Journal of Agribusiness and Rural Development

THE APPLICATION OF WEB ANALYTICS BY OWNERS OF RURAL TOURISM FACILITIES IN POLAND – DIAGNOSIS AND AN ATTEMPT AT A MEASUREMENT

Karol Król✉

University of Agriculture in Kraków, Poland

Abstract. Data analytics changes the way the enterprises operate, and allows them to both discover new business opportunities and offer innovative products and services. Data analytics can also be used to boost the performance of rural businesses. Therefore, the purpose of this study is to identify web analytics tools used by owners of rural tourism establishments in Poland. The study covered 965 websites hosted on selected paid domains. SEOptimer, an online application, and Google Tag Assistant were used in order to identify the analytical tools. Findings from the study were considered in the context of search engine optimization level and of the scale of business operations. In the sample covered by this study, a form of web analytics was found in 449 websites (46.5%); however, a data collection engine was probably implemented in 425 websites (44%). Google Analytics was the most frequent analytical tool. It was demonstrated that rural tourism establishments used analytical tools to a lesser extent than other businesses which rely on the Internet to promote and sell their products. Moreover, owners of higher-standard rural tourism establishments are more keen to use web analytics.

Keywords: data analytics, web analytics, rural tourism, optimization, Google Analytics, digital marketing

INTRODUCTION

Rural tourism establishments, especially those active in touristically attractive areas, face strong competition and therefore need to compete for effective sales of products and services. This would be impossible without promotional activities, in particular those carried out on the Internet. In order to reach the potential customer with their offer, owners of rural tourism establishments most frequently use individual websites, social media and sector portals. In most cases, this involves certain costs which primarily include hosting and domain subscription fees. However, only few of them analyze the cost efficiency of their online promotion efforts. Meanwhile, this can be done in two different ways, either by directly asking the customers what they find effective (in the case of small-scale operations) or by using web analytics, i.e. a particular type of data analysis (Król, 2017).

Data analytics allows information to be obtained and provide a basis for decision-making at the right time. Agility in using analytics and integrating data into business processes enables the creation of a long-lasting competitive edge. Analytics allows the data potential to be unleashed (Tarhan et al., 2016; Grossman, 2018).

PURPOSE AND GOALS OF THE STUDY

In the digital transformation era, the purchasing behavior and the way purchase decisions are made differ from what was experienced in the past. Consumers who are in contact with "partner brands" more and more often expect an immediate and adequate response, i.e. a reply or an offer. As a consequence, innovative enterprises

✉Karol Król, PhD, Department of Land Management and Landscape Architecture, University of Agriculture in Kraków, ul. Balicka 253c, 30-149 Kraków, Poland, e-mail: k.krol@onet.com.pl, https://orcid.org/0000-0003-0534-8471

"no longer interfere in what people are interested in, and make efforts to become the very thing what they are interested in." What has become crucial is the ability for the organizations to monitor customer behavior in real time at all contact points. Other vital aspects consist in analyzing the context and patterns of customer behavior and selecting appropriate actions at the right time. This is referred to as a "perishable insight" that can only be discovered and made actionable in real time. These needs are being met by broadly defined data analytics, including web analytics.

Due to the growing availability of new, intuitive analytical techniques and tools, no enterprise is "too small" or "too young" for analytics. Small- and medium-sized rural enterprises, including those promoting and selling goods or services online, can benefit from the development of analytical technologies, too. A research report by the Online Industry Employers' IAB Poland (Związek Pracodawców Branży Internetowej IAB Polska) and PwC points out to the relatively high popularity of e-commerce among rural entrepreneurs. Indeed, 41% of respondents sold goods and services online, usually through an individual website or other online sites such as auction sites or e-commerce platforms (Głąb, 2014).

It has been observed that most studies focus on the economic and organizational effects of implementing enterprise analytics systems, and that organizations are assessed for analytical maturity (Alles, 2018). While corporate entities are the usual subject of research, less attention is paid to rural entrepreneurs. This study attempts to fill this gap, and intends to identify the analytical tools used in the monitoring of Polish rural tourism websites. The study raised the following research questions: (1) is there a correlation between the scale of operations, expressed by the number of beds, and the use of web analytics; and (2) is data analytics used by owners of Polish rural tourism establishments? An attempt was also made to investigate the relationships between the quality of a rural tourism website, expressed by the level of search engine optimization (SEO), and the use of web analytics.

DATA ANALYTICS

Generally, analytics means a broad use of data, statistical and quantitative analyses, explanatory and predictive models, and fact-based management in decision-making and measure-taking processes (Davenport and Harris, 2010). Analytics comprises a technology, tools, a strategy, and specific goals, and is a customized solution as it requires that research methods and tools be adapted, and that business-specific measures be defined (Kaushik, 2010; Barton and Court, 2012).

Traditional analytics: Analytics 1.0

It was assumed that the Analytics 1.0 era started in mid-1950s when UPS initiated the first analytical group in the US, and lasted until mid-2000s when Internet companies began to introduce innovations based on data analysis. Analytics 1.0 was characterized by: (1) relatively small, structured sources of data mostly derived from internal systems; (2) most analytical activities being focused on descriptive analytics or reporting; (3) the development of analytical models being tedious, slow, and time-consuming; (4) a team of quantitative analysts being usually distanced from the decision-making center; (5) few organizations "competing using analytics," as data analysis played a marginal role in relation to the strategy; and (6) decisions being made based on experience and intuition. In the Analytics 1.0 era, over 90% of analytical activity involved descriptive analytics or a form of reporting. Data was mainly derived from internal transaction systems, and referred to well-known domains such as customer or product information. Reporting processes focused exclusively on the past, with no forecasts being made. The analysts were spending most of their time preparing the data for analysis. Meanwhile, little time was spent on the (mostly quantitative) analysis itself. All of this resulted in decisions being made based on experience and intuition (Davenport, 2013).

Analytics 2.0: Big Data

In the mid-2000s, the world began to pay attention to big data, thus initiating the Analytics 2.0 era. That period began with the use of fast-moving data sourced online. The Analytics 2.0 era is in progress, and involves a great deal of dispersed data. The data is frequently derived from external sources, in particular from the Internet, observations of human behavior, and various types of sensors. The new generation of quantitative analysts have begun to be referred to as "data scientists," and have become an integral part of product development teams. In the Analytics 2.0 era, freeware emerged that increased the availability and popularity of analytics. The ethos of Analytics 2.0 is reflected in the following phrases: "agile is too slow," "develop products, not

PowerPoints or reports," "information, hardware and software want to be free," and "share your big data tools with the community" (Davenport, 2013).

Analytics 3.0: Fast Impact for the Data Economy

Until recently, data analytics most frequently relied on predictive models. Currently, however, increased use is made of anomaly detection algorithms and real-time alert algorithms (Wiecka, 2018b). Large organizations are entering the Analytics 3.0 era which combines the best features of Analytics 1.0 and 2.0, i.e. large databases and traditional analytical tools. In the Analytics 1.0 era, enterprises used data warehouses and copies of operational data as a basis for analyses. In the Analytics 2.0 era, focus was on Hadoop clusters and NoSQL databases. In Analytics 3.0, new agile analytical methods and machine learning techniques have ensured a significantly quicker insight into data (Davenport, 2013).

WEB ANALYTICS

The development of (information) technologies enabled the acquisition of information on customer preferences based on an analysis of their current online behavior. The measurement of online users' activity has a great cognitive potential. As more and more new data analysis techniques and tools become available, website owners are in a better position to "hear what the users say," i.e. to analyze the user/website interactions (Kaushik, 2007; Jansen, 2009). Web analytics enables hypothesis testing, forecasting, and inference, and is meant to support decision-making processes and to discover new relationships. Web analytics can provide information on how to optimize a website, how to target an offer, and what areas require that activities be intensified and marketing expenditure be increased.

An outline of the history of web analytics

The analysis of online data was initially based on server logs which mainly contained technical rather than business information. Therefore, the analyses focused on clickstreams and such indicators as the number of hits or visits. Kaushik (2010) described such activities as the "world of reports" in which decision-makers were flooded with a great number of statistics and statements. These, however, provided no answers and solved no problems. "New web analytics" leads to practical conclusions, and goes beyond the analysis of clickstreams. Data describing user behavior and providing answers to specific questions, e.g. "where is the purchase process most often interrupted?," has become much more significant. The qualitative analysis also gained in importance: why do visitors behave in a certain way, and why do they use the website?

The first mentions of web analytics date back to early 1990s when server logs were observed to store information on user visits (hits), such as IP address, browser ID, or operating system name (Kaushik, 2007). Over time, server logs collected more and more data, which necessitated a graphical presentation. Developed in 1995, Analog is one of the first web analytics programs. It enabled the analysis of logs and processed data, and divided them according to specific metrics. Reports and statements have become more comprehensible and accessible for an average user.

In 1995–1996, web statistics became widespread owing to commonly available counters which usually presented the number of hits or visits (the company Web-Counter has significantly contributed to their dissemination). A commercial analysis of web data gained in popularity a few years later, largely owing to WebTrends software which enabled data reporting and visualization based on tables and diagrams. This has led to the use of web analytics data in decision-making processes. After 2000, server logs were replaced as data sources by the JavaScript tracking code, a technology which enabled the development of the market for operators engaged strictly in web analytics (Ołdak, 2010).

In 2005, Google released Google Analytics (GA), a milestone in the dissemination of web analytics based on the Urchin project. Soon, it gained great popularity worldwide, primarily due to its usability and unlimited availability. Alternatives to Google Analytics include the Matomo (Piwik) platform, HotJar, GoStats.pl, and stat.4u.pl.

MATERIAL AND METHODS

This study focuses on websites of Polish rural tourism establishments. Website addresses were retrieved from a directory delivered by agroturystyka.pl and the website of the Polish Federation of Rural Tourism "Hospitable Farms" (Polska Federacja Turystyki Wiejskiej "Gospodarstwa Gościnne"). The addresses were also retrieved from the Google search engine, based on organic search.

Table 1. Tools used in the website research

Tool	Tool type and aspects measured	Measurement scale
SEOptimer (online)	Identification of analytical tools implemented in the website code.	N/A
Google Tag Assistant (Google Chrome extension)	Identification of the Google Analytics tracking code; verification whether other Google tools are implemented in the website code.	N/A
SEO Analyzer (online)	Measurement of the search engine optimization (SEO) level. SEO indicator: SEO Score.	0–100
Website quality (online)	Measurement of the search engine optimization (SEO) level. SEO indicator: Website Grader Score.	0–100
Sunspot SEO Audit (online)	Measurement of the search engine optimization (SEO) level. SEO indicator: Overall SEO assessment.	0–170

Source: own elaboration.

The Google search engine was selected because of its great popularity worldwide (Dergiades et al., 2018). The study covered paid domain websites, including national (.pl, a country code top-level domain), European, commercial, regional, and other websites hosted at subscription domains.

In most cases, analytical tools are embedded in the website structure using the tagging procedure which is among the simplest and most popular methods, and consists in inserting a JavaScript tracking code to the website code (Peterson, 2004). As for Google Analytics, the tracking code contains a unique identifier, such as UA-000000-2. This is how a website is tagged; a reference to the script is used in the statistics system offered by the service provider. Since the basic structure of the tracking code is the same, it is possible to tell whether an analytical tool is implemented in the website framework.

SEOptimer, an online application (the module available in the free version), was used in order to identify the analytical tools used. Its capabilities include retrieving information on components that extend the functionality and interactivity of a website. Data retrieved in this way was verified using the Google Tag Assistant which analyzes the website code for Google components. The tests were conducted in Google Chrome (v70.0.3538.110).

In the next stage of the study, a SEO audit was conducted for the websites using three independent testing tools (Table 1). Consequently, each website was described using three SEO indicators expressed by a score measured on different scales. The final assessment of the optimization degree of each website was based on the unitarization method (normalization of variables) which allowed each website to be described with a cumulated synthetic score (F-Score). Moreover, the number of beds available at each establishment was verified when browsing the websites.

The variables were normalized using the zero unitarization method. The purpose of normalization is to make diagnostic variables usable as sub-criteria in assessing complex phenomena. Zero unitarization allows the numerical ranges of variables to be unified, and their absolute values to be converted into relative values (Król, 2019). The search engine optimization level is a variable with a stimulating effect, and its highest possible value is desirable (expected). Therefore, the following formula (1) was used in the unitarization process:

$$x'_{ij} = \frac{x_{ij} - \min_i\{x_{ij}\}}{\max_i\{x_{ij}\} - \min_i\{x_{ij}\}} \tag{1}$$

$i = 1, 2, ..., n; j = 1, 2, ..., m$
x'_{ij} – standardized diagnostic variable
x_{ij} – non-standardized diagnostic variable
$\min_i\{x_{ij}\}$ – minimum value of the non-standardized diagnostic variable
$\max_i\{x_{ij}\}$ – maximum value of the non-standardized diagnostic variable

The values of normalized variables varied in the range of 0 to 1. The F-Score index was the result of adding up the normalized values of individual measurements, and therefore could range from 0 to 3.

RESULTS

The study investigated 965 websites of rural tourism establishments. The Google Analytics (GA) tracking code was found in 415 websites (Table 2). A different type of tracking code was found in 34 websites. In total, the tracking code was identified in the structure of 449 websites (46.5% of the study sample). However, as shown by an in-depth qualitative analysis, 5 websites had a tracking code of the statistics system which expired in January 2018, whereas a critical error reported with a "No HTTP response detected" message was found in 19 installations of the GA tracking code. This means that the Google Analytics code had not been correctly implemented in the website, and that data had not been collected. In addition, in 10 cases, warnings about improper GA tracking code implementation were identified, which could have affected the quality of data collection. Therefore, although the tracking code was identified in 449 websites, data was probably collected only for 425 of them (44%).

A visit counter was found in 65 websites which accounted for less than 7% of the total sample, proving that this form of statistics is currently seldom used. Other Google tools were identified in 43 websites (approx. 4.5%), namely: Global Site Tag (which enables the measurement of micro-conversions, thus enhancing the spectrum of analyses available in GA); Google Remarketing Tag (which enables the launching of remarketing campaigns in an AdWords account); and Google Ads conversion tracking (which enables monitoring the performance of Google Ads). Owners of these websites were able to use more sophisticated forms of digital marketing.

Websites covered by this study reached a total score of 1519 in the SEO assessment, which is ca. 52.5% of the maximum score attainable. Most of them achieved an average score (1.5 to 2) (Table 3). This means that many of the websites under study required optimization, and had high untapped marketing potential and goal conversion potential.

In the study sample, a large part (ca. 36%) of rural tourism establishments offered from 10 to 19 beds. There were also many establishments (ca. 34%) which offered from 20 to 99 beds. Every fifth website failed to provide information on the number of beds offered to visitors (Fig. 1).

A medium strength of relationships (according to the Guilford scale) was noted between the level of search engine optimization (expressed by the F-Score index) and the occurrence of the GA tracking code ($r = 0.419$, $n = 965$, $\alpha = 0.05$, $r_2 = 0.175$, $p < \alpha$). This means that it was more frequent for websites with a higher F-Score to have a GA tracking code implemented. Moreover, owners of rural tourism establishments with higher-quality websites could have been more interested in the effectiveness of GA (the number of visits and hits, goal conversion, etc.). After other analytical systems (less popular than the GA), e.g. stat.4u.pl, were included in the calculations, the correlation decreased ($r = 3.88$). This

Table 2. The number and percentage of websites where a tracking code was discovered as one of web analytics systems

Web analytics system	Google Analytics	stat24.com	stat.4u.pl	Matomo (Piwik)	Other
Number of websites	415	5	21	3	5
Percentage (%)	92.43	1.11	4.68	0.67	1.11

Source: own elaboration.

Table 3. Number and percentage of websites according to F-Score

Qualitative range	0.5–1.0	1.001–1.5	1.501–2.0	2.001–2.5	2.501–3.000	Total
Number of websites	33	378	449	103	2	965
Percentage (%)	3.42	39.17	46.53	10.67	0.21	100

Source: own elaboration.
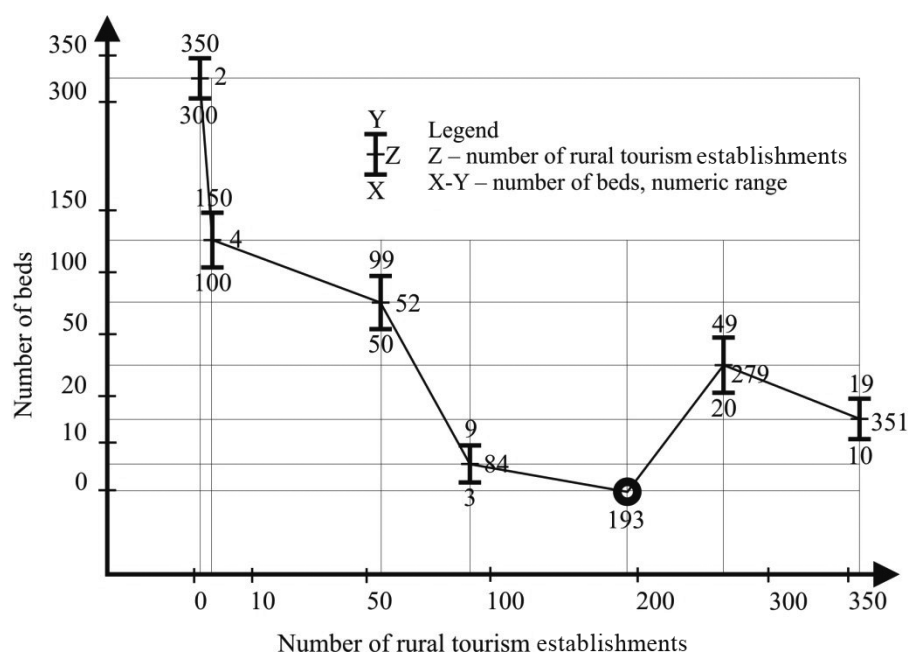
Fig. 1. Rural tourism establishments broken down by the number of beds
Source: own elaboration.

means that less popular statistics systems were used in websites with a lower F-Score.

No linear relationships were noted between the scale of business operations (expressed by the number of beds) and the occurrence of the GA tracking code ($r = 0.009$, $n = 965$, $\alpha = 0.05$, $r_2 = 0.009$, $p < \alpha$). This shows that the GA tracking code was implemented in websites run by both large- and small-sized rural tourism establishments. No linear relationship was noted between the scale of business operations and website quality (expressed with the F-Score) either.

DISCUSSION

Due to the easy access to data, software, and analytical methods, data analytics are used increasingly widely. This was confirmed in research conducted by the International Institute for Analytics. Relatively young and small companies from the Digital Natives group have often succeeded thanks to the innovativeness supported by the widespread application of data analytics. This is explained by data-based decision-making and by the democratization of analytics which consists in making

data and analytical software available to all employees (Wiecka, 2018a).

Business entities rely on analytics to various degrees, and the effectiveness of an enterprise increases depending on the extent to which it uses data in its decision-making processes (Brynjolfsson et al., 2011). An organization's analytical maturity can be assessed in various ways, including the AMA (Analytics Maturity Assessment) method and the DELTA model. A study conducted with a group of 515 respondents from 74 enterprises showed that high analytical maturity was strongly correlated with financial performance, reputation, and the power and innovativeness of a brand (Alles, 2018). In turn, periodical studies by NewVantage Partners showed that both the declared sense of urgency of the effective use of data and the conviction about the value to be provided by big data projects were on the increase. As many as 73% respondents stated that such initiatives provided them with a measurable benefit. Another study showed that the companies which used analytics were 5% more productive and 6% more profitable than ones who did not use it (McAfee and Brynjolfsson, 2012).

Business process automation is usually associated with replacing humans by machines, which can be expected to result in both a reduction in the number of jobs and an increase in the unemployment rate. This stereotype is contradicted by the results of "Analytics as a Source of Business Innovation," a study conducted with a group of over 2,500 respondents from all over the world. The study showed that a significant percentage of companies either automated (36%) of extended the functionality (41%) of business processes with the use of analytics. At the same time, 34% of them witnessed the emergence of new tasks for employees as a result of the use of analytics, while 16% of them admitted that certain previously automated tasks were assigned to employees (Wilson et al., 2017). Also, analytical enterprises experience the shortage of analytics experts capable of fully unleashing the potential behind data.

According to a study conducted by IRCenter (2015) among readers of online marketing portals, the largest group of analytical tools users was that of marketers and accounted for 43% of respondents. Marketers most frequently relied on web analytics to: monitor and report the performance of websites or web applications; optimize content; and assess the effectiveness of Google AdWords campaigns. Most respondents (61%) intended to spend more time using analytical tools which, in their opinion, would improve the effectiveness of measures taken. In a study conducted by Analyzeo (2015), nearly 92% of respondents (representatives of the e-commerce sector representatives) declared that they used web analytics to measure the effectiveness of their company websites. Just under 87% of them relied on specialized tools, of which 82% used Google Analytics.

CONCLUSIONS

A widespread use of data analytics, including web analytics, among owners of rural tourism establishments in Poland appears to be a distant prospect. Most rural tourism establishments are small- and medium-sized, often family-run enterprises. Many of them, especially agritourism farms, operate on a small scale. Their owners are usually not interested in maximizing the occupancy rate. Instead, they provide services to regular customers or attract new ones through word of mouth. Such operators may be not interested in data analytics, unlike large-size hotel chains and corporations focused on optimizing the effectiveness and efficiency of processes

and on profit maximization, market expansion and development.

This study allowed to assess only the potential behind the actual implementation of analytical tools. Irrespective of the involvement of the website owner, the proper implementation of the tracking script is a way to ensure that data is collected. However, this does not necessarily mean that data is put into use, i.e. viewed, analyzed, and relied upon in the decision-making process.

It is worth implementing an analytical tool at design stage of a website, or incorporating it in the structure of an existing one. There is a need to include web analytics in computer courses offered by agricultural advisory centers and other organizations educating farmers who are willing to use the Internet in promotional and sales activities.

Rural tourism establishments use analytical tools to a lesser extent than other businesses which rely on the Internet to promote and sell their products. This is understandable in the case of operators who benefit from tax reliefs. They usually regard tourism activities as a source of additional income which also satisfies the need for self-fulfillment, sharing their passion, interacting with people, and establishing social relations. On the other hand, the use of web analytics by larger rural tourism operators could provide information on the downsides of websites which, if overcome (optimized), could contribute to goal conversion. This can translate into more customers being attracted through the Internet, and into a constant (better) inflow of customers. Also, it can accelerate the return on amounts spent on online promotion of the business offer.

REFERENCES

Alles, D. (2018). Analytics maturity powers company performance. Portland, Oregon, US: International Institute for Analytics.

Analyzeo (2015). Analityka internetowa – przyszłość rynku e-commerce [Internet analytics – the future of the e-commerce market]. Analyzeo. Retrieved from: https://goo.gl/zhJegK [in Polish].

Barton, D., Court, D. (2012). Making advanced analytics work for you. Harv. Bus. Rev., 90(10), 78–83.

Brynjolfsson, E., Hitt, L. M., Kim, H. H. (2011). Strenght in numbers: how does data-driven decisionmaking affect firm performance? Retrieved from: https://ssrn.com/abstract=1819486

Davenport, T. H. (2013). The Rise of Analytics 3.0. How to Compete in the Data Economy. eBook: International Institute for Analytics.

Davenport, T. H., Harris, J. G. (2010). Inteligencja analityczna w biznesie. Nowa nauka zwyciężania [Competing on Analytics: The New Science of Winning]. Warszawa: MT Biznes [in Polish].

Dergiades, T., Mavragani, E., Pan, B. (2018). Google Trends and tourists' arrivals: Emerging biases and proposed corrections. Tour. Manag., 66, 108–120. DOI: https://doi.org/10.1016/j.tourman.2017.10.014

Głąb, A. (2014). Rynek biznesowy. Raport: Wpływ internetu na gospodarkę w Polsce (s. 10–12) [Business market. Report: The impact of the Internet on the economy in Poland]. Warszawa: Związek Pracodawców Branży Internetowej IAB Polska [in Polish].

Grossman, R. L. (2018). A framework for evaluating the analytic maturity of an organization. Int. J. Inf. Manag., 38(1), 45–51. DOI: https://doi.org/10.1016/j.ijinfomgt.2017.08.005

IRCenter (2015). Raport: Rynek marketing technologies w Polsce [Report: Market technologies in Poland]. Warszawa: Interactive Research Center [in Polish].

Jansen, B. J. (2009). Understanding user-web interactions via web analytics. Synthesis Lectures on Information Concepts, Retrieval, and Services. DOI: https://doi.org/10.2200/S00191ED1V01Y200904ICR006

Kaushik, A. (2007). Web Analytics: An Hour A Day. Indianapolis: Wiley Publishing, Inc.

Kaushik, A. (2010). Web Analytics 2.0. Świadome rozwijanie witryn internetowych [Web Analytics 2.0. Conscious development of websites]. Gliwice: Helion [in Polish].

Król, K. (2017). Promoting of agrotourism on the Internet – A lesson from the Visegrad Group countries. J. Agribus. Rural Dev., 4(46), 805–813. DOI: https://doi.org/10.17306/J.JARD.2017.00383

Król, K. (2019). Forgotten agritourism: abandoned websites in the promotion of rural tourism in Poland. J. Hospit. Tour. Technol., 10(3), 461–472. DOI: https://doi.org/10.1108/JHTT-09-2018-0092

McAfee, A., Brynjolfsson, E. (2012). Big Data: The Management Revolution. Harv. Bus. Rev. Retrieved from: https://goo.gl/XpevWN

Ołdak, A. K. (2010). Web Analytics. Warszawa: PARP [in Polish].

Peterson, E. T. (2004). Web analytics demystified: a marketer's guide to understanding how your web site affects your business. Celilo Group Media and CafePress.

Tarhan, A., Turetken, O., Reijers, H. A. (2016). Business process maturity models: A systematic literature review. Inf. Soft. Technol., 75, 122–134. DOI: https://doi.org/10.1016/j.infsof.2016.01.010

Wiecka, A. (2018a). Dokąd zmierza analityka w biznesie [Where is the analyst in business going]. Harv. Bus. Rev. Polska. Retrieved from: https://goo.gl/DYoLs9 [in Polish].

Wiecka, A. (2018b). Gospodarka analityczna [Analytical economy]. Harv. Bus. Rev. Polska, 182. Retrieved from: https://goo.gl/HjExAn [in Polish].

Wilson, H. J., Daugherty, P. R., Morini-Bianzino, N. (2017). The Jobs That Artificial Intelligence Will Create. MIT Sloan Manag. Rev., 58(4), 14–16.