wjert, 2019, Vol. 5, Issue 4, 123-139. *Review Article* ISSN 2454-695X

World Journal of Engineering Research and Technology
WJERT
www.wjert.org SJIF Impact Factor: 5.218

TOWARD A SMART E-PARTICIPATION TO PUBLIC PROCUREMENT GOVERNANCE USING DATA SCIENCE

Dr. Amadou Diabagate*[1] and Pr. Abdellah Azmani[2]

[1]Faculty of Mathematics and Computer Science, University Félix Houphouet Boigny.
[2]Faculty of Sciences and Technologies, University Abdel Malek Essaadi.

Article Received on 01/06/2019 Article Revised on 22/06/2019 Article Accepted on 12/07/2019

***Corresponding Author**
Dr. Amadou Diabagate
Faculty of Mathematics and Computer Science,
University Félix Houphouet Boigny.

ABSTRACT

One of the major goals of governments and public authorities around the world is the well-being of peoples and e-participation of citizens in public governance is by far one of the adequate ways to achieve it. This article focuses on e-participation specifically on its decision-making component by using data science techniques for collecting, storing and processing data generated by the e-participation of citizens in governance. It is a way toward a smart e-participation in which public authorities have in their possession a powerful decision making tool that can allows getting useful and strategic insights from e-participation data. A case study was conducted concerning the construction of Istanbul's third airport with data collected on Twitter. The aims of this work are taking into account the opinions of citizens who have to be at the hearth of each development project, making smarter the e-participation process and contributing to the satisfaction of citizens.

KEYWORDS: E-Participation, Data Science, E-Governance, Social Media Analytics, Procurement.

INTRODUCTION

Public procurement, for its great economic weight namely 20% of world GPD, are of utmost importance for governments, businesses and citizens.[1,2] The sector of public procurement faces many challenges and it is pointed out as one of the most corrupted sectors.[3,4] To solve part or all of these challenges, the e-participation of citizen in procurement governance is indubitably considered as one of the best solutions.[5]

Citizens are the first recipients of public procurement. Therefore, they are not associated to its governance in many countries. In most cases, public procurement are designed by authorities, leaded by them until achievement without involving citizens at any level of this process.

The development of Information and Communication Technologies (ICT) and their use in our lives is an irreversible process. Thus, e-governance, across the world, is considered as a priority by governments, businesses and civil societies.[6] E-participation, which allows a participation of people using ICT typically internet, is an important part of e-government and can be a powerful tool that can allow citizen’s participation in public procurement management in particular and governance in general.[7] Today, e-participation has become more and more a reality because of the emergence of advanced digital societies.

To improve the process of e-participation as well as achieve a smart e-participation that can provide strategic information for decision making, the integration of data science concepts and methods is indubitably one of the best ways.[8,9] Data science has emerged as a growing new discipline that uses essentially mathematics and computer science to build a smart decision support tools.[10]

The goal of this work is to bring a contribution to citizen e-participation by building a decision support platform for citizens and public authorities. This IT solution integrates the concepts and methods of data science namely data visualization, natural language processing, social media analytics and sentiment analysis. This IT solution, on the one hand, will take into account the opinions of citizens concerning public procurement management and, on the other hand, it will help public authorities to have a better understanding of citizens’ expectation.

The remainder of this paper is structured as follows. Section 2 exposes the concepts of e-participation and data science and the link between them. Section 3 explains public procurement and its characteristics. The results of our work concerning the contribution of data science concepts and its methods in the emerging of a smart e-participation are presented in section 4. The paper ends with concluding remarks and avenues for future research in section 6.

E-Participation and Data Science

E-participation and data science are independent concepts. However, data science can be used in many contexts including e-participation for recovery, processing and analysis of data to facilitate the decision process.

The concept of e-participation

Public participation which is defined as the redistribution of power from government to citizen is an effective way to empower citizens by taking into account their opinions and expectations in the decision making.[11]

The process of public participation, according Arnstein, can be divided into three main steps: non-participation, tokenism and citizen power. The step of non-participation has manipulation and therapy as subs-steps that are not intended to enable the participation of citizen but allow the authorities to educate and prepare citizen to public participation. Regarding tokenism, it has three sub-steps namely informing, consulting and placation that allow citizens to heard and be heard even though their opinions will not be taken into consideration. Concerning the step of citizen power, it includes the sub-steps of partnership, delegated power and citizen control. This step is a top level of public participation and it allows the citizen to become real actors in public governance by exercising the power through authorities.[12]

The Organization for Economic Cooperation and Development (OECD) and the International Association for Public Procurement (IAP2), also, respectively introduced their classifications of public participation. According OECD, public participation can be categorized into three levels namely information, consultation and active participation. IAP2 proposed, in 2007, the new classification of public procurement steps that includes information, consultation, involvement, collaboration and empowerment.

There are some similarities and differences between the steps of public procurement proposed by Arnstein, OECD and IAP2. The step of non-participation and its sub-steps of Arnstein's model have not been represented in OECD and IAP2 models. Information and consultation steps in the models of OECD and IAP represent the step of tokenism in Arnstein's model. The step of citizen power in Arnstein's model corresponds respectively to active participation in OECD's model and involvement, collaboration and empowerment in IAP2's model.

In these three models, when going up, the role of citizen in public participation becomes more active and decisive by moving from simple information consumer to decision maker.

E-participation as a new form of public participation uses ICT to enhance the implication of citizen as stakeholders through deliberation and active decision-making initiatives.[13,15] E-participation can also "refers to citizens' voluntary participation and involvement in public administration affairs and public decision-making through the use of web-based applications provided by government".[16] and citizen.

E-participation increases the participation of citizen by deleting the constraints of time and space of classic public participation. To be effective, it involves the engagement of citizens and government and it allows establishing connection between citizens and between citizens and government. It will be more attractive for citizen if their opinions are really taken into consideration and they perceive the results of their engagement on the governance actions.[17]

E-participation is adopted by many governments around the world through IT solutions like online forums, virtual discussions room and electronic polls.

Taking into consideration OECD's classification (enabling, engaging and empowerment), Machintosh proposed e-enabling, e-engaging and e-empowerment as the level of e-participation.[18]

Using IAP's classification of public participation (information, consultation, involvement, collaboration and empowerment), Tambouris *et al.* proposed five levels for e-participation namely e-information, e-consultation, e-involvement, e-collaboration and e-empowerment.[19]

Understanding of data science

The term data science is used to designate an interdisciplinary field which allows discovering knowledge from data.[20] To achieve this, this new field uses two major scientific disciplines namely mathematics (in general statistics) and computer science and to which can be added the expertise in a given field of application.[21]

The branches of mathematics, statistics and computer science used in data science are numerous, among them signal processing, probabilistic models, statistical learning, data mining, computer programming, uncertainty modelling, data warehousing, high-performance computing, etc. Also, data science field has many techniques such as machine learning, natural language processing, data analysis, web scraping, social media analytics, sentiment analysis, data visualization (data visualization), etc.[22]

Statistics is at the heart of data science because it offers the majority of scientific methods for data processing. Thus, the statistician can easily become familiar with data science by appropriating the concepts and tools of this new discipline and by perfecting himself in certain areas of computer science.[23] Given the growing need for data scientists and the coming together of data science and statistics, many proposals stipulate the integration of data science into training programs for statisticians.[24]

The term "data scientist" is used to designate a specialist in data science. The data scientist is distinguished by his ease with mathematics including statistics and his computer skills in particular computer programming.[25,27] According to IBM: “a data scientist represents an evolution from the business or data analyst role. The formal training is similar, with a solid foundation typically, in computer science and applications, modeling, statistics, analytics, and math’’.[25] However, given the interdisciplinary nature of data science[28], collaborative work between people of different skills (statistician, computer scientist, specialist in the field of application) is the ideal approach.

Public Procurement

Public procurement is an onerous contract concluded between a contracting authority (public service) and a natural or legal person called an entrepreneur, supplier or service provider for the performance of works, the delivery of supplies or the provision of services.

It can also be defined as a contract concluded between one or more economic operators (public or private entities) and one or more contracting authorities (State, local authorities, public law organisms and associations formed by these communities or institutions).

By their nature and purpose, public procurement can be grouped into three broad categories:

Public works contracts: these contracts deal with the execution or design of works such as the construction of buildings, roads and other that meet the needs specified by the contracting authority.

Public service contracts: they relate to the delivery of benefits relating to "priority" or "non-priority" services. The so-called "priority" services, limited by the texts, are subject to the rules of public procurement. This is such maintenance or repair services, land transport services, electronic communications, advertising, roads, market research, training, etc. The other services, deemed "non-priority", such as hotel and restaurant services, rail transport, legal services, social and health services or any other service may be placed under a lighter procedure.

Public supply contracts: these contracts are for the acquisition of supplies, their installation and sometimes for their maintenance. They cover all contracts outside works and services contracts relating to the purchase, leasing, rental, hire-purchase and delivery of products.

Data Science for a Smart E-Participation In Public Procurement Governance

This approach consists of using data science techniques to analyse the opinions of citizens regarding the governance of public procurement. Its purpose is to take into account these opinions and to make citizens new actors in the management of public contracts.

To implement this approach, R and its data science packages have been chosen, especially those dedicated to natural language processing, text mining, social media analytics, sentiment analysis and data visualization. The main R's packages used are:

- Twitte R: a package that allows you to connect R to Twitter and retrieve its data;
- NLP: R package dedicated to natural language processing;
- Tm: popular and widely used package for text mining;
- Ggplot2: R package dedicated to data visualization;
- Sentiment: R package for sentiment analysis that integrates the concepts of machine learning.

For illustration, tweets about the construction project of the third airport of Istanbul have been used. This latter is poorly perceived by some persons and associations who consider it as a threat to the environment.

Data recovery and preparation

```
#Load the packages
library(twitteR)
library(RColorBrewer)
library(wordcloud)
library(ggplot2)
library(NLP)
library(tm)
library(devtools)
library(Rstem)
library(sentiment)

#File required for windows user
download.file(url="http://curl.haxx.se/ca/cacert.pem",
destfile="cacert.pem")

#Access keys to twitter documentation
 consumer_key <- '***********************************'
consumer_secret <- '*********************************'
access_token <- '***********************************'
access_secret <- '**********************************'
setup_twitter_oauth(consumer_key,consumer_secret,access_token,
access_secret)

#####################        Download        the        tweets
#####################
Airport_P  <-  searchTwitter("Istanbul  New  Airport",  lang  =
'en')

#save tweets as text
Airport _P_text <- sapply(Marche_P, function(x) x$getText())

################### Clean  and  prepare  tweets  for  analysis
####################
# Remove retweet entities
Airport_P_text   =   gsub("(RT|via)((?:\\b\\W*@\\w+)+)",   "",
Airport_P_text)

# Remove at people
Airport_P_text = gsub("@\\w+", "", Airport_P_text)

# Save tweets in .csv file
write.csv(Airport_P_text,   file   =   '../Airport_P_text.csv',
row.names=F)
```

Fig. 1: R code for retrieving and cleaning tweets.

The code in Fig. 1 allows you to connect to twitter documentation through identifiers, to load tweets with specific keywords and hashtags, to clean and prepare tweets for processing and to save them in a .csv file.

Application of data visualization

Data visualization is a data science method that consists of making graphical representations of data to improve understanding, communication and decision-making. In other words, it is the art and practice of collecting, analyzing and graphing information (data).[29,31] This is not a new technique but it is growing today thanks to its key role in decision-making systems and its use in Big Data and natural language processing.[32,33] Indeed, for decision makers, who are largely not statisticians, the visualization of data allows them to have a good readability of a given situation.

The application of data visualization in this study is to develop some graphical representations. Thus, after cleaning up and preparing tweets for analysis, the data visualization stage will help better understand of the tweets via a graphical representation of key information. Fig. 2 illustrates the R code used to visualize data in the corpus which contains all the tweets collected.

```
######     Data     visualization :     wordcloud,     histogram
##############

#Create the corpus
Marche_P_text_corpus <- Corpus(VectorSource(Marche_P_text))

# Make the Wordcloud
Marche_P_text_corpus       <-       tm_map(Marche_P_text_corpus,
function(x)removeWords(x,stopwords()))
pal2 <- brewer.pal(8,"Dark2")
wordcloud(Marche_P_text_corpus,       min.freq=2,max.words=300,
colors=pal2)

#Build the tdm (term document matrix)
tdm <- TermDocumentMatrix(Marche_P_text_corpus)

#Most frequent terms
(tt <- findFreqTerms(tdm, lowfreq=30))

# Count the number of occurrences of the most frequent words
termFrequency <- rowSums(as.matrix(tdm[tt,]))
termFrequency

# Build the histogram
qplot(names(termFrequency),     termFrequency,     geom="bar",
stat="identity")
```

Fig. 2: R code for data visualization.

Fig. 3: Wordcloud representing the most common words in tweets.

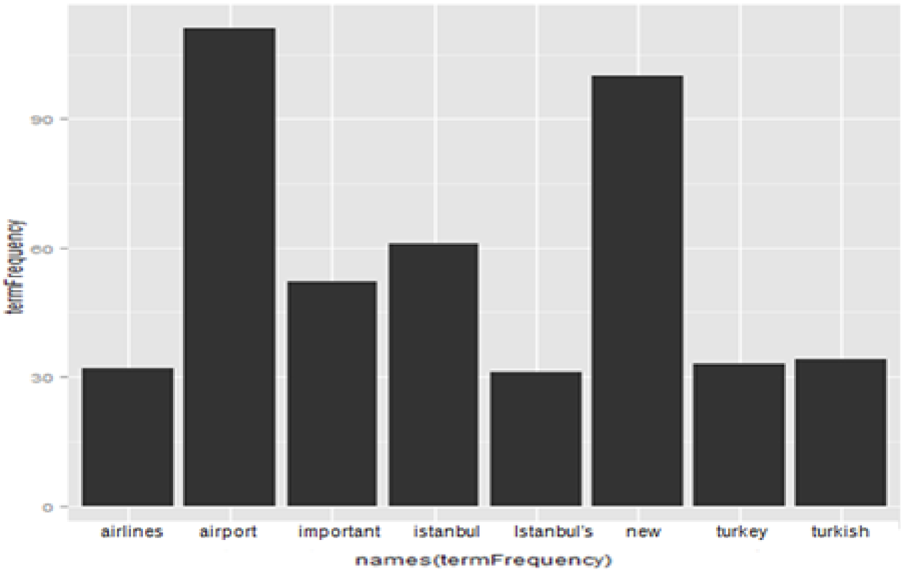

Fig. 4: Histogram of the most common words.

The wordcloud, popular in text mining, was used to represent the most frequent words (Fig. 3). In addition, a histogram (Fig. 4) has been made to highlight the words having a frequency equal to at least 30.

The histogram and the wordcloud both show that the words "airport" and "new" are the most frequent.

Application of social media analytics and sentiment analysis

The proposed e-participation approach of involving citizens in the governance of public procurement integrates social media analytics. The latter contains all practices for collecting data from blogs, social networks (Twitter, Facebook, Instagram, Youtube, linkedIn, Flickr, etc.) and analyze them for descriptive and decision-making purposes.[34,38] Also called social media intelligence, it is used by professionals in many fields: business, marketing, politics (popularity rating), transportation, sales, etc.[39,43] Its most common use is to identify the opinions (sentiments) of blogs and social networks users, which establishes a close link between social media analysis and sentiment analytics.[44]

Human beings, in a general and natural way, always have an opinion on a given situation. The sentiment analysis is a data science method whose purpose is the analysis of opinions, sentiments, attitudes, emotions from written language (text).[45,47] making it a method belonging to the big family of natural language processing methods.[48]

The sentiment analysis is a recent method. This is because statisticians used surveys and opinion polls to identify people's opinions about a given situation, the lack of statistical methods of extracting opinions from written documents (texts) and the lack of powerful computers capable of storing and implementing greedy methods in terms of calculations.[49] Sentiment analysis has flourished with the advent and development of the web and social media,[50] the availability of more powerful computers in terms of storage capacity and RAM and advances in statistical methods.[38]

Many tools have been developed to make social media analytics and sentiment analysis among which packages of R and Python modules that have the advantage of being generally free. In this study, the package "sentiment" of R that integrates the machine learning method «Naïve Bayes» has been used.

Machine learning denotes a set of methods aimed at analyzing, interpreting and predicting a phenomenon based on a sample of observations described by J independent or explanatory variables.

Let $\mathcal{L}_n = \{(x_1, y_1), \dots, (x_n, y_n)\}$ be a sample of training; i.e. a series of independent and identically distributed (i.i.d) random variables having the same distribution as a random vector (X, Y) which distribution is unknown. Let $\mathbf{X}$ and $\mathbf{Y}$ be respectively spaces in which live the random variables X and Y.

For every new input $\mathrm{x} = (x_1, \dots, x_n) \in \mathbf{X}$, the aim of supervised machine learning is to predict $\hat{\mathrm{y}} \in \mathbf{Y}$. Depending on the space $\mathbf{Y}$, supervised learning is:

- A regression if $y \in \mathbb{R}$;
- A classification if $y \in \{C_k\}_{1 \le k \le K}$ where C_k represent the classes (categories).

A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumption. Given a problem instance to be classified $X = (x_1, x_2, \dots, x_n)$ representing n independent variables, it assigns to this instance probabilities:

$$P(C_k | x_1, \dots, x_n) \text{ , for each of k classes} \quad (1)$$

Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable.

Naïve-Bayes has been used as an effective classifier. It is easy to construct Naïve Bayes classifier as compared to other classifiers because the structure is given a priori and hence no structure learning procedure is required. Naïve Bayes assumes that all the features are independent of each other.

An advantage of the naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters necessary for classification. Naïve Bayes is based on Bayes theorem defined as follows:

$$P(C_k | x) = \frac{P(C_k) * P(x | C_k)}{P(x)} \quad (2)$$

By using Bayes theorem and under the independence assumption, the probabilities define in x become as follows:

$$P(C_k | x_1, \dots, x_n) = \frac{P(C_k) * \prod_{i=1}^{n} P(x_i | C_k)}{P(x_1, \dots, x_n)} \quad (3)$$

To build the Naïve Bayes classifier, the Naïve Bayes probabilities model defined in y is combined with a decision rule. For example, by using the maximum a posteriori decision rule, the Naïve Bayes classifier is the function that assigns a class label $\hat{y} = C_k$ for some k as follows:

$$\hat{y} = argmax \left\{ P(C_k) * \prod_{i=1}^{n} P(x_i | C_k) \right\}_{1 \le k \le K} \quad (4)$$

Fig. 5 shows the R code for the implementation of the Naïve Bayes by using "classify_polarity" method of the package "sentiment", which allows subdividing the tweets into three categories (negative, neutral, positive). This code, also, counts the number of tweets in each category.

Fig. 6 shows the code used to visualize the opinions. Of all the tweets collected, 95% are positive about the construction project of Istanbul's third airport, 4% are negative and 1% is neutral. If this trend is confirmed with a large number of tweets, the political authorities can use this result as an argument to justify the citizens' support for this project. Fig. 7 shows a histogram of the polarity of opinions from tweets.

```
################ Sentiment analysis (polarity) ##############
# classify polarity
class_pol          =          classify_polarity(Marche_P_text,
algorithm="bayes")

# get polarity best fit
polarity = class_pol[,4]

# data frame with results
sent_df     =     data.frame(Marche_P_text,     polarity=polarity,
stringsAsFactors=FALSE)

# Number of occurrences in modalities in "Polarity"
table(sent_df[2])
```

Fig. 5: R code for sentiment analysis.

```
# data visualization: plot distribution of polarity
ggplot(sent_df, aes(x=polarity)) +
  geom_bar(aes(y2=..count.., fill=polarity)) +
  scale_fill_brewer(palette="RdYy") +
  labs(x="polarity categories", y="number of tweets")
  ggtitle("Sentiment  Analysis:  histogram  of  opinion  (e-
particpation)")+
  theme(plot.title = element_text(size=12))
```

Fig. 6: R code for viewing results on a histogram.

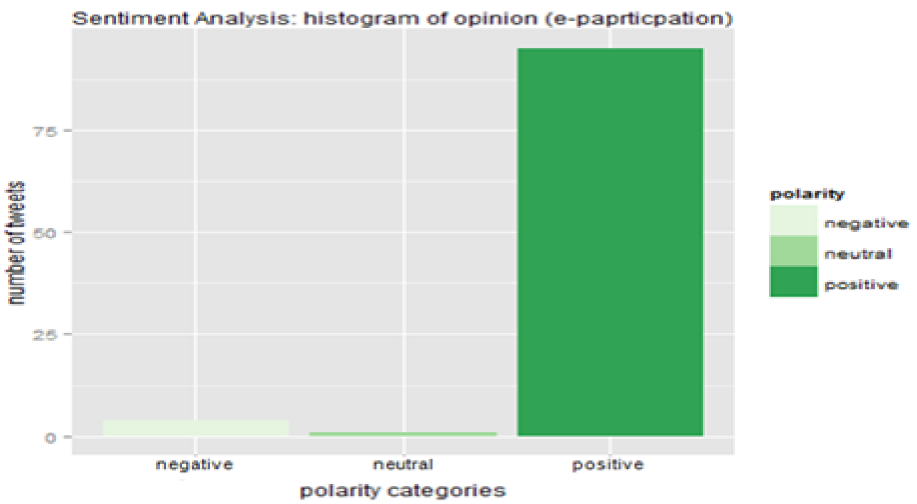

Fig. 7: Histogram of opinions.

Conclusion and perspectives

All public authorities claim that their main goal is to work for people happiness and well-being. This goal cannot be really achieved without people participation to the main activities of public governance. For this purpose, since ancient societies, political leaders to ensure the contribution of people have implemented some mechanisms. Today, with the advent of information and communication technologies, governments and citizen have at their disposition a powerful tool that can help them improve and make better the citizen participation in governance.

E-participation process and the data it generates can be used to make powerful and strategic decision-making tools, which can allow achieving its main objective: citizen participation and taking into account of their opinions. Moreover, political leaders and citizen can use these decision-making tools to reach others goals such as political stability, measuring of polarity rating, transparency and good governance.

This paper proposes, in the context of e-participation, decision support tool by using data science methods namely data visualization, sentiment analysis and social media analytics. In that sense, twitter data have been collected, stored and processed with a view to identifying and understanding citizen opinions and expectation.

In terms of perspectives, the concept of Big Data will be applied because of the huge quantity of data that can be produced by the e-participation of citizens to public procurement governance. To that end, a Hadoop or Spark cluster will be built to stock, manage and process the data. In addition, for real time processing and providing real time information, a Storm cluster will be deployed.

REFERENCES

1. Auriol, E. "Corruption in procurement and public purchase," *Int. J. Ind. Organ.*, 2006; 24(5): 867–885.
2. Koning, P. and Van de Meerendonk, P. "The impact of scoring weights on price and quality outcomes: An application to the procurement of Welfare-to-Work contracts," *Eur. Econ. Rev.*, 2014; 71: 1–14.
3. Auriol, E. "The Cost of Corruption in Procurement and Public Purchase,", 2004.
4. Hessami, Z. "Political corruption, public procurement, and budget composition: Theory and evidence from OECD countries," *Eur. J. Polit. Econ.*, 2014; 34: 372–389.
5. Sharma, G. "Public Participation and Ethical Issues on E-governance: A Study Perspective in Nepal,", 2014; 12(1): 82 - 96.
6. Yadav, N. and Singh, B. "E-Governance: Past, Present and Future in India," International Journal of Computer Applications, 2012; 53(7): 36-48.
7. Oyediran, S. and Akintola, A. "A survey of the state of the art of e-tendering in nigeria," *Journal of Information Technology in Construction (ITcon)*, 2011; 16: 557-576.
8. Agasisti, T. and Bowers, J. "Data Analytics and Decision-Making in Education: Towards the Educational Data Scientist as a Key Actor in Schools and Higher Education Institutions,", 2017; 23.
9. Aerts, H. "Data Science in Radiology: A Path Forward," *Clin. Cancer Res.*, 2018; 24(3): 532–534.
10. Chou, S. et al. "Democratizing Data Science."
11. Arnstein, R. "A Ladder Of Citizen Participation," *J. Am. Inst. Plann.*, 1969; 35(4): 216–224.
12. Fedotova, O. and Teixeira, L. and Alvelos, H. "E-participation in Portugal: evaluation of government electronic platforms," *Procedia Technology*, 2012; 5: 152–161.
13. Sæbø, Ø. and Rose, J. and Skiftenes Flak, L. "The shape of eParticipation: Characterizing an emerging research area," *Gov. Inf. Q.*, 2008; 25(3): 400–428.
14. Organisation for Economic Co-operation and Development, Ed., *Promise and problems of e-democracy: challenges of online citizen engagement*. Paris: OECD, 2003.

15. Wimmer, A. "Ontology for an e-participation virtual resource centre," in *Proceedings of the 1st international conference on Theory and practice of electronic governance*, 2007; 89–98.
16. Kim, S. and Lee, J. "E-Participation, Transparency, and Trust in Local Government," *Public Adm. Rev.*, 2012; 72(6): 819–828.
17. Komito, L. "e-Participation and Governance: Widening the net," *Electron. J. E-Gov.*, 2005; 3(1): 39–48.
18. Macintosh, A. "Characterizing e-participation in policy-making," in *System Sciences, Proceedings of the 37th Annual Hawaii International Conference on*, 2004; 1–10.
19. Tambouris, E. et al., "Introduction eParticipation," in DEMO-net booklet series, 1, 2007.
20. Jifa, G. and Lingling, Z. "Data, DIKW, Big Data and Data Science," *Procedia Comput. Sci.*, 2014; 31: 814–821.
21. Shi, Y. et al. "Explore New Field of Data Science Under Big Data Era: Preface for ICDS 2014," *Procedia Comput. Sci.*, 2014; 30: 1–3.
22. O'Halloran, S. et al. "Data Science and Political Economy: Application to Financial Regulatory Structure," *RSF Russell Sage Found. J. Soc. Sci.*, 2016; 2(7): 87-109.
23. Besse, P. et al. "Big Data Analytics-Retour vers le Futur 3*; De Statisticien à Data Scientist*," *Systèmes d'information et big data*, 2014; 19(3): 93-108.
24. Hardin, J. et al. "Data Science in the Statistics Curricula: Preparing Students to Think with Data," *ArXiv Prepr. ArXiv14103127*, 2014.
25. Ecleo, J. "Surveying LinkedIn Profiles of Data Scientists: The Case of the Philippines," *Procedia Comput. Sci.*, 2017; 14: 53–60.
26. Ford, H. "Big Data and Small: Collaborations between ethnographers and data scientists," *Big Data Soc.*, 2014; 1(2): 1-3.
27. Davenport, H. and Patil, J. "Data scientist," *Harv. Bus. Rev.*, 2012; 90(10): 70–76.
28. Chatfield, T. et al. "Data scientists as game changers in big data environments," *25th Australasian Conference on Information Systems*, 8^{th} -10^{th} Dec, Auckland, New Zealand, 2014.
29. Schwabish, A. "An Economist's Guide to Visualizing Data," *J. Econ. Perspect.*, 2014; 28(1): 209–234.
30. Heckerman, D. et al. "Dependency networks for inference, collaborative filtering, and data visualization," *J. Mach. Learn. Res.*, 2000; 1: 49–75.

31. Liu, L. et al. "Analyzing Personality through Social Media Profile Picture Choice," *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM),* Cologne, Germania, 17th-20th May, 2016; 211-220.
32. Gorodov, Y. and Gubarev, V. "Analytical Review of Data Visualization Methods in Application to Big Data," *J. Electr. Comput. Eng.*, 2013; 1–7.
33. Wang, L. et al. "Big Data and Visualization: Methods, Challenges and Technology Progress," *Digital Technologies,* 2015; 1(1): 33-38.
34. Pick, B. et al. "A Spatial and Regression Analysis of Social Media in the United States Counties," *Pre-ICIS Workshop on Locational Analytics and Big Data*, Fort Worth, 2015; 21.
35. Sie, L. et al. "Social network analysis for technology-enhanced learning: review and future directions," *Int. J. Technol. Enhanc. Learn.*, 2012; 4(3/4): 172.
36. Coppersmith, G. et al. "Exploratory Analysis of Social Media Prior to a Suicide Attempt," *Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, San Diego, California, June 16, 2016; 106–117.
37. Ratkiewicz, J. et al. "Detecting and Tracking Political Abuse in Social Media," *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain, 2011; 298-304.
38. Sinha, V. et al. "The contemporary framework on social media analytics as an emerging tool for behavior informatics, hr analytics and business process," *Management*, 2012; 17: 65-84.
39. Ruiz Sánchez, T. et al. "Social networks, big data and transport planning," *CIT2016 – XII Congreso de Ingeniería del Transporte*, València, Italia, 2016.
40. Gil de Zúñiga H. and Diehl, T. "Citizenship, Social Media, and Big Data: Current and Future Research in the Social Sciences," *Soc. Sci. Comput. Rev.*, 2017; 35(1): 3–9.
41. Yin, J. et al. "Using Social Media to Enhance Emergency Situation Awareness: Extended Abstract," *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, Buenos Aires, Argentina, 25th - 31th, 2015; 4234-4238.
42. Michaelidou, N. et al. "Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands," *Ind. Mark. Manag.*, 2011; 40(7): 1153–1159.
43. Wang, Y. et al. "Unsupervised Sentiment Analysis for Social Media Images," *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence,* Buenos Aires, Argentina, 25th - 31th, 2015; 2378-2379.

44. Pak, A. and Paroubek, P. “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” *IJARCCE*, 2016; 5(12): 320–322.
45. Wang, H. et al. “A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle,” *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, Republic of Korea, 8-14 July, 2012; 115–120.
46. Habernal, I. et al. “Sentiment Analysis in Czech Social Media Using Supervised Machine Learning,” *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity*, Atlanta, Georgia, 14 June, 2013; 65–74.
47. Wilson, T. et al. “Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis,” *Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, Vancouver, Canada, 6-8 October, 2005; 347–354.
48. Godbole, N. et al. “Large-Scale Sentiment Analysis for News and Blogs,” ICWSM’2007 Boulder, *Colorado, USA*, 2017; 4.
49. Xie, B. et al. “Sentiment analysis of Twitter data,” *Proceedings of the Workshop on Language in Social Media*, Oregon, Portland, 23 June, 2011; 30–38.
50. Mohammad, M. et al. “NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets,” *Proceedings of the Seventh International Workshop on Semantic Evaluation*, Atlanta, Georgia, USA, 14th-15th June, 2013; 321–327.