Archeologia e Calcolatori
35.2, 2024, 19-28
doi 10.19282/ac.35.2.2024.03

AMPLIFYING ARCHAEOLOGICAL OUTREACH: THE IMPACT OF WIKIPEDIA AND COLLABORATIVE INITIATIVES ON CULTURAL TOURISM

1. Preface

The advent of digital platforms has revolutionized the way information is disseminated, and Wikipedia has emerged as a pivotal tool in democratizing access to knowledge (Hinnosaar *et al.* 2021; Marwick, Smith 2021). In the realm of archaeology, leveraging such platforms can enhance public engagement and cultural understanding. In this paper, we outline the collaborative efforts undertaken to utilize Wikipedia as a tool for archaeological outreach. This narrative forms the first part of a comprehensive analysis, with the second segment dedicated to statistical analysis. The projects discussed herein are part of an ongoing initiative, and although the data are not yet complete, the initial insights have prompted us to present these preliminary findings. We will base our analysis on statistical theory to ensure a robust discussion.

P.G.

2. Introduction: Wikipedia as a catalyst for archaeological outreach

This research is part of broader collaborative efforts involving academic institutions and local government aimed at democratizing information and bolstering cultural tourism through digital means. Our engagement with Wikipedia began with a focus on the archaeological heritage of Verona, under the auspices of the Soprintendenza Archeologia, Belle Arti e Paesaggio per le Province di Verona, Rovigo e Vicenza (Superintendency VR-RO-VI) and the University of Verona. This project, initiated in 2017, aimed at enhancing the visibility and accessibility of archaeological sites through the creation of Wikipedia articles. Furthermore, the ongoing project combines the creation and enhancement of Wikipedia entries with the collection of site visitation data and employs statistical methods to analyze relationships and assess impact and symbiosis between them. This paper presents a work in progress of the efforts made and the results obtained so far.

2.1 *The Public Archaeology in Verona initiative (2017) and its outcomes*

The initial project was centered around Verona's Roman heritage, thanks to a collaboration among the University of Verona, the Superintendency VR-RO-VI and Wikimedia Italia. This initiative witnessed the creation of Wikipedia articles for key archaeological sites, notably Corte Sgarzerie. Within the

Fig. 1 – Visit to archaeological area of via San Cosimo, Verona (Wikipedia e Archaeology Project) (Patafisik, via Wikimedia Commons).

first year following these articles' inception, data indicated a 5.9% increase in visitor numbers, underscoring the project's immediate impact (Cenci, Grossi 2019). Even when we had an interruption of activities during the Covid-19 lockdown, the observation was repeated and an important increase in visits was found after 5 years.

2.2 *The Wikipedia and Archaeology GLAM Project (2023): expansion and engagement*

Building upon the success of the initial efforts, the Wikipedia and Archaeology GLAM Project (https://it.Wikipedia.org/wiki/Progetto:GLAM/Wikipedia_e_archeologia) was launched to further enhance and create Wikipedia articles and was extended to all the archaeological areas of the Superintendency VR-RO-VI, not only those of Verona. This phase was marked by a series of workshops, photographic campaigns, and guided tours, designed to engage the community and improve the quality and accuracy of the information provided on Wikipedia (Fig. 1). The project was promoted by Wikimedia Italia, with the approval and support of Superintendency VR-RO-VI.

2.3 *The Wikipedia and Archaeology across borders GLAM Project (2024): analyzing long-term and international impacts*

The ongoing project involves further translation of articles into multiple languages and continuous monitoring of site visits in relation to Wikipedia

articles views. This work aims to provide a more comprehensive understanding of how sustained digital outreach can influence both, public engagement with site visitation over a longer period and cultural awareness. It is important to note that we face challenges, particularly with the variability in data recording practices at different sites, which complicates long-term analysis. Future efforts will aim to standardize data collection methods and expand the geographical scope of the project to improve the reliability and applicability of our findings. The forthcoming analysis is part of this project. It incorporates a detailed statistical evaluation of the data collected throughout the course of the past projects. By combining these findings with the narrative outlined in this section of the article, we aim to provide a comprehensive overview of the impact of Wikipedia on public engagement with archaeological heritage.

P.G.

3. Context and hypothesis

Since the rise of accessible information, people have access to more data and knowledge on several topics, from hard science like quantum physics, to humanities like archaeology. With the explosion of free information immediately accessible on either computers or mobile devices people are able to know about what organizations, and institutions are doing, and promoting. This covers all demographics, even the older generations as shown in Chakraborty *et al.* 2013, in fact the rapid development of Internet information technologies has radically altered the landscape for collecting and processing information. The spheres in which information is available and shared are many, for example academic, economic or societal, as mentioned by Buhalis *et al.* 2011; Shamir *et al.* 2015; Griggio *et al.* 2019 among others. One of the fields rarely investigated is cultural tourism, for example visits to archaeological sites. In this section, we explore how accessible digital information impacts the physical visits to archaeological sites and vice versa, how visits can trigger further access to accessible information and hence increase cultural awareness and understanding, completing a cognitive circle of physical and digital knowledge. Our hypothesis is two-fold: on the one hand, the more digital information is viewed, the more this information influences decisions on visiting the physical sites and on the other hand, the physical visits to sites trigger the search for further information.

3.1 *Data*

The data was collected for two sites: 1) Corte Sgarzerie and 2) Criptoportico Romano Vicenza. The data about the Wikipedia views came from Wikipedia reports on pageviews from the Wikipedia article https://it.wikipedia.org/wiki/Corte_Sgarzerie (*Corte Sgarzerie, Digital* 2024) and

from the Wikipedia article https://it.wikipedia.org/wiki/Criptoportico_romano_(Vicenza) (*Criptoportico, Digital* 2024); the data about the site visits came from the official reports of the areas (*Corte Sgarzerie, Physical* 2024; *Criptoportico, Physical* 2024). The aggregation level was monthly for Corte Sgarzerie and annually for Criptoportico Romano. We also split the data on Wikipedia views between platforms (desktop and mobile) to enhance the analysis. Some of the data on site visits was incomplete, for example lack of data points for 2022 for Corte Sgarzerie, however there is a work in progress, which is part of the central objectives of the entire project, to have complete datasets for future analysis.

3.1.1 Methodology

Firstly we analyzed the distribution of data, and established its relevant parameters like mean, median, range, standard deviation and interquartiles. We employed the Jarque-Bera test to assess normality, and adhered to standard definitions for mean, median, and standard deviation. Our analysis involved identifying outliers and anomalies in the data, such as the impact of Covid-19 on visitation. Any data points falling outside three standard deviations from the mean were treated as outliers. Then we performed data engineering techniques and treated outliers and anomalies accordingly so we could withdraw results with statistical confidence. Anomalies and outliers were extracted from the main *corpus* of data. We also performed a paired t-test for statistical significance on the impact of digital information versus physical visits to sites by comparing visits when the Wikipedia articles were not existent and when they were up and consulted.

After calculating correlation coefficients between datasets, a behavioral pattern was established. To achieve this, we shifted the data by one period (lagged correlation) and formulated our hypotheses that: 1) Wikipedia views preceded visits to actual sites as implied by Boukas, 2008 and 2) that visits to sites triggered further curiosity and research on Wikipedia. It was used the standard Pearson Correlation Coefficient formula.

4. Findings

4.1 *Corte Sgarzerie*

For Corte Sgarzerie, we omitted the Covid-19 months when the site was closed, including the months from May to August 2021 which still had Covid-19 effects on visitors. Data passed the Jarque-Bera test so we assumed normality, however we also worked with median and interquartile ranges. We omitted the months when the site was closed, these were January 2014, August 2014, November 2014, and August 2015, and also found 1 outlier outside 3 standard deviations from the mean in 100 data points, which was

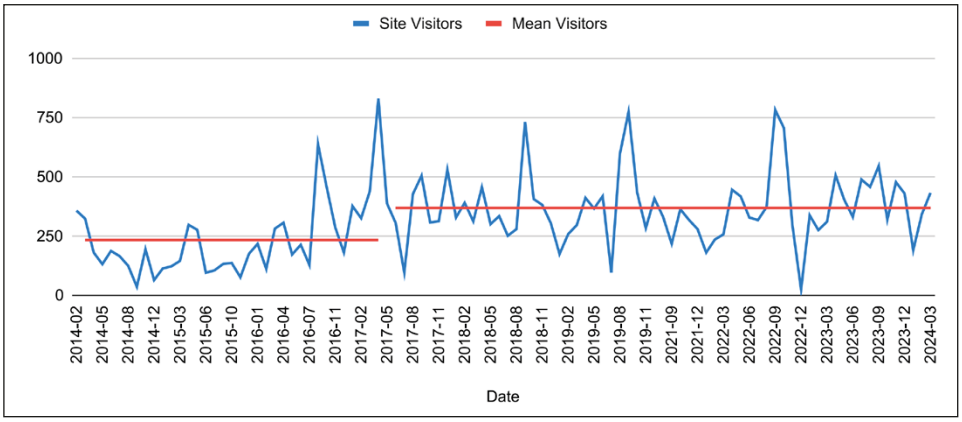

Fig. 2 – Site visitors and mean visitors before and after Wikipedia page was up for Corte Sgarzerie.

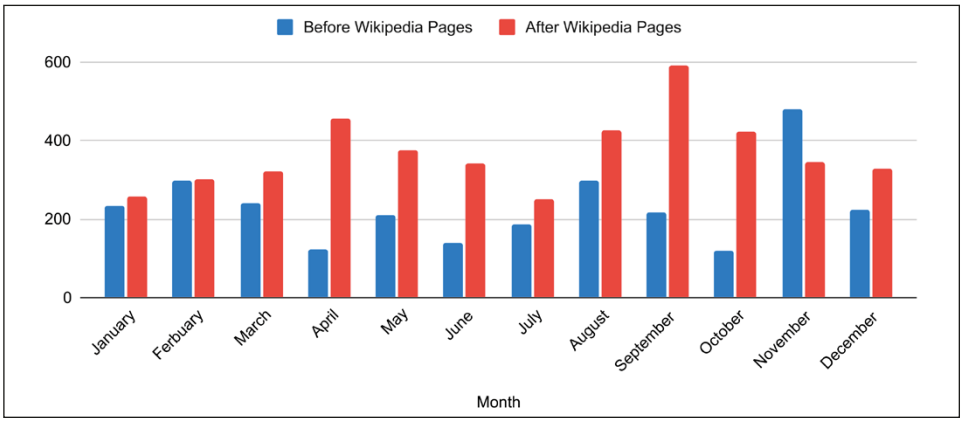

Fig. 3 – Site visitors before Wikipedia pages and after Wikipedia pages monthly comparison for Corte Sgarzerie, excluding the Covid-19 months.

omitted as it accounted for 1% of the total data. This outlier corresponded to March 2014 and was due to a FAI (Fondo Ambiente Italiano) event that spiked the number of visitors in that month. Finally we filtered the data to analyze only ordinary visits to the site. The paired t-test with H_0: $\mu_{before} = \mu_{after}$ versus H_1: $\mu_{before} < \mu_{after}$ was below the p-value 0.5 so we consider that the average visitors before and after is statistically meaningful. The average went from 233 visitors per month to 368 accounting to an increase of 58% (Fig. 2).

Moreover, analyzing month *versus* month the average percentage change found was up by 88%, being April the month with the highest

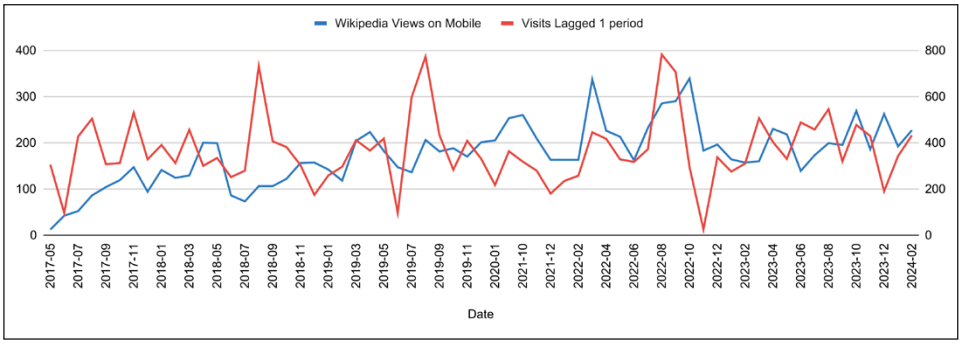

Fig. 4 – Wikipedia views on mobile and visits lagged by 1 period for Corte Sgarzerie.

uplift (Fig. 3). The correlation coefficient between total Wikipedia views and the 1-month shifted data was 0.14 which shows a medium causation compared to the correlation of 0.13 of the non-shifted data. An interesting finding was the stronger correlation between the platform used (desktop or mobile) and the number of visits to the actual site. Mobile users showed a correlation coefficient of 0.19 over all the period against the lagged data, which we consider medium-to-strong given the nature of the phenomenon (human behavior).

We also found that the use of mobile devices to access information surpassed the desktop users in general over the period (2014 to 2024) and 2022 registered the highest correlation between mobile users and visitors (lagged 1 period) of 0.55 and the total users (desktop and mobile together) had a correlation coefficient on that year of 0.28 (Fig. 4).

4.2 *Criptoportico Romano*

For Criptoportico Romano, with only 6 data points available due to annual data aggregation, our focus was solely on correlating Wikipedia visits with site visits. As school visits accounted to 38% of the total visits, we assumed that a lag of one period (one year) was appropriate for the analysis as not only educational coursework is due during the following period, but also school trips require a certain amount of logistics that normally happens in the previous quadrimester. The Wikipedia article for this site has been up and running during all the years of the analysis (2018-2023). What we found is that the correlation coefficient assuming this 1-period lag was 0.72 (perfect correlation is 1), so there is a strong correlation, especially given the analysis of human behavior (Fig. 5).

It is clear from the chart that the behavior is similar, leading us not only to correlate but establish causation between the two (one causes or

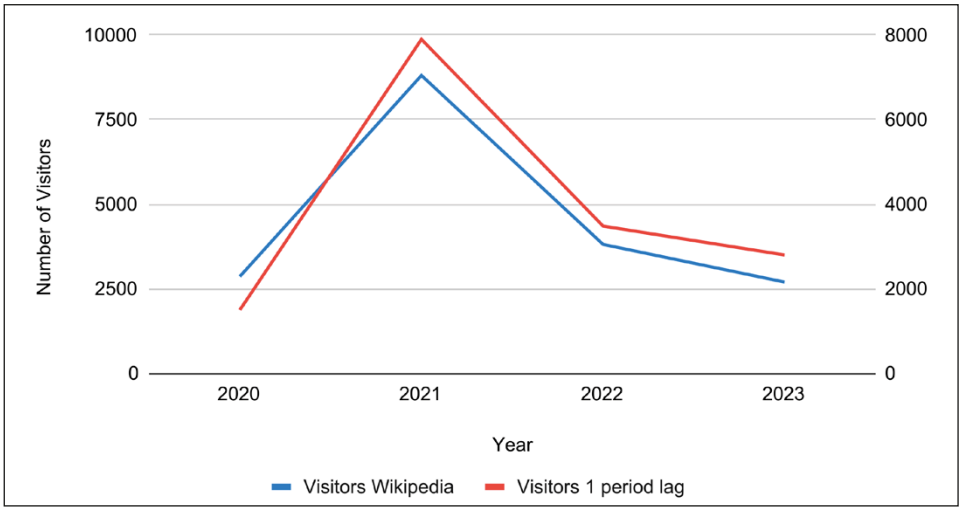

Fig. 5 – Total Wikipedia visitors *versus* site visitors lagged by 1 period for Criptoportico Romano.

strongly influences the other). For this we assume the principle of causal analysis in statistics which involves four elements: 1) correlation, 2) sequence in time (causes must occur before their proposed effect), 3) a plausible mechanism for an observed effect to follow from a possible cause, and 4) eliminating the possibility of alternative causes (https://en.wikipedia.org/wiki/Causal_analysis). We cover all the four elements in analysis addressing the points as follows: 1) calculating correlation coefficient, 2) moving the data 1 period to establish a time sequence, 3) a fair amount of visitors are schools, being a plausible cause influencing the time-sequence and the behavior, and 4) we did not find any special events that could have influenced the behavior otherwise.

S.P.

5. Conclusions

Open access plays a pivotal role in the dissemination of cultural heritage knowledge, with Wikipedia standing out as a key tool for such endeavors. In Italy, several projects aim to harness Wikipedia to promote cultural heritage awareness. All of them are significant, however for conciseness, we will only mention 'Tutti i Musei su Wikipedia' (https://www.wikimedia.it/musei), conducted by Wikimedia and ICOM. Through the described project, Wikipedia has proven to be a powerful tool for archaeological outreach, significantly affecting both educational access and cultural tourism. As a

conclusion, we can see that open access information from Wikipedia influenced meaningfully the number of visits to the actual site at Corte Sgarzerie: the average monthly visitors before the Wikipedia article was created was 233 versus 368 which was the average monthly visitors after the Wikipedia article was up, which scored a percentage change of 58%. The difference month to month from before and after the Wikipedia articles were created is consistent with the previous result, being the highest differences in the months of April, September and October for Corte Sgarzerie. Following the widely used Causal Analysis in statistics, there is a strong correlation of 0.55 between the 1-period lagged data and the number of Wikipedia views on mobile devices in 2022 for Corte Sgarzerie and 0.28 between the same lagged data and total views on all platforms for 2022, while the correlation for all the period was 0.14 confirming this amount of causation between the views and the visits to the site. In the case of Criptoportico Romano, the correlation is strong (coefficient was 0.72), confirming the assumption of causation. Moreover because of the 1-period lag correlation we can assume that a number of visits to the actual site triggered further research on Wikipedia articles.

To sum up, these preliminary results demonstrate the potential of collaborative digital initiatives in fostering a deeper public connection with cultural heritage, highlighting the role of free access to information in promoting local and historical awareness.

P.G., S.P.

Piergiovanna Grossi
Dipartimento di Lingue e Letterature Straniere
Università degli Studi di Verona
piergiovanna.grossi@univr.it

Serge Plata
Infosys Consulting
serge.plata@infosysconsulting.com

Acknowledgements

We extend our gratitude primarily to Claudia Cenci, with whom the project was started in 2017, to Brunella Bruno, Giovanna Falezza, Giulia Pelucchini of the Superintendency VR-RO-VI, Annarosa Tricomi, Marzia Bersani of Archeonaute Onlus, Manuela Mantiero of Museo Diocesano Vicenza, for their unwavering support and significant contributions to this research. Their cooperation has been indispensable in the creation of the articles and collection and analysis of data, and their dedication greatly enhances the scope and impact of our work. A heartfelt thanks also goes to the archaeologists who provided guided tours and bibliographies, to Federico Benvenuti, Marco Chemello, and all the other contributors of Wikipedia who collaborated on the Project, that for brevity we can't mention here.

REFERENCES

Boukas N. 2008, *Cultural tourism, young people and destination perception: A case study of Delphi, Greece*, Thesis in Management Studies, University of Exeter (https://www.proquest.com/openview/c3d3732d8cf6c57204db302a31da144d/1?pq-origsite=gscholar&cbl=51922&diss=y).

Buhalis D., Leung D., Law R. 2011, *eTourism: Critical information and communication technologies for tourism destinations*, in Y. Wang, A. Pizam (eds.), *Tourism Destination Marketing and Management: Collaborative Strategies*, Wallingford, CABI Publishing, chapter 13.

Cenci C., Grossi P. 2018, *Public archaeology a Verona: un percorso sperimentale lungo la via Postumia*, in P. Basso, B. Bruno, C. Cenci, P. Grossi (eds.), *Verona e le sue strade. Archeologia e valorizzazione*, Verona, Cierre Edizioni, 223-232.

Chakraborty R., Vishik C., Rao R. 2013, *Privacy preserving actions of older adults on social media: Exploring the behavior of opting out of information sharing*, «Decision Support Systems», 55, 4, 948-956.

Corte Sgarzerie, Digital 2024, *Wikipedia page views reports* (based on visits to https://it.wikipedia.org/wiki/Corte_Sgarzerie, https://pageviews.wmcloud.org/?project=it.wikipedia.org&platform=all-access&agent=all-agents&redirects=0&start=2017-04&end=2024-03&pages=Corte_Sgarzerie).

Corte Sgarzerie, Physical 2024, *Area archeologica di Corte Sgarzerie, Verona. Dossier eventi e presenze*, Report 2014-2024, Archeonaute Onlus, Soprintendenza ABAP per le Province di Verona, Rovigo e Vicenza.

Criptoportico, Digital 2024, *Wikipedia page views reports* (based on visits to https://it.wikipedia.org/wiki/Criptoportico_romano_(Vicenza), https://pageviews.wmcloud.org/?project=it.wikipedia.org&platform=mobile-web&agent=all-agents&redirects=0&start=2015-07&end=2024-03&pages=Criptoportico_romano_(Vicenza).

Criptoportico, Physical 2024, *Criptoportico Romano, Vicenza, Monitoraggio visite*, Report 2018-2023, Museo Diocesano di Vicenza, Soprintendenza ABAP per le Province di Verona, Rovigo e Vicenza.

Griggio C., Midas Nouwens M., McGrenere J., Mackay E. 2019, *Augmenting couples' communication with lifelines: Shared timelines of mixed contextual information*, in S. Brewster, G. Fitzpatrick (eds.), *CHI 19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow 2019)*, New York, Association for Computing Machinery, Paper n. 623, 1-13 (https://doi.org/10.1145/3290605.3300853).

Hinnosaar M., Hinnosaar T., Kummer M., Slivko O. 2021, *Wikipedia matters*, «Journal of Economics & Management Strategy», 32, 3, 1-13 (https://doi.org/10.1111/jems.12421).

Marwick B., Smith P. 2021, *World Heritage sites on Wikipedia: Cultural heritage activism in a context of constrained agency*, «Big Data & Society», January-June, 1-19 (https://journals.sagepub.com/doi/full/10.1177/20539517211017304).

Shamir N., Shin H. 2015, *Public forecast information sharing in a market with competing supply chains*, «Management Science», 62, 10 (https://doi.org/10.1287/mnsc.2015.2261).

ABSTRACT

This paper examines the impact of Wikipedia and Wikimedia Foundation projects on archaeological heritage dissemination and cultural tourism enhancement. It focuses on two initiatives: a 2017 Public Archaeology event and a 2023 project extending into 2024. The 2017 initiative, in collaboration with the University of Verona and the Soprintendenza Archeologia, Belle Arti e Paesaggio per le Province di Verona, Rovigo e Vicenza, involved creating

Wikipedia entries for key archaeological sites in Verona. Visitor data analysis, from these sites, showed an increased number of visits after the entry creation. The ongoing 2023 project, backed by Wikimedia Italia and the Superintendency, aims to refine Wikipedia entries for all supervised sites. Preliminary assessments indicate enhanced visitor engagement correlating with article updates. The study particularly analyzes visitor data for Corte Sgarzerie (Verona) and Criptoportico Romano (Vicenza), assessing both electronic and physical visits. Results indicate a significant statistical increase in site visits, following the availability of Wikipedia articles. Further, a significant correlation was found between electronic access to articles and actual visits to these sites. The findings suggest a substantial impact of Wikipedia on cultural tourism, highlighting the role of open access, free software, and open data in promoting and preserving knowledge.