

Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world

AMPS, Architecture_MPS, Parade, American University in Dubai
Dubai: 22-24 November, 2020

DEFINING CULTURAL DATA SCIENCE

Author:

VISHAL KUMAR

Institution:

THE BARTLETT CENTRE FOR ADVANCED SPATIAL ANALYSIS, UCL

INTRODUCTION

Since 2019 a new term called “cultural data science” has emerged as a result of the increased utilisation of data within the fields of art and humanities, cultural studies and the built environment. It differs from a similar term called “cultural analytics” in that the latter is principally concerned with the “analysis and visualization of large cultural data sets and flows”¹. Cultural data science, on the other hand, can be thought of as an extension of cultural analytics, urban studies and social science with a particular focus on applying advanced computational methods to various forms of cultural data to test hypotheses and devise new frameworks for policy, planning and market action within the creative and cultural industries.

At present there is no clear definition of cultural data science, therefore, the aim of this paper is to present the first definition of the field. Adapted from Irizarry², cultural data science is defined as the umbrella term used to describe the entire complex and multistep process used to extract value from cultural data with real-world implications to influence policy, placemaking and market dynamics within the creative and cultural industries. This new concept is a unification of several academic fields and hopes to respect and honour the ideas, methodologies and tools developed by these adjacent fields. Its intention is to assist purpose-driven collaboration between public and private organisations and has the potential to become increasingly important for stakeholders in the built environment and policy makers who want to apply novel frameworks and evidence-based methods to influence the artistic and cultural vibrancy of place, space, societies and our local economies.

This paper explains how the importance placed on the creative and cultural industries as key drivers of our economies and the increased volume and velocity of data within the industry has spurred a need for cultural data science, especially in the post-COVID-19 pandemic era where many aspects of the industry have moved online and have become digital. The paper first explains why cultural data science is required, it then discusses the distinct aspects of cultural data science, a definition of cultural data science is then provided for the first time, and finally it describes how the theory and methodologies can be applied and its limitations.

WHY WE NEED CULTURAL DATA SCIENCE

¹ Lev Manovich. Cultural Analytics: white paper (5/2007; latest update 11/2008). With contributions by Noah Wardrip-Fruin.

² Irizarry, R.A., 2020. The role of academia in data science education. Harvard Data Science Review, 2(1).

Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world

AMPS, Architecture_MPS, Parade, American University in Dubai
Dubai: 22-24 November, 2020

The cultural and creative industries have received much attention since the turn of the millenia as an important driver of our economies and stimulant for societal and urban change; more so in the developed world, but increasingly in the developing world too. As summed up by Montalto et al³, the sector is known to be responsible for constituting important parts of local identity and quality of life, creativity and innovation, tourism, urban growth and regeneration, and well-being. Consequently, particular attention has been given to the industry by international organisations such as the United Nations, OECD and European Commission, and governments of the United Kingdom and Australia have produced dedicated strategy documents for the industry such as the Creative industries: Sector Deal⁴ and Strategy: 2020-2024⁵, respectively. Moreover, city-wide cultural megaevents such as the European City of Culture and UK City of Culture and policy initiatives such as the London Borough of Culture and Cultural Creative Spaces & Cities⁶ illustrate the desire of policy makers to utilise targeted event-driven cultural programs and initiatives to stimulate culture-induced change.

It has become more common, and in some cases it has become a requirement, to undertake thorough and rigorous evidence-driven evaluation research prior to and following such initiatives, whether it is a mega cultural event, a sweeping cultural policy program, or a local creative initiative. A clear illustration of this is the London Borough of Culture Evaluation Outcome Framework⁷ developed by the Audience Agency in 2018. Furthermore, the Coventry City of Culture Trust has devised a Performance Measurement and Evaluation Strategy document⁸ alongside a theory of change for Coventry 2021 UK city of Culture. The justification for undertaking evaluations of this nature is to demonstrate the impact of cultural events and initiatives at a time when they have “increasingly come under the scrutiny of funders, policy-makers and planners”⁹. All of this is part of a broader trend of using data-driven methods to measure the impact of the cultural and creative industries. The UNESCO 2030 Cultural Indicators¹⁰ are a bold step in this direction: “the development of a new framework for measuring and collecting data on culture is foundational both for advocating for the role of culture in the SDGs, as well as for integrating culture into development plans and policies at the national and urban levels and within the United Nations Development Assistance Frameworks (UNDAF)”. Additionally, the European Commission’s Cultural and Creative Cities Monitor¹¹, developed in 2017 and updated in 2019, is motivated by the desire to create composite indicators to benchmark the cultural vibrancy, creative economy and enabling environment of 200 European cities against one another.

Alongside the growing interest in measuring and demonstrating the industry’s impact to our economies and societies, an increasing volume and velocity of data is being generated from, with and by the creative and cultural industries. Indeed, it is noted by Bruce, Malcolm & O’Neill that “the driving force behind the industry’s continued success is the growth in digital content consumption, along with the rapid development of new platforms and digital technologies, especially through the

³ Montalto, Valentina, Carlos Jorge Tacao Moura, Sven Langedijk, and Michaela Saisana. "Culture counts: An empirical approach to measure the cultural and creative vitality of European cities." *Cities* 89 (2019): 167-185.

⁴ Creative industries: Sector Deal, <https://www.gov.uk/government/publications/creative-industries-sector-deal/creative-industries-sector-deal-html#further-information>

⁵ Strategy: 2020-2024,

<https://www.australiacouncil.gov.au/workspace/uploads/files/strategic-plan-discussion-pape-5c94330fc8729.pdf>

⁶ Cultural Creative Spaces & Cities, website: <https://www.spacesandcities.com/>

⁷ The Audience Agency, https://www.london.gov.uk/sites/default/files/lboc_outcomes_framework3.pdf

⁸ Coventry 2021, Performance Measurement and Evaluation Strategy document:

<https://warwick.ac.uk/about/cityofculture/researchresources/2021-pme-strategy-oct19-online.pdf>

⁹ Langen, Floris, and Beatriz Garcia. *Measuring the impacts of large scale cultural events: a literature review*. Liverpool: Impacts 08, 2009.

¹⁰ The UNESCO 2030 Cultural Indicators, <https://unesdoc.unesco.org/ark:/48223/pf0000371562>

¹¹ European Commission’s Cultural and Creative Cities Monitor, <https://composite-indicators.jrc.ec.europa.eu/cultural-creative-cities-monitor/>

Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world

AMPS, Architecture_MPS, Parade, American University in Dubai
Dubai: 22-24 November, 2020

expansion of devices connected to the internet”¹². Manovich describes a world where “billions of people create media, sharing them, and having access to trillions of media objects”¹³. Our global culture was already moving online before the COVID-19 pandemic struck; the rate of change sharply accelerated as a result of it.

Despite top-down strategies and requirements for evaluation, and with a new wealth of data being generated within the industry, cohesive data-driven initiatives are not being utilized from the bottom-up by most cultural and creative institutions. Bruce, Malcolm & O’Neill explain that “many cultural arts organisations generate value through the relationships they create and the networks they sustain, but far too often this data is not clearly articulated or evidenced to leverage insight, support and business opportunities”¹⁴. The institutions that made significant investments in upgrading their data infrastructure and making their labour pool more data literate while these seismic data trends were taking place would have survived, and even thrived, during the pandemic. For example, the National Theatre increased its YouTube subscriber count from 76,000 subscribers to 700,000 during the first lockdown in Britain in 2020, and received 10 million views on NT at Home in 6 weeks on YouTube compared to 10 years for NT Live, their own platform¹⁵. Initiatives such as Europeana “empowers the cultural heritage sector in its digital transformation”¹⁶. Yet for many, old legacy data systems, a lack of data literacy and even resistance to new technologies has severed handicapped data adoption, and the requirement to undertake evaluation has become a burden.

Clearly there has become an increased desire by the creative and culture industries to learn from their data, and advanced computational and data science methods offer an opportunity to analyse this data at scale. This has led to a term called *cultural data science* to appear on the international stage, and has even been turned into a university bachelors course at Aarhus University¹⁷. However, the term itself, theories and concepts of *cultural data science* are yet undefined and not clearly articulated for people to utilize it to its full potential. The industry needs a unifying set of frameworks, instructions and series of steps to explain how value can be extracted from cultural data by using advanced computational methods and how the results of their research can be used to answer policy questions and how data-driven insights can be used to defend against scrutiny and champion change.

HOW CULTURAL DATA SCIENCE IS DISTINCT

Before diving into how cultural data science can be applied by those in the creative and cultural industries, it’s important to first clarify what the term means. The best way to explain what cultural data science means is by explaining what it does not mean; *cultural data science* is not *cultural analytics*, much like *data science* is not *data analytics*. The nuance is very subtle, perhaps a matter of semantics to some people, but there are differences.

First, one must understand the dichotomy between *data analytics* and *data science*. Data analytics is often referred to as a “process” or “procedure” that transforms raw data into meaningful information to find patterns¹⁸, whereas data science, a term that still lacks consensus¹⁹, found its roots in academia and is often referred to as a “discipline” or a “study” to “extract knowledge from

¹² Fraser Bruce, Jackie Malcolm & Shaleph O’Neill (2017) Big Data:

Understanding how Creative Organisations Create and Sustain their Networks, The Design Journal, S436

¹³ Manovich, Lev. Cultural analytics. MIT Press, 2020, p25

¹⁴ Fraser Bruce, Jackie Malcolm & Shaleph O’Neill (2017) Big Data:

Understanding how Creative Organisations Create and Sustain their Networks, The Design Journal, S435

¹⁵ YouTube Spotlight UK. “National Theatre x YouTube.” September 22, 2020. Video, 1:35.

<https://youtu.be/2NQE1IcRPuI>.

¹⁶ Europeana, accessed 28 December 2020, <https://pro.europeana.eu/about-us/mission>

¹⁷ <https://bachelor.au.dk/en/supplementary-subject/culturaldatascience/>

¹⁸ Power, D.J., Heavin, C., McDermott, J. and Daly, M., 2018. Defining business analytics: an empirical approach. Journal of Business Analytics, 1(1).

¹⁹ Irizarry, R.A., 2020. The role of academia in data science education. Harvard Data Science Review, 2(1).

Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world

AMPS, Architecture_MPS, Parade, American University in Dubai
Dubai: 22-24 November, 2020

information”²⁰. Data science unifies “statistics, data analysis, machine learning and their related methods”²¹ to question and test hypotheses with data by writing algorithms and building statistical models. In fact, a Harvard Data Science Review paper written by Meng explains that data science is “best understood as a collection of disciplines with complementary foundations, perspectives, approaches, and aims, but with a shared grand mission”²². Furthermore, Irizarry extends Meng’s definition to describe data science as an “umbrella term” to describe “the entire complex and multistep processes used to extract value from data” with “real-world implications”²³.

The term “cultural analytics”, invented by Lev Manovich in 2005, is “the exploration and analysis of massive cultural data sets of visual material”. The way cultural analytics is framed in Manovich’s 2007 whitepaper is about “mining data and then visualizing data”, furthermore, Manovich (2015) explains that cultural analytics is interested in the “patterns that can be derived from the analysis of large cultural datasets”. Here we see that the term cultural analytics is very similar to that of data analytics in that it is a process of discovering insights and patterns from cultural data.

Although similar, cultural data science, on the other hand, aims to be grounded as a science, one that poses questions, discusses and applies interdisciplinary methodologies to test hypotheses and extract value from cultural data to understand and positively influence the underpinnings of human culture and society. Cultural analytics presently lacks frameworks for influencing policy and instead focuses on “analyzing cultural trends”. Cultural data science differs from cultural analytics in its focus on real-world, practical application, and hope to work alongside decision-makers and government agents to help guide cultural policy.

WHAT IS CULTURAL DATA SCIENCE

At present, there is no clear definition of cultural data science, therefore, this work presents the first definition of the field. Adapted from Irizarry²⁴, *cultural data science* is defined as the umbrella term used to describe the entire complex, multistep process used to extract value from cultural data with real-world implications to influence policy, placemaking and market dynamics within the creative and cultural industries. This process is initiated by the desire to answer a pertinent question within the creative and cultural industries using data and involves collecting and cleaning any form of cultural data that is relevant to the question, exploring and applying advanced computational methods to that data to test hypotheses, and presenting the findings of the study to stakeholders as well as devising new frameworks to influence policy and market dynamics within the industry. Figure 1 illustrates the seven-step *cultural data science* process to undertake research within the field.

Cultural data science is a unification of several academic fields (Figure 2) and hopes to respect and honour the ideas, methodologies and tools developed by these adjacent fields. In the broadest sense, cultural data science is at the union of arts & humanities, social science, computer science and mathematics; more specifically, it can be thought of as an extension of cultural analytics, urban studies and computational social science. Computational techniques are used to analyse both quantitative and qualitative data where machine learning and deep learning techniques can analyse text or visual data.

²⁰ Dhar, V., 2013. Data science and prediction. *Communications of the ACM*, 56(12), pp.64-73.

²¹ Hayashi, C., 1998. What is data science? Fundamental concepts and a heuristic example. In *Data science, classification, and related methods* (pp. 40-51). Springer, Tokyo.

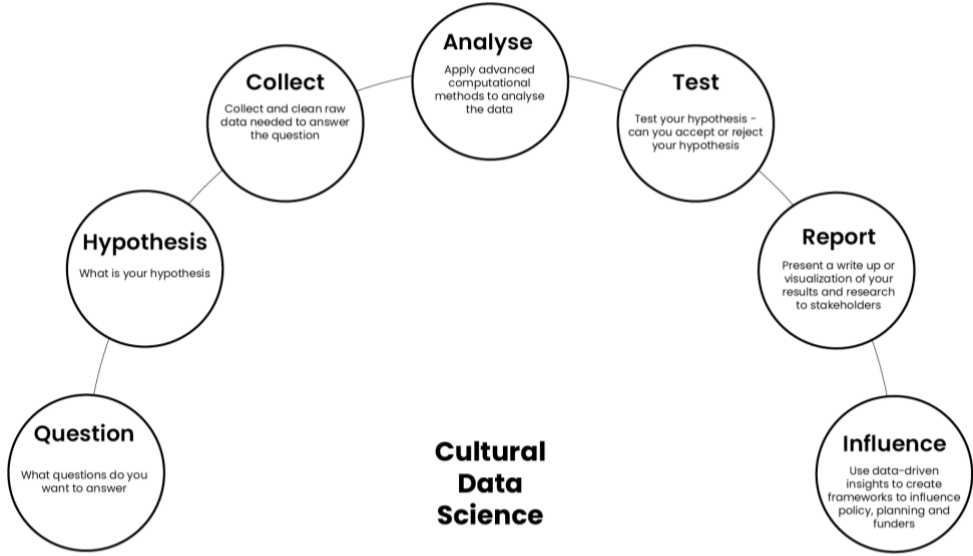
²² Meng, X.L., 2019. Data science: An artificial ecosystem. *Harvard Data Science Review*, 1(1).

²³ Irizarry, R.A., 2020. The role of academia in data science education. *Harvard Data Science Review*, 2(1).

²⁴ Irizarry, R.A., 2020. The role of academia in data science education. *Harvard Data Science Review*, 2(1).

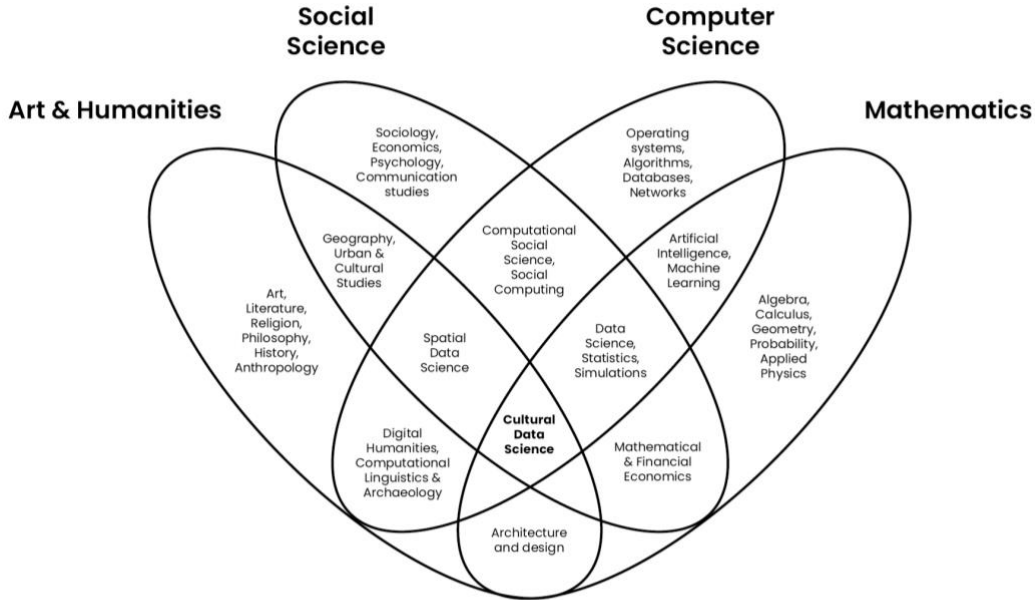
Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world

AMPS, Architecture_MPS, Parade, American University in Dubai
Dubai: 22-24 November, 2020



vishalkumar.london ©

Figure 1: illustrates a seven-step process to undertake cultural data science.



vishalkumar.london ©

Figure 2: A Venn diagram explaining cultural data science

Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world

AMPS, Architecture_MPS, Parade, American University in Dubai
Dubai: 22-24 November, 2020

HOW CAN CULTURAL DATA SCIENCE BE APPLIED

Cultural data science cannot be undertaken without cultural data. But what exactly is cultural data? Handwerker and Wozniak describe cultural data as those that “reflect the social (interactive) process by which we construct our knowledge of each other and the way these social processes work”²⁵. As our culture has moved more online, Manovich describes the four categories to explain our global digital culture that can be analyzed computationally at scale: media; behaviours; interactions; events²⁶. Kumar describes several categories of cultural data within the creative and cultural industries²⁷: historical cultural datasets - data about traditional cultural heritage and artefacts and the digitization of collections in museums and libraries, both visual data and textual data; contemporary cultural datasets - activity, user-created content and interactions on social media; art market transactions - data about art and cultural objects that are traded; socio-economic cultural datasets - data about our economies, preferences and societies that are usually collected using surveys; spatial cultural datasets - the location of cultural buildings, public spaces and public art; new forms of cultural data - data such as digital assets used in augmented and virtual reality.

Where does the cultural data come from? Institutions within the cultural and creative industries have their own proprietary data on audiences, tickets sales and cultural artifacts, and usually this data is stored in a data warehouse; many of these systems are old legacy data warehouses but some institutions have moved to the cloud²⁸. Some institutions such as The Metropolitan Museum of Art²⁹, Victoria and Albert Museum³⁰ and The Wellcome Collection³¹ have decided to open up these data to the public via an application programming interface (API). International institutions, national governments and city authorities are also collecting data and making them openly accessible through online portals. For example, portals by Arts Council England³², data.london³³, and make.open.data³⁴ offer a trove of publicly available data. Social media platforms and tech companies such as Twitter, Foursquare, Google Places and YouTube offer access to user-generated data through APIs. More recently, new mobile technologies and sensors are collecting data too³⁵.

Once cultural datasets have been collected, organised and cleaned, various computational methods can be applied to them to extract value and understand their economic, social, cultural, or environmental impact. First exploratory data analysis and the methods suggested by cultural analytics can help understand patterns and trends in the data. Then advanced computational methods can be applied to extract value from cultural data: statistics can be used to perform econometrics to understand the relationship between variables; spatial data science can undertake geographic-weighted regressions to understand the spatial relationship between variables; natural language processing can

²⁵ Handwerker, W. Penn, and Danielle F. Wozniak. "Sampling strategies for the collection of cultural data: an extension of Boas's answer to Galton's problem." *Current Anthropology* 38, no. 5 (1997): p870.

²⁶ Manovich, Lev. *Cultural analytics*. MIT Press, 2020, p75

²⁷ Kumar, Vishal 2020. What is Cultural Data?, accessed 5 January 2021, <https://medium.com/@vishalkumarlondon/what-is-cultural-data-89c18612b109>

²⁸ "Museum of London Protects Collection Databases and Advances Cloud Strategy with Rubrik", accessed 02 January 2021, <https://www.rubrik.com/press-release/museum-of-london-protects-collection-databases-and-advances-cloud-strategy-with-rubrik/>

²⁹ "The Metropolitan Museum of Art Collection API", accessed 02 January 2021, <https://metmuseum.github.io/>

³⁰ "Victoria and Albert Museum API Documentation", accessed 02 January 2021, <https://www.vam.ac.uk/api>

³¹ "Developers - Wellcome Collection", accessed 02 January 2021, <https://developers.wellcomecollection.org/>

³² Arts Council England, Our Data, accessed 02 January 2021, <https://www.artscouncil.org.uk/research-and-data/our-data>

³³ Data.London, accessed 02 January 2021, <https://data.london.gov.uk/topic/art-and-culture>

³⁴ Open Cultural Data Worldwide, accessed 02 January 2021, <https://make.opendata.ch/wiki/data:glam>

³⁵ Blanke, Ulf, Gerhard Tröster, Tobias Franke, and Paul Lukowicz. "Capturing crowd dynamics at large scale events using participatory gps-localization." In 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp. 1-7. IEEE, 2014.

Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world

AMPS, Architecture_MPS, Parade, American University in Dubai
Dubai: 22-24 November, 2020

execute linguistic analysis to learn from text; computer vision techniques can learn from images and videos; and, network science can understand how data is connected. Moreover, data visualizations can be used to present the findings of the research; data visualizations are increasingly being used to present ubiquitous data information³⁶.

The process of collecting cultural data and extracting value from that data by applying advanced computational methods to answer pertinent questions in the cultural and creative industries will lead to novel insights being drawn. When those insights are written up, reported and presented they can be used to have meaningful data-driven conversations and discussions with policy makers, planners and funders, which allows for arguments to be made about the importance of the cultural and creative industries to drive significant change. The aim of *cultural data science* is to assist purpose and data-driven collaboration between public and private organisations and has the potential to become increasingly important for stakeholders in the built environment and policy makers who want to apply novel frameworks and evidence-based methods to influence the artistic and cultural vibrancy of place, space, societies and our local economies. For example, if a city were to take a data-driven evidence-led approach to implementing a new cultural policy in combination with architects, developers, local residents and cultural institutions, the entire procedure can be assisted and guided by the methodologies, frameworks and toolkits from cultural data science.

It's important for institutions to clearly plan and articulate a cultural data science strategy before beginning any project. This allows for the right data collection, data infrastructure and data analysis techniques to be to be implemented and then applied for evaluation, and can lead to the creation of indicators or benchmarking metrics to compare the value of the project against others.

LIMITATIONS OF CULTURAL DATA SCIENCE

The enthusiasm for the "datafication" of culture and society is not shared by all. Helen Kennedy's Post, Mine, Repeat explains that research should be weary of the negative consequences of the desire to mine social media data because of its effects on civil liberty and data privacy. Ethical considerations should certainly be consulted when undertaking cultural data science, and the theory could be strengthened by providing best practices. Yet, innovative civilian data initiatives such as the data commons, where citizens come together to pool their personal data, are being tested and trialed in European Cities such as Barcelona and Helsinki and offer a glimpse of how data relations can become more equitable in the future leading to better quality data generated for research purposes³⁷. Additionally, there is a large extent to which cultural, creative and artist value cannot only be captured using objective methods³⁸, therefore, cultural data scientists must work with and alongside more subjective and emotional representations of value. There are also more practical limitations in terms of the lack of data literacy and data infrastructure within the creative and cultural industries. In the short term, it will be difficult to transition to a world that is completely data driven, but institutions should plan to be more data driven over time.

CONCLUSION

In conclusion this paper presented the first definition of the term *cultural data science*, an important step in genesis of a new field of research. Future research should dive deeper into exactly

³⁶ Living with data visualisation, <https://livingwithdata.org/previous-research/living-with-data-visualisation/>

³⁷ Calzada, Igor. "(Smart) citizens from data providers to decision-makers? The case study of Barcelona." *Sustainability* 10, no. 9 (2018): 3252.

³⁸ Rife, Mary Lou, D. King, S. Thomas, and R. Li. *Measuring Cultural Engagement: A Quest for New Terms, Tools, and Techniques: Summary of a Joint Research Symposium Titled "Measuring Cultural Engagement Amid Confounding Variables: A Reality Check,"* Held at the Gallup Headquarters in Washington, DC, June 2-3, 2014, Co-sponsored by the National Endowment for the Arts and the Arts & Humanities Research Council's Cultural Value Project. National endowment for the arts, 2014. page 1

Rapid Cities - Responsive Architectures: A conference examining design, planning & construction in the modern world

AMPS, Architecture_MPS, Parade, American University in Dubai
Dubai: 22-24 November, 2020

how these methods can be executed and what technologies can be used to perform cultural data science alongside relevant case studies. Finally, ethical considerations should be explored and best practices should be suggested.

BIBLIOGRAPHY

Blanke, Ulf, Gerhard Tröster, Tobias Franke, and Paul Lukowicz. "Capturing crowd dynamics at large scale events using participatory gps-localization." In 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp. 1-7. IEEE, 2014.

Bruce, Fraser, Jackie Malcolm, and Shaleph O'Neill. "Big data: Understanding how creative organisations create and sustain their networks." *The Design Journal* 20, no. sup1 (2017): S435-S443.

Calzada, Igor. "(Smart) citizens from data providers to decision-makers? The case study of Barcelona." *Sustainability* 10, no. 9 (2018): 3252.

Dhar, V., 2013. Data science and prediction. *Communications of the ACM*, 56(12), pp.64-73.

Handwerker, W. Penn, and Danielle F. Wozniak. "Sampling strategies for the collection of cultural data: an extension of Boas's answer to Galton's problem." *Current Anthropology* 38, no. 5 (1997): 869-875.

Hayashi, C., 1998. What is data science? Fundamental concepts and a heuristic example. In *Data science, classification, and related methods* (pp. 40-51). Springer, Tokyo.

Irizarry, R.A., 2020. The role of academia in data science education. *Harvard Data Science Review*, 2(1).

Langen, Floris, and Beatriz Garcia. *Measuring the impacts of large scale cultural events: a literature review*. Liverpool: Impacts 08, 2009.

Manovich, Lev. *Cultural analytics*. MIT Press, 2020, p75

Meng, X.L., 2019. Data science: An artificial ecosystem. *Harvard Data Science Review*, 1(1).

Montalto, Valentina, Carlos Jorge Tacao Moura, Sven Langedijk, and Michaela Saisana. "Culture counts: An empirical approach to measure the cultural and creative vitality of European cities." *Cities* 89 (2019): 167-185.

Power, D.J., Heavin, C., McDermott, J. and Daly, M., 2018. Defining business analytics: an empirical approach. *Journal of Business Analytics*, 1(1), pp.40-53.

Rife, Mary Lou, D. King, S. Thomas, and R. Li. *Measuring Cultural Engagement: A Quest for New Terms, Tools, and Techniques: Summary of a Joint Research Symposium Titled "Measuring Cultural Engagement Amid Confounding Variables: A Reality Check," Held at the Gallup Headquarters in Washington, DC, June 2-3, 2014, Co-sponsored by the National Endowment for the Arts and the Arts & Humanities Research Council's Cultural Value Project*. National endowment for the arts, 2014.