 2019; Arnaboldi et al., 2017; Jeacle, 2017b, 2012; Jeacle and Miller, 2016; Eschenbrenner et al., 2015; Jeacle and Carter, 2014), the high subjectivity of online UGCs and their strict connection with user engagement, preferences and experiences (see also Begkos and Antonopoulou, 2020) has until now limited the number of studies relying on the textual opinions of online users for PM. Attempts in this direction have mainly focused on PM of social media, since these platforms open the possibility to measure online users' validation through *counts* of stars, likes, comments and shares to online contents (e.g., Agostino and Arnaboldi, 2017; Oviedo-García et al., 2014; Bonsón and Ratkai, 2013).

Rather than looking merely at *numbers* of users' reactions, making use of online user-generated *texts*, like online messages, social media posts and online reviews, also poses the methodological challenge of modelling human language expressed not only in long documents but also in short text messages, which may contain ambiguous and informal expressions (Alnajran et al., 2019; Fisher, 2019). On this line, this paper aims to understand the role for PM of online user-generated *texts*, voluntary contributions of online users to the online conversation (Peddibhotla and Subramani, 2007), to contribute filling in the gap in literature that connects PM, online data generated by users and text modelling (see also Figure 2).

Cultural institutions have been selected as a relevant setting where to ground this research for at least three main reasons. First, these institutions are nowadays pushed to modify their relationship with the public, moving from an authoritative approach to a more participatory one (Agostino and Arnaboldi, 2020; Bonet and Négrier, 2018). Indeed, these institutions are the more and more recognizing a central role to users (OMC, 2014; Kotler and Kotler, 2000; Falk et al., 1998) and encouraging cultural participation also through online channels, asking online users to interact, engage and creatively contribute to the generation of online contents (ICOM, 2020; ICOM & OECD, 2020; NEMO, 2020; Department for Digital, Culture, Media and Sport, 2019; Agostino and Arnaboldi, 2020; Bonet and Négrier, 2018; Crossick, 2018; Nielsen, 2015; UNESCO, 2012, 2009). Second, the debate on online contents' democratization (Jeacle, 2017a) and on the potentialities of

employing data coming from online users in cultural sector (e.g., Agostino et al., 2020; Mihelj et al., 2019) offers to date just few evidences of the actual usage of online UGCs to inform decisions considering the actual voices of online users (e.g., Marchiori and Cantoni, 2015; Cao and Nguyen, 2012). Third, despite the recent growing interest towards PM in cultural institutions (e.g., Pierotti et al., 2019; Chiaravalloti, 2014; Deidda Gagliardo et al., 2012; Antonelli et al., 2007), experts still evidence the need for additional investigation of these aspects in the cultural field (Agostino and Arnaboldi, 2020; Jeacle, 2017b, 2012; Lapsley and Rekers, 2017; Jeacle and Miller, 2016). This is also connected to the fact that, rather than offering services, cultural institutions offer cultural experiences (Throsby, 2009; Anton et al., 2018) which are valued by users through subjective experiences (Trinh & Ryan, 2017; Falk & Dierking, 2016; Scott, 2004). Considering all these aspects can represent a challenge for the development and implementation of PM systems (Agostino and Arnaboldi, 2020; Jeacle and Miller, 2016) and may require the development of innovative performance metrics (Simpson, 2017; Jacobsen, 2014).

In light of these considerations, this work aims at answering the overarching research question

(RQ) What is the role of online User-Generated Contents for Performance Management of cultural institutions?

articulating it in four sub-research questions (see also **Table 1**), designed to target the knowledge gaps that emerge from a narrative literature review of extant literature connected to PM, text modelling and online data generated by users, with particular reference to the specific context of cultural institutions (see also **Key Literature**).

A knowledge gap emerging from literature is connected to the lack of clear methodological guidelines on how to measure and elaborate online UGTs to support PM tasks. Although recent contributions recognize the potentiality of using online UGTs to inform organizations (e.g., Sagnika et al., 2020; van Dieijen et al., 2020; Tazibt and Aoughlis, 2019; Honari, 2017) and the importance of looking at comments and feedbacks of customers in implementing performance measurement systems (Fisher, 2019), few attempts in PM literature elaborate vast amounts of users' short-text comments for performance measurement of organizations (e.g., Chai, 2017; McAuliffe and Blei, 2008). Moreover, extant literature shows a lack of structured procedures to address the challenges of undergoing the whole process of collecting, managing, maintaining, elaborating and interpreting unstructured data such as online UGCs, but rather focus on specific sub-phases of this process (e.g., Wang et al., 2019; Yang et al., 2019). To fill in this gap, this study aims at answering the sub-research question

(RQ₁) Which mathematical representations of online user-generated texts are most appropriate to address performance measurement tasks?

to offer a methodological contribution to extant performance measurement literature by proposing a methodological framework that highlights the necessary choices to be made to accurately select the most

adequate text embedding and the phases to be followed to exploit the properties of online UGTs in supporting the desired PM task.

Another knowledge gap emerging from the analysis literature is associated to the lack of a link between the adoption of online UGTs for decision making of cultural institutions. Although cultural institutions are the more and more oriented towards their audience and public (Agostino and Arnaboldi, 2020) and many contributions recognize the potential of adopting online UGCs to increase the knowledge of users and to better inform the decision making of cultural institutions (e.g. Capodiecì et al., 2019; Fu et al., 2018; Marchiori and Cantoni, 2015), it is not yet clear neither how cultural institutions could exploit the information conveyed by online UGCs nor whether these institutions are actually willing to exploit these information to identify, measure and monitor their performances. Among the studies concerned with cultural institutions which consider the perspectives of users, just few studies exploit the own words of cultural users to evaluate dimensions of cultural experiences (e.g., Dimache et al., 2017; Jagger et al., 2012; Laing et al., 2014; Craig Wight, 2020; Simeon et al., 2017; Su & Teng, 2018), with only an extremely narrow set of studies entering the realm of automatic analysis of online UGTs (i.e., Kirilenko et al., 2021; Taecharungroj & Mathayomchan, 2019). This research aims to enter this emerging stream of literature addressing the following research questions:

(RQ₂) To what extent do museum quality dimensions evaluated from online reviews in a “bottom-up” fashion differ from those identified in a “top-down” fashion?

(RQ₃) Which are (if any) the differences in latent dimensions of museum visit experience perceived by reviewers who write reviews in the local language compared to those of reviewers who write in other languages?

In particular, answering RQ₂ allows to discuss the potential of informing decision making through “bottom-up” and unsupervised approaches to online UGTs, since traditional “top-down” approaches limit the decision makers’ knowledge on users’ interests. Indeed, “top-down” approaches that screen public opinions in search of specific dimensions defined by the policy makers are not able to grasp the variety of aspects and the detail with which online users evaluate the cultural experience. On the contrary, the specific application to museums shows that “bottom-up” approaches to the analysis of online UGTs reveal museum quality dimensions implicitly referred by online users and additional dimensions with respect to those set by policy makers. The answer to RQ₃ further underlines the value that cultural institutions could extract from the analysis of online UGTs, showing that, without the need of having access to highly personal user profile’s information, cultural managers are able to identify the main themes of interest of cross-cultural groups of online reviewers, just accessing the text of the online description of the museum experience.

Since the answer to each of the two latter research questions (RQ₂ and RQ₃) also details the analytical steps adopted to represent and process online UGTs to better understand users’ characteristics, these two studies also contribute filling in the first knowledge gap highlighted above, by providing real examples on how online

UGTs can be used to enrich the information available to the decision maker, taking also into account the perspectives of online users of cultural institutions expressed by means of their own online voices.

To complement the previous analyses, the last research question is aimed at considering the perspective of cultural institutions towards the exploitation of this new role of users and of online UGCs in supporting the definition, measurement and monitoring of the performances of cultural institutions:

(RQ₄) How do managers of cultural institutions design, use, report and trust online UGCs as management tools?

This leads to the discussion of ways through which managers of cultural institutions could actually implement performance management systems which include online UGCs and exploit the information contained in online UGTs.

2. Relevance

This research aims to understand the role that online UGCs assume for PM of cultural institutions, investigating ways to account for measures of thoughts, perceptions and needs expressed online by cultural users.

From an academic perspective, this work aims at contributing to PM literature, exploring the possibility of including textual expressions of online users' perceptions into a system for the management of performances of institutions. Indeed, although prior researches in PM literature recognize the potential of exploiting information deriving from contents shared online by users (e.g., Lassila et al., 2019; Arnaboldi et al., 2017; Jeacle, 2017b, 2012; Jeacle and Miller, 2016; Eschenbrenner et al., 2015; Jeacle and Carter, 2014), the high subjectivity of online UGCs and their strict connection with user engagement, preferences and experiences (see also Begkos and Antonopoulou, 2020) has until now limited the number of studies relying on the textual opinions of online users for performance management. Moreover, the choice of focusing on voluntary contributions of online users to the online conversation (Peddibhotla and Subramani, 2007) poses the methodological challenge of modelling human language expressed not only in long documents but also in short text messages, which may contain ambiguous and informal expressions (Alnajran et al., 2019; Fisher, 2019). Answering the first three aforementioned research questions (RQ₁ – RQ₃), this study contributes in guiding from a methodological and an empirical perspective the exploitation of online UGTs for PM purposes. Specifically, the implementation of this PM system to the specific case of cultural institutions addressed through RQ₂ and RQ₃ is expected to offer an innovative methodological perspective to PM of cultural institutions, on the line of a more participatory approach of Culture, accounting for the perspectives of large amounts of actual cultural users.

From a managerial perspective, the answer to RQ₄ aims at discussing the actual exploitation of online UGCs as tools for PM of cultural institutions, understanding their potential and limitations also in the light of cultural managers' perspectives. This is expected to offer cultural managers a new way to monitor

performances, fostering an even more participatory approach towards their public also thanks to the inclusion of the thoughts of large sets of cultural audiences within performance management practices.

3. Key Literature

A narrative literature review⁴ of extant literature connecting the topics of performance management, text modelling, online data generated by users, in the context of cultural institutions (Figure 2) is here shortly discussed.

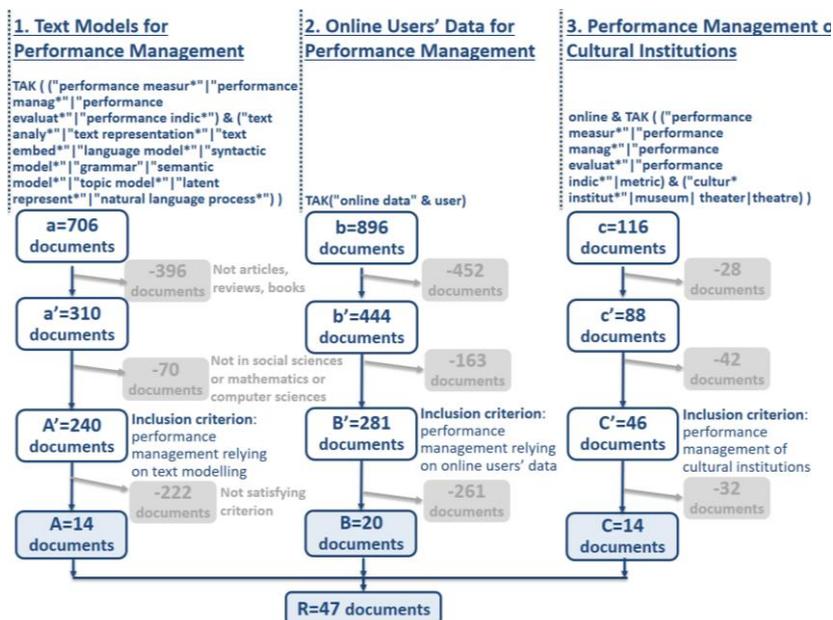
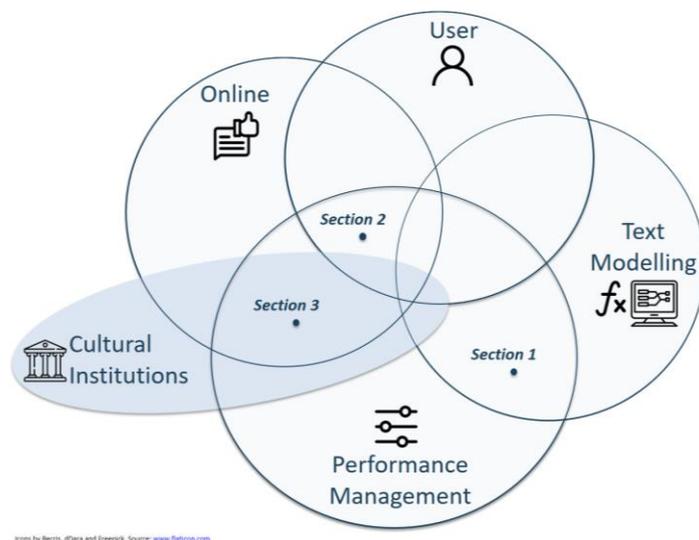


Figure 2 (top) Graphical representation of main topics (white bubbles) and sector (light-blue bubble) of interest of current research, with specification of corresponding section of narrative literature review in *italic*. **(bottom)** Literature review prism for each of the sections of narrative review.

⁴ From a methodological point of view, the same review procedure (see Figure 2) has been applied to each of the three searches. Specifically, starting from contributions presenting at least one of the specified search keywords in title, abstract or keywords, the search sample has been narrowed considering only peer-reviewed contributions in Mathematics, Computer Science or any discipline classified by Scopus as pertaining to Social Science.

3.1. Text Models for Performance Management

Although as separate disciplines both performance management and text modelling have been largely studied, literature connecting the two is relatively young and still little explored, with few examples that analyse text generated by *organizations*, such as text of strategic, financial or logistic reports (e.g., Kinra et al., 2020; del Mar Roldán-García et al., 2019; Fisher et al., 2016; Jeske and Liu, 2007), few studies that make use of text generated by *internal users of organizations*, such as feedbacks of employees (Ben-Hador, 2016) or written descriptions of key performance indicators by managers (van der Aa et al., 2017) and a limited number of contributions that analyses text generated by *external users with respect to the organization*, that is online user-generated text, like online reviews of actual or potential customers (Sagnika et al., 2020) or social media comments of online users (van Dieijen et al., 2020). The analysis of such literature also shows that extant contributions do not share neither the same methodology for text representation nor the same analytical purpose. For example, representing online user-generated texts through topic models has been recognized to be helpful in identifying and summarizing relevant contents to inform decisions (Tazibt and Aoughlis, 2019), while enriching the user-generated text embedding for sentiment tasks has shown to potentially increase the understanding of online users' opinions and of customers' satisfaction towards specific products or services (Sagnika et al., 2020). Moreover, although recent contributions recognize the importance of looking at user-generated texts such as comments of customers in implementing performance measurement systems (Fisher, 2019), just few attempts in performance management literature elaborate vast amounts of users' short-text comments for the measurement of performances of organizations (e.g., Chai, 2017; McAuliffe and Blei, 2008)

3.2. Online Users' Data for Performance Management

Extant literature connecting the worlds of online users' data with performance management literature can be divided according to the type of online user data, i.e. *personal information*, *movement* and *online UGC*. While studies of online users' behaviours relying on user's *personal information*, such as social media profile information, require managers and analysts to access rich users' information (e.g., Howden et al., 2014; Rosen et al., 2011), a recent stream of studies of online users' behaviours has been relying on digital traces of users' *movements*, coming from digital spaces, like answers to surveys on online purchase behaviours for measuring brand perception (e.g., Tanusondjaja et al., 2015), or from physical one (e.g., Hashemi and Kamps, 2018). Rising from the idea that clicking on buttons limits user's expression and action and could result in an inaccurate proxy of the actual intentions of users (Mavragani and Tsagarakis, 2019; Stamou et al., 2008), various recent studies pursue the analysis of online data directly generated by users, namely *online UGCs*, such as photos, videos, audio or written texts. Indeed, since neither researchers nor platforms guide the development of such online contents, various authors have recognized the richness of these data, which hinder perspectives, ideas and perceptions of users (e.g., Jeffrey et al., 2019; Suppala and Rao, 2019; Oteros-Rozas et al., 2018; Seresinhe et al., 2018; Soler and Gémar, 2017; Xiao et al., 2017). However, coupled to this richness, comes the difficulty of relying on variegated forms of unstructured data, like texts, images, audio and video (Preis et al., 2020; Liu et al., 2016) and the need to focus on how to actually exploit such information

to support organizations in taking informed decisions (Hashemi and Kamps, 2018). With particular reference to user-generated texts, various studies have been conducted to understand users' needs, desires and experiences based on the free-text they deliver online (e.g., Suppala and Rao, 2019; Xing et al., 2019; Yang et al., 2019; Li et al., 2018; Pera and Ng, 2017; Liu et al., 2016; Ramanujam and David, 2016; Marchiori and Cantoni, 2015; Capra and Pérez-Quñones, 2005), but, even if the diffusion of these studies have incentivized businesses and organizations to exploit the latent potential of online UGCs (Agathangelou et al., 2018; Wu et al., 2017), little exists in this direction in performance management literature.

3.3. Performance Management of Cultural Institutions

The analysis of literature connecting the worlds of online data and performance management of cultural institutions evidences two main streams. On the one hand, a stream discusses around the definition of *metric and measures* aligned with the specific mission of cultural institutions, debating the difficulties for cultural institutions to adapt their performance measures in accordance to the specific mission of the cultural institution (Gałęcka and Smolny, 2019; Hatton, 2012) and the necessity to overcome too simplified metrics, move for instance beyond simple statistics on visitors of cultural institutions (Simpson, 2017). On the other hand, studies discuss the possibilities opened by digital technologies to enrich the set of *data sources* that can be used to feed a performance management system of cultural institutions, underlying the role played by new technologies for instance in monitoring the quality of artistic collection and guaranteeing a fast intervention in case of structural instabilities (Hess et al., 2015) and in managing artistic collection through digital archives (Maggio et al., 2017). Exploiting the perspectives and opinions of the audience of cultural institutions establishing an open dialogue with their public, Capodiecì et al. (2019) and Hashemi and Kamps (2018) offer examples of the potential role of online users' data in supporting cultural institutions, measuring the online discussion around cultural contents through social media data (Capodiecì et al., 2019) or exploiting the interactions of users with IoT technologies within cultural exhibitions to enrich visitors' experience (Hashemi and Kamps, 2018).

In the specific context of museums, various studies have investigated the dimensions of museum experience and the perception of visitors using different methodologies, but mostly relying on surveys that measure a predefined set of dimensions of museum visit experience (e.g., Nowacki & Kruczek, 2021; Oren et al., 2021; Richards et al., 2020). Very little has been yet done in the direction of exploiting the own-words of visitors expressed through online reviews to inform managers decisions (e.g., Craig Wight, 2020; Kirilenko et al., 2021; Simeon et al., 2017; Su & Teng, 2018; Taecharunroj & Mathayomchan, 2019), especially with reference to the adoption of statistical models for the identification of dimensions of museum visit from the content of reviews (i.e., Kirilenko et al., 2021; Taecharunroj & Mathayomchan, 2019). Notwithstanding the pivotal role played by visitors for museums (e.g., Craig Wight, 2020; Deligiannis et al., 2020; ICOM, 2007; Villaespesa & Álvarez, 2020), the assessment of dimensions of museum visit experience through the online voices of visitors is still in an embryonal phase (i.e., Kirilenko et al., 2021; Simeon et al., 2017; Su & Teng, 2018; Taecharunroj & Mathayomchan, 2019). Specifically, the adoption of automated models for the

identification of the dimensions of museum visit experience remains an exception rather than a praxis (i.e., Kirilenko et al., 2021; Taecharungroj & Mathayomchan, 2019).

4. Research Methodology

From a methodological viewpoint, the answer to the aforementioned research questions is developed in four main steps (see also Table 1):

- (a) Critical review of extant mathematical models for representation of semantic and syntactic properties of online text and of text-based indicators used to translate online user-generated texts into performance measures;
- (b) Quantitative analysis of online UGCs, including data collection and statistical analysis of text generated online by users, ranging from text of online reviews extracted from querying platforms (e.g., TripAdvisor) to posts from social media platforms (e.g., Facebook, Instagram, Twitter);
- (c) Qualitative analysis of cultural managers' perspectives towards usage of online UGCs for performance measurement;
- (d) Development and application of a framework for management of performances of cultural institutions, based on quantitative and qualitative results of analysis conducted on specific cultural institutions.

Specifically, phase (a) consists in the definition of a taxonomy of mathematical models used in extant literature to embed semantical and syntactical properties of text and of text-based indicators to translate text into performance measures. On top of this, a critical review is expected to guide the selection of the appropriate model for the representation of online text according to the specific PM task to be pursued.

In methodological phase (b), quantitative analyses of online UGCs are expected to offer a practical example of the usability of the results of the analyses conducted in phase (a). Therefore, this phase is articulated into the following main steps:

- identification of sample of cultural institutions;
- manual collection of online channels and platforms (e.g., websites or social media pages or online querying platforms) from which data need to be downloaded;
- definition and implementation of online data crawler for data collection;
- definition of time-horizon and frequency of data collection;
- actual download of online data;
- data pre-processing and cleaning;
- data analysis, including quantitative analysis of online text (e.g. topic models, like Latent Dirichlet Allocation);
- critical analysis and interpretation of results;
- reporting and discussion of results.

Qualitative analyses of cultural managers' perspectives towards the usage of online UGCs for performance measurement (phase c) are aimed at exploring the position of managers with regard to current usage but also of perspective potential and actual usage of online UGCs within a performance management system for cultural institutions. Specifically, adopting multiple case study methodology will require:

- identification of sample of cultural institutions;
- identification of key actors to be interviewed within organizations;
- data collection, through primary (e.g., interviews) and secondary sources (e.g., archival data);
- qualitative data analysis;
- critical analysis and discussion of results.

Mixing quantitative and qualitative methods, phase (d) consists in the development and in the empirical application of a framework for performances management of specific cultural institutions based also on textual forms of online UGCs. This phase is expected to showcase the applicability of the framework to specific selected cultural institutions and to offer the opportunity to describe the entire pipeline required for the analyses of online UGCs to support the management of performances, including collection and management, pre-processing and elaboration, interpretation and reporting of statistical analysis of online text generated by users.

The empirical setting of this research (see also Table 1) is constituted by museums and opera theatres, two particular types of cultural institutions whose socio-economic activities have beneficial influence on the community (Rentschler & Potter, 1996). Due also to the recent push of the European Commission towards cultural and creative sector, this thesis frames the analyses within the European context, with a special focus on Italy which has also been recognized to be the first country worldwide in terms of UNESCO World Heritage sites.

Driven by the exploratory nature of this research, qualitative analyses intend to target both opera theatres and museums, to contribute to the debate on the specificity of perspectives across cultural institutions (Hatton, 2012). The possibility to establish contacts with specific cultural institutions is expect to be favoured also by the connection of this thesis with the Observatory in Digital Innovation in Heritage and Culture.

With specific reference to the quantitative analyses of online UGCs, museums are particularly suited for the purposes of this research. Indeed, museums have the peculiarity of offering cultural experiences rather than services (Throsby, 2009; Anton et al., 2018), characterised by a strong social component (Throsby, 2009) and they are recently moving towards a more participatory approach (e.g., Agostino 2020; Minkiewicz et al., 2014; Kéfi and Pallud, 2011), allowing online users not only to access virtually to artistic collections (Gil-Fuentetaja and Economou, 2019), but also giving them an active role in shaping cultural experience (Galani and Kidd, 2019; Hashemi and Kamps, 2018; Ritzer et al., 2012). More specifically, Italian public museums constitute a favourable empirical setting to ground this research, since the Italian Ministry for Cultural Heritage and Tourism has recently launched a call to encourage museums nurturing audience engagement also through online channels such, as social media platforms (MIBACT, 2019), and it has also provided museums

with an online platform to monitor their online performances, in terms of online activity, followers and engagement.

5. Expected Outputs

First, in answering the research question RQ₁ this work is expected to offer a critical review of state-of-the-art mathematical models for text representation, allowing to highlight assumptions and limitations associated to each text model. This is expected to guide towards the definition of a methodological framework to select an appropriate model for online UGTs analysis according to the specific performance management task of interest.

Second, applying the framework developed to answer RQ₁, this work addresses research questions RQ₂ and RQ₃ to show in practice the necessary phases to be followed to exploit a particular type of highly accessible online UGTs, i.e. online reviews, and provide evidence of the usability of online UGTs as tools for PM of cultural institutions. In particular, this allows to identify the performance dimensions of specific types cultural institutions (i.e. museums) based only on the online description of users' cultural experiences. Specifically, the answer to RQ₂ contributes to the discussions on consequences of adopting traditional "top-down" approaches, instead of moving forward to more inclusive and participatory "bottom-up" approaches, showing that online UGTs can actually provide valuable insights for both the evaluation and identification of performances moving beyond the typical service-centred approaches. The answer to RQ₃ offers the possibility to dig into the online users perspectives, understanding differences in preferences and experiences of online users exploiting only highly accessible forms of online data voluntarily shared online by users, namely recognizing cultural groups of users simply on the language in which online users write online.

Lastly, answering the research question RQ₃ this work expects to present a framework for PM of cultural institutions based also on online UGCs and to discuss its usability directly with cultural managers, exploring their perspectives towards the design, usage, reporting and trust of performance management tools based on online UGCs.

6. Potential Contributions

First, answering the research question RQ₁ this work aims at contributing to literature connecting the two disciplines of PM and text modelling, offering clear technical guidelines on which model to employ according to the specific PM task and practical suggestions on usability of user-generated texts in PM (Fisher, 2019).

Second, addressing research question RQ₂ and RQ₃, this work contributes in filling in the gap in literature connected to the actually exploitation of online UGTs to support decision maker of cultural institutions, widening the understanding on cultural users' perspectives. Providing examples in this direction contributes to the definition of guidelines to facilitate the management of these types of online data and therefore support an actual and efficient exploitation of the latent potential that these types of data convey for performance management and for the understanding of users' characteristics. Beyond contributing in filling in the gap in

the adoption of automated models for the identification of the dimensions of museum visit experience (i.e., Kirilenko et al., 2021; Taecharunroj & Mathayomchan, 2019), each of the two research questions has additional contributions. Specifically, the second research question RQ₂ is expected to contribute to the debate on the measurement of dimensions of museum experience comparing the perspectives of policy makers and users, while the few studies interested in this topics do not compare the two perspectives. The answer to the third research question RQ₃ is aimed at providing an empirical examples of the cross-cultural differences in the dimensions of museum visit experience distinguishing museums visitors based on their language, hence empirically showing that language is a valid proxy for cultural background (e.g., Trinh & Ryan, 2017) that can be used as alternative to the nationality or country of origin of users.

Third, answering the research question RQ₄ this work contributes to literature discussing the potential of adopting online UGCs to increase the knowledge of users and to better inform these institutions (e.g. Capodiecì et al., 2019; Fu et al., 2018; Marchiori and Cantoni, 2015), proposing an *actual* implementation of the usage of online UGCs for performance management of cultural institution, discussing its strengths and limitations.

References

- Agag G.M., El-Masry A.A. (2017). Why Do Consumers Trust Online Travel Websites? Drivers and Outcomes of Consumer Trust toward Online Travel Websites. *Journal of Travel Research*
- Agathangelou P., Katakis I., Koutoulakis I., Kokkoras F. & Gunopulos D. (2018). Learning patterns for discovering domain-oriented opinion words. *Knowledge and Information Systems*
- Agostino D. (2018) Can Twitter Add to Performance Evaluation in the Area of Performing Arts? Reflections from La Scala Opera House. Routledge, *Journal of Arts Management Law and Society*
- Agostino D. & Arnaboldi M. (2020) From preservation to entertainment: Accounting for the transformation of participation in Italian state museums. *Accounting History*
- Agostino D. & Arnaboldi M. (2019) What can social media data add to the knowledge of arts and humanities? an empirical investigation on twitter at teatro alla scala. Springer Nature, *Knowledge Management and Organizational Learning*
- Agostino D. & Arnaboldi M. (2017). Social media data used in the measurement of public services effectiveness: Empirical evidence from Twitter in higher education institutions, *Public Policy and Administration, Vol. 32, No.4*, pp. 296-322
- Agostino D.& Arnaboldi M. (2016) A Measurement Framework for Assessing the Contribution of Social Media to Public Engagement: An empirical analysis on Facebook, *Public Management Review*
- Agostino D. & Sidorova Y. (2017), How social media reshapes action on distant customers: some empirical evidence, *Accounting, Auditing and Accountability Journal*, Vol. 30 No. 4, pp. 777-794.
- Agostino D. & Sidorova Y. (2016) A performance measurement system to quantify the contribution of social media: new requirements for metrics and methods. Emerald Group Publishing Ltd., *Measuring Business*

Excellence

- Alnajran N., Crockett K., McLean D., Latham A. (2019) An Empirical Performance Evaluation of Semantic-Based Similarity Measures in Microblogging Social Media. Proceedings - 5th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2018
- Antón, C., Camarero, C., & Garrido, M.-J. (2018). Exploring the experience value of museum visitors as a cocreation process. *Current Issues in Tourism*, 21(12), 1406-1425.
- Antonelli V, D'alessio R & Iuliano G (2007) Art and accounting history: The Teatro San Carlo of Naples, 1737-1786. *Accounting, Business & Financial History* 17(1): 165–186.
- Arnaboldi M., Busco C. & Cuganesan S. (2017), “Accounting, accountability, social media and big data: revolution or hype?”, *Accounting, Auditing and Accountability Journal*, Vol. 30 No. 4, pp. 762-776.
- Athwal N., Istanbuluoglu D. & McCormack S.E. (2019) The allure of luxury brands’ social media activities: a uses and gratifications perspective. *Information Technology and People*
- Begkos C. & Antonopoulou K. (2020) Measuring the unknown: Evaluative practices and performance indicators for digital platforms, *Accounting, Auditing and Accountability Journal*
- Ben-Hador B. (2016) Coaching executives as tacit performance evaluation: a multiple case study, *Journal of Management Development*
- Bertacchini E. & Morando F. (2013) The future of museums in the digital age: New models for access to and use of digital collections. *International Journal of Arts Management*
- Bialecki M., O’Leary S. & Smith D. (2017), “Judgement devices and the evaluation of singularities: the use of performance ratings and narrative information to guide film viewer choice”, *Management Accounting Research*, Vol. 35, pp. 56-65.
- Bonet L & Négrier E (2018) The participative turn in cultural policy: Paradigms, models, contexts. *Poetics* 66: 64–73.
- Bonsón E., & Ratkai M. (2013). A Set of Metrics to Assess Stakeholder Engagement and Social Legitimacy on a Corporate Facebook Page, *Online Information Review Vol. 37, No.5*, pp. 787–803
- Cao T.-D. & Nguyen Q.-M. (2012). Semantic approach to travel information search and itinerary recommendation. *International Journal of Web Information Systems*
- Capodieci A., Elia G., Grippa F. & Mainetti L. (2019) A network-based dashboard for cultural heritage promotion in digital environments. *International Journal of Entrepreneurship and Small Business*
- Capra III R.G. & Pérez-Quiñones M.A. (2005) Using Web search engines to find and refind information. *Computer*
- Caylor M., Cecchini M., Winchel J. (2017) Analysts' qualitative statements and the profitability of favorable investment recommendations. *Accounting, Organizations and Society*
- Chai, C. (2017) Statistical issues in quantifying text mining performance. PhD Dissertation. Department of Statistical Science, Duke University, Durham.
- Chiaravalloti F. (2014). Performance evaluation in the arts and cultural sector: A story of accounting at its

- margins. *Journal of Arts Management Law and Society*, 44(2), 61–89.
<https://doi.org/10.1080/10632921.2014.905400>
- Craig Wight, A. (2020). Visitor perceptions of European Holocaust Heritage: A social media analysis. *Tourism Management*, 81(May). <https://doi.org/10.1016/j.tourman.2020.104142>
- Crossick G (2018) Museums, audiences and collections: Thinking about historical context. *Cultural Trends* 27(4): 232–238.
- Deidda Gagliardo E., Madonna S. & Pierotti M. (2012) Evolution of Governance Models and Accounting Information System. the Case of Teatro Alla Scala. Roma: Rirea.
- del Mar Roldán-García M., García-Nieto J., Maté A., Trujillo J., Aldana-Montes J.F. (2019) Ontology-driven approach for KPI meta-modelling, selection and reasoning, *International Journal of Information Management*
- Department for Digital, Culture, Media and Sport, Gov. UK, (2019). Culture is Digital. <https://www.gov.uk/government/publications/culture-is-digital/culture-is-digital-june-2019-progress-report>
- Dimache, A., Wondirad, A., & Agyeiwaah, E. (2017). One museum, two stories: Place identity at the Hong Kong Museum of History. *Tourism Management*, 63, 287–301.
<https://doi.org/10.1016/j.tourman.2017.06.020>
- Eschenbrenner B., Nah F.F.-H. & Telaprolu V.R. (2015) Efficacy of social media utilization by public accounting firms: Findings and directions for future research, *Journal of Information Systems*
- EU Proposal Regulation (2018) Proposal for a Regulation of the European Parliament and of the Council Establishing the Creative Europe Programme (2021 to 2027) and Repealing Regulation (EU) No 1295/2013, COM/2018/366 final - 2018/0190 (COD), Link: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A366%3AFIN>
- Falk, J. H., & Dierking, L. D. (2016). The Museum Experience Revisited. In *The Museum Experience Revisited*. <https://doi.org/10.4324/9781315417851>
- Falk J.H., Moussouri T., & Coulson D. (1998). “The effect of visitors ‘agendas on museum learning”, *Curator: The Museum Journal*, Vol 41, No2, pp. 107-120.
- Fisher I.E., Garnsey M.R., & Hughes M.E. (2016) Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research. *Intelligent Systems in Accounting, Finance and Management*
- Fisher N.I. (2019) A comprehensive approach to problems of performance measurement, *Journal of the Royal Statistical Society. Series A: Statistics in Society*
- Fu D., Hong Y., Wang K. & Fan W. (2018). Effects of membership tier on user content generation behaviors: evidence from online reviews. *Electronic Commerce Research*
- Galani, A., Kidd, J. (2019). Evaluating digital cultural heritage “in the wild”: The case for reflexivit *Journal on Computing and Cultural Heritage*, 12 (1).

- Gałecka M. & Smolny K. (2019) Criteria for the optimal financing model of public theatres. Review of Economic Perspectives
- Gil-Fuentetaja, I., Economou, M., (2019). Communicating Museum Collections Information Online: Analysis of the Philosophy of Communication Extending the Constructivist Approach. *Journal on Computing and Cultural Heritage*, 12(1), 1-16.
- Goepp V., Matta N., Caillaud E., Feugeas F. (2019) Systematic community of Practice activities evaluation through Natural Language Processing: Application to research projects, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*
- Hashemi S.H. & Kamps J. (2018) Exploiting-behavioral in online news consumption: Evidence from survey and navigation data in Spain. *New Review of Hypermedia and Multimedia*
- Hatton A. (2012) The conceptual roots of modern museum management dilemmas. *Museum Management and Curatorship*
- Hess M., Korenberg C., Ward C., Robson S. & Entwistle C. (2015) Use of 3D laser scanning for monitoring the dimensional stability of a Byzantine ivory panel. *Studies in Conservation*
- Honari A. (2017) The formation of consensus in Iranian online communities, *Online Communities as Agents of Change and Social Movements*
- Howden C., Liu L., Li Z.Y., Li J.X. & Antonopoulos N. (2014) Virtual vignettes: The acquisition, analysis, and presentation of social network data. *Science China Information Sciences*
- ICOM, International Council of Museums. (2020). How to Reach Your Public Remotely. <https://icom.museum/en/covid-19/resources/how-to-reach-your-public-remotely/>
- ICOM & OECD, International Council of Museums & Organisation for Economic Co-operation and Development. (2020) Coronavirus (COVID-19) and museums: impact, innovations and planning for post-crisis. Webinar: April 10, 2020. <https://icom.museum/en/covid-19/resources/icom-oecd-webinar/>
- Jacobsen J.W. (2014) The community service museum: Owning up to our multiple missions. *Museum Management and Curatorship*
- Jagger, S. L., Dubek, M. M., & Pedretti, E. (2012). ‘It’s a personal thing’: Visitors’ responses to Body Worlds. *Museum Management and Curatorship*, 27(4). <https://doi.org/10.1080/09647775.2012.720185>
- Jeacle I. (2017a), “Constructing audit society in the virtual world: the case of the online reviewer”, *Accounting, Auditing and Accountability Journal*, Vol. 30 No. 1, pp. 18-37.
- Jeacle I. (2017b), “Managing popular culture”. *Management Accounting Research*, Vol. 100 No. 35, pp. 1-4.
- Jeacle I. (2012), “Accounting and popular culture: framing a research agenda”, *Accounting, Auditing and Accountability Journal*, Vol. 25 No. 4, pp. 580-601.
- Jeacle I. & Carter C. (2014) Creative spaces in interdisciplinary accounting research, *Accounting, Auditing and Accountability Journal*
- Jeacle I. & Carter C. (2011), “In TripAdvisor we trust: rankings, calculative regimes and abstract systems”, *Accounting, Organizations and Society*, Vol. 36 No. 4, pp. 293-309.

- Jeacle I. & Miller P. (2016), “Accounting, culture, and the state”, *Critical Perspectives on Accounting*, Vol. 37, pp. 1-4.
- Jeffrey H. L., Ashraf H. & Paris C.M. (2019): Hanging out on Snapchat: disrupting passive covert netnography in tourism research, *Tourism Geographies*, DOI: 10.1080/14616688.2019.1666159
- Jeske D.R., Liu R.Y. (2007) Mining and tracking massive text data: Classification, construction of tracking statistics, and inference under misclassification, *Technometrics*
- Ji Y. & Sproul G. (2013) FreeMobility-Dynamic in online news consumption: Evidence from survey and navigation data in Spain. *Wireless Communications and Mobile Computing*
- Kinra A., Hald K.S., Mukkamala R.R., Vatrappu R. (2020) An unstructured big data approach for country logistics performance assessment in global supply chains, *International Journal of Operations and Production Management*
- Kirilenko, A. P., Stepchenkova, S. O., & Dai, X. (2021). Automated topic modeling of tourist reviews: Does the Anna Karenina principle apply? *Tourism Management*, 83(September 2020). <https://doi.org/10.1016/j.tourman.2020.104241>
- Kotler N. & Kotler P. (2000). “Can museums be all things to all people?: Missions, goals, and marketing's role”, *Museum Management and Curatorship*, Vol.18, No.3, 271-287.
- Kornberger M., Pflueger D. & Mouritsen J. (2017), “Evaluative infrastructures: accounting for platform organization”, *Accounting, Organisations and Society*, Vol. 60, pp. 79-95.
- Laing, J., Wheeler, F., Reeves, K., & Frost, W. (2014). Assessing the experiential value of heritage assets: A case study of a Chinese heritage precinct, Bendigo, Australia. *Tourism Management*, 40, 180–192. <https://doi.org/10.1016/j.tourman.2013.06.004>
- Lapsley I & Rekers JV (2017) The relevance of strategic management accounting to popular culture: The world of West End Musicals. *Management Accounting Research* 35: 47–55.
- Leoni G. & Parker L.D. (2019), “Governance and control of sharing economy platforms: hosting on Airbnb”, *The British Accounting Review*, Vol. 51 No. 6, pp. 100814.
- Lassila E.M., Moilanen S. & Järvinen J.T. (2019) Visualising a “good game”: analytics as a calculative engine in a digital environment, *Accounting, Auditing and Accountability Journal*
- Li S., Cheng X., Su S., Sun H. (2017) Exploiting organizer influence and geographical preference for new event recommendation, *Expert Systems*
- Li Y., Peng Y., Zhang Z., Yin H. & Xu Q. (2019). Matching user accounts across social networks based on username and display name. *World Wide Web*
- Li Y., Peng Y., Ji W., Zhang Z. & Xu Q. (2017) User Identification Based on Display Names Across Online Social Networks. *IEEE Access*
- Li Y., Zhang Z., Peng Y., Yin H. & Xu Q. (2018). Matching user accounts based on user generated content across social networks. *Future Generation Computer Systems*
- Liu X., Singh P.V. & Srinivasan K. (2016). A structured analysis of unstructured big data by leveraging cloud

computing. Marketing Science

- MacKenzie J. (2017) Identifying informational norms in Mumsnet Talk: A reflexive-linguistic approach to internet research ethics. *Applied Linguistics Review*
- Maggio A., Kuffer J., Lazzari M. (2017) Advances and trends in bibliographic research: Examples of new technological applications for the cataloguing of the georeferenced library heritage, *Journal of Librarianship and Information Science*
- Marchiori E. & Cantoni L. (2015) The role of prior experience in the perception of a tourism destination in user-generated content. *Journal of Destination Marketing and Management*
- Mavragani A. & Tsagarakis K.P. (2019). Predicting referendum results in the Big Data Era. *Journal of Big data*
- McAuliffe, J. D. and Blei, D. M. (2008) Supervised topic models. In *Advances in Neural Information Processing Systems*, pp. 121–128.
- McDaid E., Boedker C. & Free, C. (2019), “Close encounters and the illusion of accountability in the sharing economy”, *Accounting, Auditing and Accountability Journal*, Vol. 32 No. 5, pp. 1437-1466.
- MIBACT, Ministero per i beni e le attività culturali e per il turismo (2019), Piano Triennale per la Digitalizzazione e l’Innovazione dei Musei, <http://musei.beniculturali.it/wp-content/uploads/2019/08/Piano-Triennale-per-la-Digitalizzazione-e-l%E2%80%99Innovazione-dei-Musei.pdf>
- Mihelj S, Leguina A & Downey J (2019) Culture is digital: Cultural participation, diversity and the digital divide. *New Media & Society* 21: 1465–1485.
- Minkiewicz, J., Evans, J., & Bridson, K. (2014). How do consumers co-create their experiences? An exploration in the heritage sector. *Journal of Marketing Management*, 30(12), 30–59.
- Munley M. (2011) The Human Origins Initiative (HOI) What Does It Mean To Be Human? Evaluation #1 Key Findings Report prepared for Smithsonian National Museum of Natural History. Washington, DC.
- NEMO, Network of European Museum Organisations. (2020). Survey on the Impact of the Corona Crisis on Museums in Europe. Results updated at 6 April 2020. <https://www.ne-mo.org/news/article/nemo/nemo-publishes-results-of-survey-on-the-impact-of-the-corona-crisis-on-museums-in-europe.html>
- Nielsen J. K. (2015). The relevant museum: Defining relevance in museological practices. *Museum Management and Curatorship*, 30(5), 364-378.
- OECD, Organisation for Economic Co-operation and Development. (2007). Participative Web: User-Created Content. *Policy*. <https://doi.org/10.1787/9789264037472-en>
- OMC, Open Method of Coordination. (2014). *Report of Open Method of Coordination group on the role of public arts and cultural institutions in the promotion of cultural diversity and intercultural dialogue. January, 2014*. https://ec.europa.eu/culture/policy/strategic-framework/european-coop_en
- Oteros-Rozas E., Martín-López B., Fagerholm N., Bieling C. & Plieninger T. (2018). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five

European sites. *Ecological Indicators*

- Oviedo-García A.M., Muñoz-Expósito M., Castellanos-Verdugo M. & Sancho-Mejías M. (2014). Metric proposal for customer engagement in Facebook. *Journal of Research in Interactive Marketing, Vol. 8, No.4*, pp. 327-344.
- Kéfi, H., & Pallud, J. (2011). Digital heritage the role of technologies in cultural mediation in museums: An actor-network theory view applied in France. *Museum Management and Curatorship, 26(3)*, 273-289.
- Pangallo M. & Loberto M. (2018) Home is where the ad is: online interest proxies housing demand. *EPJ Data Science*
- Park K., Peng L. (2018) A development of LDA topic association systems based on spark-hadoop framework, *Journal of Information Processing Systems*
- Peddibhotla N. & Subramani M. (2007), “Contributing to public document repositories: a critical mass theory perspective”, *Organization Studies, Vol. 28 No. 3*, pp. 327-346.
- Pera M.S. & Ng Y.-K. (2017) Using online data sources to make query suggestions for children. *Web Intelligence*
- Peral J., Ferrández A., Mora H., Gil D., Kauffmann E. (2019) A Review of the Analytics Techniques for an Efficient Management of Online Forums: An Architecture Proposal, *IEEE Access*
- Pierotti M., Capocchi A. & Orlandini P. (2019) Mutual aid for the arts in nineteenth- and twentieth-century Milan: The ‘Pio Istituto Teatrale’, *Accounting History*
- Preis T., Botta F. & Moat H.S. (2020). Sensing global tourism numbers with millions of publicly shared online photographs. *Environment and Planning A (EPA): Economy and Space 2020, Vol. 52(3)* 471–477 DOI: 10.1177/0308518X19872772
- Quattrone P. (2016), “Management accounting goes digital: will the move make it wiser?”, *Management Accounting Research, Vol. 31, June*, pp. 118-122.
- Ramanujam K.S. & David K. (2016) POBSIRS-Personalized in online news consumption: Evidence from survey and navigation data in Spain. *International Journal of Control Theory and Applications*
- Reisenbichler M., Reutterer T. (2019) Topic modeling in marketing: recent advances and research opportunities, *Journal of Business Economics*
- Rentschler R. & Potter B. (1996). Accountability versus artistic development. *Accounting, Auditing & Accountability Journal, Vol. 9 Iss 5 pp. 100 – 113* <http://dx.doi.org/10.1108/09513579610151971>
- Ritzer, G., Dean, P., & Jurgenson, N. (2012). The coming of age of the prosumer. *American Behavioral Scientist, 56(4)*, 379-398.
- Rosen D., Barnett G.A. & Kim J.H. (2011) Social networks and online environments: when science and practice co-evolve. *Social Network Analysis and Mining*
- Sagnika S., Mishra B.S.P., Meher S.K. (2020) Improved method of word embedding for efficient analysis of human sentiments, *Multimedia Tools and Applications*
- Sang J., Xu C., Lu D. (2012) Learn to personalized image search from the photo sharing websites, *IEEE*

- Scott S.V. & Orlikowski W.J. (2012), “Reconfiguring relations of accountability: materialization of social media in the travel sector”, *Accounting, Organizations and Society*, Vol. 37 No. 1, 26-40.
- Scott A.J. (2004). Cultural-products industries and urban economic development: prospects for growth and market contestation in global context. *Urban Aff. Rev.* 39 (4), 461–490.
- Seresinhe C.I., Moat H.S. & Preis T. (2018) Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science*
- Simeon, M. I., Buonincontri, P., Cinquegrani, F., & Martone, A. (2017). Exploring tourists’ cultural experiences in Naples through online reviews. *Journal of Hospitality and Tourism Technology*, 8(2), 220–238. <https://doi.org/10.1108/JHTT-10-2016-0067>
- Simpson A.J. (2017) Beyond visitor statistics: value propositions and metrics for university museums and collections. *Museum Management and Curatorship*
- Soler I.P., G emar G. (2017) Brand Equity Research Using Online Customer Ratings of Spanish Hotels. *International Journal of Tourism Research*
- Stamou S., Kozanidis L., Tzekou P. & Zotos N. (2008) A-recommendation in online news consumption: Evidence from survey and navigation data in Spain. *Journal of Digital Information Management*
- Su, Y., & Teng, W. (2018). Contemplating museums’ service failure: Extracting the service quality dimensions of museums from negative on-line reviews. *Tourism Management*, 69(August 2017), 214–222. <https://doi.org/10.1016/j.tourman.2018.06.020>
- Suppala K. & Rao N. (2019). Sentiment analysis using na ive bayes classifier. *International Journal of Innovative Technology and Exploring Engineering*
- Tanusondjaja A., Greenacre L., Banelis M., Truong O. & Andrews T. (2015). International brands in emerging markets: the myths of segmentation. *International Marketing Review*
- Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, 75(April), 550–568. <https://doi.org/10.1016/j.tourman.2019.06.020>
- Tazibt A.A., Aoughlis F. (2019) Latent Dirichlet allocation-based temporal summarization, *International Journal of Web Information Systems*
- Trinh, T. T., & Ryan, C. (2017). Visitors to Heritage Sites: Motives and Involvement—A Model and Textual Analysis. *Journal of Travel Research*, 56(1), 67–80. <https://doi.org/10.1177/0047287515626305>
- Throsby, C. D. (2009). Introduction and Overview. In V. Ginsburgh & C. D. Throsby (Eds.), *Handbook of the Economics of Art and Culture*. Amsterdam: North-holland.
- UNESCO, United Nations Educational, Scientific and Cultural Organization. (2012). *Measuring cultural participation (Vol. 2)*. Montreal, Canada: UNESCO Institute for Statistics.
- UNESCO, United Nations Educational, Scientific and Cultural Organization. (2009). *Measuring cultural participation. Framework for cultural statistics handbook no. 2. Paris: UNESCO.*
- van der Aa H., Leopold H., del-R o-Ortega A., Resinas M., Reijers H.A. (2017) Transforming unstructured natural language descriptions into measurable process performance indicators using Hidden Markov

- van Dieijen M., Borah A., Tellis G.J., Franses P.H. (2020) Big Data Analysis of Volatility Spillovers of Brands across Social Media and Stock Markets, *Industrial Marketing Management*
- Wang W., Ying S., Lyu J. & Qi X. (2019) Perceived image study with online data from social media: the case of boutique hotels in China. *Industrial Management and Data Systems*
- Wang Z., Zhang J., Ji S., Meng C., Li T., Zheng Y. (2020) Predicting and ranking box office revenue of movies based on big data, *Information Fusion*
- Windhager F., Salisu S. & Mayr E. (2019) Exhibiting uncertainty: Visualizing data quality indicators for cultural collections. *Informatics*
- Wu F., Huang Y. & Yuan Z. (2017) Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources. *Information Fusion*
- Xiao J., Li X., Chen S., Zhao X., Xu M. (2017) An inside look into the complexity of box-office revenue prediction in China. *International Journal of Distributed Sensor Networks*
- Xing L., Deng K., Wu H., Xie P., Gao J. (2019). Behavioral habits-based user identification across social networks. *Symmetry*
- Yang B., Liu Y., Liang Y. & Tang M. (2019). Exploiting user experience from online customer reviews for product design. *International Journal of Information Management*
- Yu H. (2011) Selective sampling techniques for feedback-based data retrieval. *Data Mining and Knowledge Discovery*
- Zhang J.-D., Chow C.-Y. (2016) CRATS: An LDA-Based Model for Jointly Mining Latent Communities, Regions, Activities, Topics, and Sentiments from Geosocial Network Data, *IEEE Transactions on Knowledge and Data Engineering*
- Zhang P., Xia J., Li R. (2013) Personalized multimedia information retrieval based on user profile mining, *Journal of Networks*
- Zhao Y. (2017) Online privacy protection: A legal regime for personal data protection in China, *Chinese Legal Reform and the Global Legal Order: Adoption and Adaptation*
- Zhou H., Yu H. & Hu R. (2017) Topic evolution based on the probabilistic topic model: a review. *Frontiers of Computer Science*

Table 1 Tentative structure of Ph.D. thesis, from literature gaps to expected outputs through research questions, methodological phases and empirical setting.

	Literature Gap	Research Question	Methodology	Empirical Setting	Expected Output	
[Paper 1]	Difficulties of undergoing the whole process of collecting, managing, maintaining, elaborating and interpreting unstructured online UGCs to support decisions (Preis et al., 2020; Hashemi and Kamps, 2018; Liu et al., 2016; Wang et al., 2019; Yang et al., 2019)	RQ1: Which mathematical representations of online user-generated texts are most appropriate to address performance measurement tasks?	[Phase a] Narrative Literature Review	Extensive analysis of interdisciplinary literature on mathematical models for text representation and text-based performance indicators	Performance management of cultural institutions, such as museums and opera theatres, as representatives of institutions that offer cultural experiences	Critical review of state-of-the-art models for text representation, highlighting assumptions and limitations associated to each text model
			[Phase d] Framework development			Highlight the methodological steps and choices to be undertaken in framing online UGCs within a system for measurement of performances
[Paper 2]	Exploit highly accessible and unmediated online user-generated data, especially in textual form, to better understand users' characteristics (e.g., Li et al., 2019; Seresinhe et al., 2018). The adoption of automated models for the identification of the dimensions of museum visit experience remains an exception rather than a praxis (i.e., Kirilenko et al., 2021; Taecharunroj & Mathayomchan, 2019).	RQ2: To what extent do museum quality dimensions evaluated from online reviews in a "bottom-up" fashion differ from those identified in a "top-down" fashion?	[Phase b] Quantitative analysis	Collection, pre-processing and cleaning of online reviews, including unsupervised statistical analysis of text through topic modelling technique (Latent Dirichlet Allocation), and a non-overlapping multiclass supervised keyword-based classification algorithm based on a predefined list of keywords defined by the policy maker	14,250 TripAdvisor reviews written in Italian of 100 Italian public museums collected along the year 2019	Compare bottom-up and top-down approaches to highlight the differences between the two approaches, from a modelling perspective and in terms of quality dimensions identified
			[Phase d] Framework application			

	Literature Gap	Research Question	Methodology	Empirical Setting	Expected Output	
[Paper 3]	<p>The adoption of automated models for the identification of the dimensions of museum visit experience remains an exception rather than a praxis (i.e., Kirilenko et al., 2021; Taecharunroj & Mathayomchan, 2019). Although language is recognized as a valid proxy for cultural background (e.g., Trinh & Ryan, 2017), the literature lacks in providing empirical examples of the cross-cultural differences in the dimensions of museum visit experience distinguishing museums visitors based on their language.</p>	<p>RQ3: Which are (if any) the differences in latent dimensions of museum visit experience perceived by reviewers who write reviews in the local language compared to those of reviewers who write in other languages?</p>	<p>[Phase b] Quantitative analysis</p>	<p>Collection, pre-processing and cleaning of online reviews, including unsupervised statistical analysis of text through topic modelling technique (Latent Dirichlet Allocation)</p>	<p>Relying only on the language in which visitors of cultural institutions review online, identify the main themes of interest (i.e. topics) discussed online by different language groups of cultural reviewers and analyse the heterogeneity of interests across groups</p>	
			<p>[Phase d] Framework application</p>	<p>Follow the methodological steps defined in proposed framework, justifying the choices undertaken to make use of online UGCs of textual form to support the management of performances</p>		<p>36,460 TripAdvisor reviews, written in Italian and non-Italian language, of the 30 top-visited Italian museums in 2019 collected along the year 2019</p>
[Paper 4]	<p>Potential of adopting online UGCs (e.g. Capodieci et al., 2019; Fu et al., 2018; Marchiori and Cantoni, 2015) but lack of actual usage for measurement of performances of institutions</p>	<p>RQ4: How do managers of cultural institutions design, use, report and trust online UGCs as management tools?</p>	<p>[Phase c] Qualitative analysis</p>	<p>Identify key actors from a selected sample of cultural institutions to analyse the perspectives towards the usage of online UGCs through an explorative multiple case study, combining both primary and secondary sources</p>	<p>Interviews with directors and managers of Italian public museums and European opera theatres</p>	<p>Explore and discuss the position and level of trust of managers of cultural institutions towards potential usability of proposed framework for performance measurement based on online UGCs, with particular reference to text</p>
			<p>[Phase d] Framework application and development</p>	<p>Connect the choices undertaken by cultural experts to the methodological steps and choices defined within the methodological framework, discussing differences and proposing possible modifications to proposed framework</p>		