Current Practice and Challenges of Data Use and Web Analytics in Online Participations

Igor SEROV [a], Maria LEITNER [a,1], Stefanie RINDERLE-MA [b]

[a] *AIT Austrian Institute of Technology GmbH, Digital Safety & Security Department*
[b] *University of Vienna, Faculty of Computer Science*

Abstract. Information system design and implementation are key factors for electronic participatory processes and procedures. How information systems are designed does not only affect the procedures but also influences the trust building between organizers, operators and participants. In addition, the implementation often has to adhere to legal standards. In this paper, we aim to investigate current practice of data use in online participations. In particular, a qualitative analysis is conducted and 18 online participations are investigated on their data use, i.e. use of participant information, cookies and web analytics. The results show that most projects require and request data during site visits (e.g., IP address, browser type) and for active participation (e.g., name, email). The real benefit, however, for the use of web analytics is often unclear. Furthermore, often proprietary solutions for web analytics are used, even tough open source solutions (i.e. that store data locally) exist. For future projects, it is recommended to not only define but also keep privacy policies updated (according to the used technology) and to specify the purpose and goals of using web analytics.

Keywords. E-participation, Cookies, Data Use, Web Analytics, Security

Introduction

Information and Communication Technologies (ICT) are a key factor for the success and acceptance of e-participation as shown for example in [13,15]. Electronic, participatory decision processes are dependent on the underlying technologies used. Hence, the decisions for information system design (e.g., designing an architecture, define software engineering tools and methods) and furthermore the concrete information system implementation (e.g., choosing data formats and protocols) have critical impact on the e-participation solutions such as the acceptance, privacy or usability.

---

[1] Corresponding Author: Maria Leitner, AIT Austrian Institute of Technology GmbH, Digital Safety & Security Department, Donau-City-Straße 1, 1220 Vienna, Austria; E-mail: maria.leitner@ait.ac.at.

Furthermore, the technology has an influence on the participation procedures itself and the participants. In particular, trust plays a significant role for (electronic) participation initiatives (e.g., [14,6,16]). For example in [17], a security analysis is performed to identify potential risks and shortcomings that could have an affect on the trust of people towards e-participation technologies. In fact, the authors define security requirements and preventive measures to minimize the risk of exposure and to enable "*trust by design*".

The e-participation domain is particularly interesting as trust is likely to be an important factor when participants join and partake in initiatives. Technology is a key factor to provide stable, running and trustworthy implementations for these participatory decision processes. The exposure to threats and attacks might result into a decrease of trust in the cause and implementation. For example, a website defacement that changes the website interface for a few hours or the loss of personal data might lead to a loss of reputation and even more critical the motivation for people to join online participations. That is why adequate implementations should carefully consider the data required during a regular website visit (without active participation such as IP address or browser type) and during active participation (e.g., name, address) as well as the used web technologies.

In this paper, we investigate how current online participations are designed and developed in terms of data use and web analytics. In particular, we aim to analyze a set of online participations and evaluate what data is collected and for which use (if this can be determined), in particular the data collection *during a site visit (without active participation)* or *during active participation*. Furthermore, we aim to address current web technologies in online participations, i.e. the use of cookies and web analytics in this context. These features are important as they are often closely related to data protection laws and the technological feasibility. With this paper, we aim to address the following research questions: (A) How does current practice handle data use and web analytics in online participations? (B) How does current practice declare data use during a site visit (without active participation) or during active participation (e.g., in a privacy policy)? (C) What are recommendations or observations for data collection and the use of cookies, web analytics in online participations?

With this approach, we aim to identify challenges of data usage and raise awareness in e-participation. So far, research has acknowledged that security and privacy principles are important for the trust within e-participation (e.g., [20]), however, it is not clearly defined how this should be handled nor how the usage of data (e.g., with web analytics) influences the trust of participants. Furthermore, benefits of the use of web analytics for citizens or providers is yet undefined. This paper aims to start a discussion on best practices and to provide insights from current practice for future developments.

The rest of the paper is structured as follows: Section motivates the topic. Section 1 investigates related background research. Section 2 outlines the methodology used in the qualitative analysis. Furthermore, Section 3 summarizes the main results of the analysis. Lastly, Section 4 specifies observations and recommendations and concludes the paper.

1. Background

Trust between operators and participants can be promoted by choosing secure and stable technologies [17,7]. Standard web technologies are cookies and web analytics, for example. In e-participation, the use of cookies or web analytics have been proposed (e.g., [8]) but have not been investigated towards best practices. In this paper, we use the term *data* as "*a set of values of information*". These values could be for example names, address, etc. Which and how data is used in online participations is specified as "*data use*" in this paper. Data is used by web technologies such as cookies or web analytics.

A *cookie* is "*a text file that is placed on user's computer hard drive by a Web site when the user visit that site*" [3]. Cookies are intended to make the Internet surfing easier and more comfortable. Based on [3] and [5], there are two types of cookies: **Persistent cookies** "*help identify a unique browser to the website, inasmuch as they are the closest thing to tracking a person or unique visitors*" according to [5]. Persistent Cookies do not contain any personal information. **Session cookies** help the website to keep track of user movements in the website without repeatedly requiring a user's authentication and expire when the user closes the browser or logs out. Cookies can be subject to hijacking or other threats (e.g., [11,2,4]) that is why their usage must be carefully planned and implemented in all application domains (e.g., e-participation, e-government, e-commerce).

Web analytics "*is the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing web usage*" [19]. Different metrics exist for web analytics: For example, basic building blocks include page views, visits, unique visitors. Building on these blocks, further metrics such as returning visitors, repeating visitors, visitor referrer, page exit ratio or page view per visits can be established. Several open source software (e.g., Piwik, AWStats, Open Web Analytics), proprietary software (e.g., Mint, Splunk) and (proprietary) software as a service (e.g., Google Analytics, Adobe Analytics, Webtrends) for web analytics exist. In the following, we will describe the software found during our analysis: **Google Analytics** is a *software as a service* for web analytics (see `http://www.google.com/analytics/`). Google Analytics provides website owners JavaScript libraries to record information what user has seen/ done on the website. Google Analytics uses HTTP cookies for these purposes. Google Analytics is ease of use, flexible and easy to configure [9]. **Piwik** is an open source program for web analytics (see `http://piwik.org/`). Unlike other web analytics software, Piwik can be directly hosted on the client server and tracked data can be stored within an inhouse database. Therefore, clients have full control and access to the tracked data. Piwik uses also HTTP cookies. Piwik does not share tracked data with advertising companies.

2. Methodology

In this paper, we address the three research questions: (A) How does current practice handle data use and web analytics in online participations? (B) How does current practice declare data use during a site visit (without active partici-

pation) or during active participation (e.g., in a privacy policy)? and (C) What are recommendations or observations for data collection and the use of cookies, web analytics in online participations?

Questions A, B and C are answered within three steps: (1) manual search, (2) qualitative analysis, and (3) manual check. A (1) *manual search* was conducted to identify potential online participations using Google search. Therefore, keywords such as "*e-participation website*", "*e-participation platform*" were used. The search resulted in a set of 18 online participations (i.e. online participations are synonymously named as websites or platforms throughout the paper) as shown in Table 1. Based on this set, a (2) *qualitative analysis* was conducted between December 2015 and January 2016. In particular, we analyzed the following characteristics of the websites:

- *Level of participation*: Participatory processes can be divided in four tiers (see [12]): Information, consultation, cooperation and co-decision. We applied these levels of participation to categorize the 18 online participations.
- *Procedure:* This characteristic measures the usage of an electronic (i.e. online) or offline channels. A combination of both can be possible [1].
- *Duration*: Duration measures the temporal runtime of the participatory process. We adapted the duration from [1] to: permanent, periodical or onetime procedures.
- *Data use*: This feature analyzes the data that is collected during a site visit (i.e. no active participation, e.g., IP address or browser type) or active participation (e.g., name, email).
- *Web analytics*: This category refers to the collection and analysis of web data for the purpose of understanding and optimizing web usage (e.g., geolocation). In particular, we will investigate which software for web analytics is used and declared in the online participation.

In addition, we performed a (3) *manual check*. We used the declared information (e.g., in privacy policies about cookies and software for web analytics) and compared it to the real usage. Therefore, we investigated the privacy settings in the browser (in our case Firefox 44.0.2) to identify created cookies during the a site visit.

Several limitations could be identified in our study. The selection of platforms was reduced to the spoken languages of the authors (English or German). Hence, only English or German online participations could be selected. Furthermore, the authors emphasized on online participations in Europe, however, two projects were selected that are located in the USA. Keeping these reasons in mind, a qualitative analysis was performed to determine qualitative features (e.g., the use of web analytics). However, this leaves room for quantitative analysis in future work.
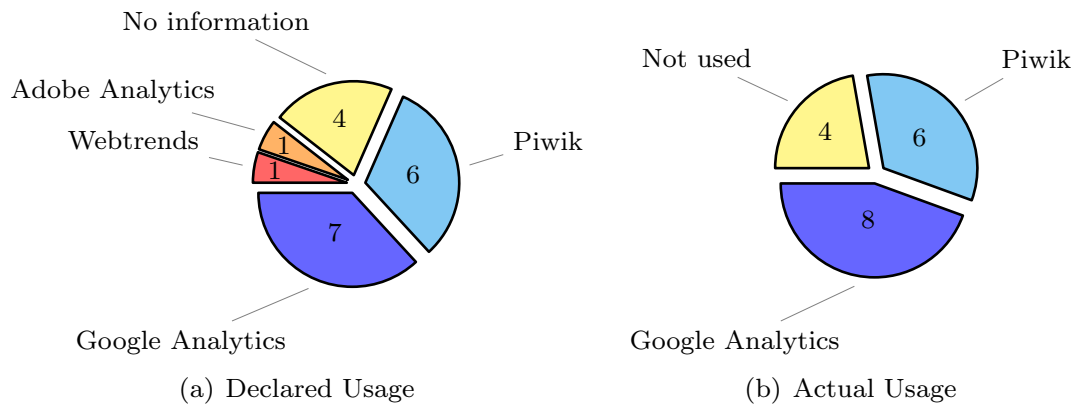
3. Results

This section summarizes the main results of our analysis for research questions (A) and (B) (see Section 2). In particular, we give an overview of the investigated platforms and analyze their use of data, cookies, web analytics.

Table 1. List of Online Participations

No.	Site	CC	URL
1	**Departament of Justice**	US	`http://www.justice.gov/oip/` `make-foia-request-doj`
2	**Direkt zur Kanzlerin**	DE	`http://www.direktzurkanzlerin.de/`
3	**Petitions UK Government**	UK	`https://petition.parliament.uk/`
4	**Züri wie neu**	CH	`https://www.zueriwieneu.ch/`
5	**Bristol Consultation Hub**	UK	`https://bristol.citizenspace.com/`
6	**Der Online-Dialog der Stadt Köln**	DE	`https://buergerhaushalt.stadt-koeln.de/2015/`
7	**Digital Agenda Wien**	AT	`https://www.digitaleagenda.wien/de`
8	**Planning Portal**	UK	`http://www.planningportal.gov.uk/`
9	**Stadtdebatte Berliner Mitte**	DE	`http://stadtdebatte.berlin.de/`
10	**Regulation.gov**	US	`http://www.regulations.gov/`
11	**Abgeordneten-watch**	DE	`http://www.abgeordnetenwatch.de/`
12	**Your voice in Europe**	EU	`http://ec.europa.eu/yourvoice`
13	**Bewegt Politik Campact**	DE	`https://www.campact.de/`
14	**WeAct**	DE	`https://weact.campact.de/`
15	**Kommission Lagerung hoch radioaktiver Abfallstoffe**	DE	`http://www.bundestag.de/endlager/`
16	**Online-Konsultation Publikationssystem**	DE	`http://konsultation.publikationssystem.de/`
17	**Bürgerhaushalt Lichtenberg**	DE	`https://www.buergerhaushalt-lichtenberg.de/`
18	**Frankfurt fragt mich**	DE	`https://www.ffm.de/frankfurt/de/home`

3.1. Data Use

This aspect is concerned with what data is collected during a regular (web) site visit and during the participation. From the selected examples, we observed that allmost all of online participations collect anonymous information (see Section 3.3). Considering this, it is not surprising that IP, browser type, time of a visit and visited pages are often logged. In particular, during site visit the following data is collected (the number in brackets represent the sum of online participations that collect the piece of data): IP (15), browser type (9), time of visit (9), visited pages (7), 7/OS (7), referring site (6), ISP name (5) and two websites specified no information for data collection (2). However, how this information is exactly used and if the information provides actual benefits to the web usage (e.g., usability improvements) is not clear and can be subject to future work. Furthermore, we examined data requested during active participation; i.e. when a participant actively gets involved such as filling out a form or registers to make or rate a comment. In this case, personal data such as name, email, address or telephone number are requested and can be mandatory or optional. The exact results for active participation are: email (18), name (16), postcode (6), address (3), country (3), telephone number (2), district (2), gender (2), age group (2) and background (2).

(a) Declared Usage

(b) Actual Usage

Figure 1. Use of Web Analytics

3.2. Cookies

Cookies are frequently used in online participations. The results of the analysis show that almost all websites declare online that they are using session (15) and persistent (16) cookies. Only few websites provide concrete information about the use of cookies: used types of cookies, purposes and expiration time. Some website do not have information about cookies, however, they provide information about the usage of web analytics software and according to this information we can suppose/expect that they are using session and persistent cookies. We reviewed the declaration by manually reviewing the cookie use during a site visit. Persistent (17) and session (18) cookies are used in the analyzed websites but are sometimes not declared on the website.

3.3. Web Analytics

Web analytics are frequently used in websites for e-participation as shown in Figure 1. In Figure 1(a), it can be seen that almost all websites are using web analytics for collection and analysis of web data. However, we wanted to assess the actual usage too (see Figure 1(b)). In particular, we assessed which cookies were created during the site visit and further assess if these cookies belong to Google[2] or Piwik[3], for example. The results show that not all websites declare their web analytics usage correctly. 14 out of 18 websites use cookies for web analytics. However, we cannot identify whether this is intended or unintended in this study. To do this, a in-depth-analysis with contacting the providers would be required.

4. Observations, Recommendations and Conclusion

Trust and transparency play an important role in the context of e-participation (see [20]). For example, trust can be established by designing and implementing

---

[2]Google analytics uses the following cookies: _ga, _gat, __utma, __utmt, __utmb, __utmc, __utmz, __utmv (see `https://developers.google.com/analytics/devguides/collection/analyticsjs/cookie-usage#gajs`, visited on 11.03.16).

[3]Piwik uses _pk_ref, _pk_cvar, _pk_id, _pk_ses or piwik_ignore (see `http://piwik.org/faq/general/faq_146/`, visited on 11.03.16).

tools that use secure and privacy-aware technologies or by providing updated privacy policies. To summarize the findings for research question (C) (see Section 2), we derived three recommendations that can boost trust and transparency in e-participation based on the results of the study.

- *Security-aware information system design and implementation*: Building a secure and privacy-aware information system starts with the design and implementation. Building security and privacy measures into information systems are important particularly for systems that cooperate and interact with the public. Incidents or other events are likely to have a high public coverage and might affect the intended project unexpectedly. Therefore, it is important to use - already during the software engineering - a privacy-aware or security-aware approach (e.g., [18]). Furthermore, the information system design should already cover and describe policies for data use and web analytics.
- *Use of web analytics*: The use and benefit of web analytics in e-participation is not yet clearly determined. Apart from the general benefits such as usability improvements or analysis of web experience, it can be easily misused and lead to threats (see e.g., [10]). Furthermore, information about web analytics tools should be clearly documented in the terms of service of web applications. We found mixed declarations on how web analytics is used (or how long data is going to be stored). Moreover, it is astonishing that operators tend to use the proprietary software than open source software. However, future developments should also consider open source solutions; they often provide the opportunity to own all data generated with web analytics (as opposed to proprietary solutions).
- *Transparent and consistent privacy policy*: Operators should provide comprehensive and updated information about the used technologies (e.g., cookies or web analytics). Updates and modifications of the privacy policy should be marked or highlighted; so that participations can understand and identify updates easily. For example, this includes information about the usage of cookies (e.g., purpose and expiration date).

Although these recommendations seem very generic at first, they seem to be needed as current practice shows different results. For example, policies are not up-to-date and do not apply to the used technology.

In summary, this paper described and analyzed current practice of data use and web analytics. The results showed that online participation policies often not fully declare their use of cookies and web analytics. This could be better maintained in permanent online participations. In the long run, this can contribute to providing more secure participatory processes and promote trust between operators and participants. For future work, we aim to investigate how data use and privacy policies can affect the trust of participants in online participations. This should provide insights on how these policies should be visualized and defined.

Acknowledgements

References

[1] Brückner, M., Märker, O.: E-Partizipation: Elektrifizierung der Bürgerbeteiligung. Standort - Zeitschrift für Angewandte Geographie pp. 112–119 (2015)

[2] Bugliesi, M., Calzavara, S., Focardi, R., Khan, W.: Automatic and Robust Client-Side Protection for Cookie-Based Sessions. In: Engineering Secure Software and Systems, pp. 161–178. No. 8364 in LNCS, Springer (Feb 2014)

[3] Danesh, A., Mehrassa, A., Lau, F.: Safe and Secure: Secure Your Home Network, and Protect Your Privacy Online. Sams Publishing (2002)

[4] De Groef, W., Devriese, D., Nikiforakis, N., Piessens, F.: Secure multi-execution of web scripts: Theory and practice. Journal of Computer Security 22(4), 469–509 (Jan 2014)

[5] Kaushik, A.: Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity. Serious skills, Wiley (2010)

[6] Kim, S., Lee, J.: E-Participation, Transparency, and Trust in Local Government. Public Administration Review 72(6), 819–828 (Nov 2012)

[7] Leitner, M., Rinderle-Ma, S.: A systematic review on security in Process-Aware Information Systems – Constitution, challenges, and future directions. Information and Software Technology 56(3), 273–293 (Mar 2014)

[8] Loukis, E., Charalabidis, Y.: Active and Passive Crowdsourcing in Government. In: Policy Practice and Digital Science. No. 10 in Public Administration and Information Technology, Springer International Publishing (2015)

[9] Marek, K.: Using Web Analytics in the Library. Library technology reports : expert guides to library systems and services, American Library Association (2011)

[10] Mayer, J.R., Mitchell, J.C.: Third-Party Web Tracking: Policy and Technology. In: 2012 IEEE Symposium on Security and Privacy (SP). pp. 413–427 (May 2012)

[11] Park, J.S., Sandhu, R.: Secure cookies on the Web. IEEE Internet Computing 4(4), 36–44 (Jul 2000)

[12] Parycek, P.: Positionspapier zu E-Democracy und E-Participation in Österreich. White Paper EDEM-1.0.0 (Jun 2008), `http://reference.e-government.gv.at/fileadmin/_migrated/content_uploads/EDEM-1-0-0-20080525.pdf`

[13] Peristeras, V., Mentzas, G., Tarabanis, K., Abecker, A.: Transforming E-government and E-participation through IT. IEEE Intelligent Systems 24(5), 14–19 (Sep 2009)

[14] Rousseau, D.M., Sitkin, S.B., Burt, R.S., Camerer, C.: Not So Different After All: A Cross-Discipline View Of Trust. Academy of Management Review 23(3), 393–404 (Jan 1998)

[15] Scherer, S., Wimmer, M.A.: Analysis of Enterprise Architecture Frameworks in the Context of e-Participation. In: Proc. of the 12th Annual Int'l Digital Government Research Conf.: Digital Government Innovation in Challenging Times. pp. 94–103. ACM (2011)

[16] Scherer, S., Wimmer, M.A.: Trust in e-Participation: Literature Review and Emerging Research Needs. In: Proc. of the 8th International Conference on Theory and Practice of Electronic Governance. pp. 61–70. ICEGOV '14, ACM (2014)

[17] Scherer, S., Wimmer, M.A.: Vertrauensförderung in E-Partizipation. Datenschutz und Datensicherheit - DuD 39(5), 295–302 (Apr 2015)

[18] Terbu, O., Hötzendorfer, W., Leitner, M., Bonitz, A., Vogl, S., Zehetbauer, S.: Privacy and security by Design im agilen Softwareprozess. In: IRIS 2016. pp. 457–464. OCG (2016)

[19] Web Analytics Association: Web analytics definitions (2008), `http://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAnalyticsDefinitions.pdf`

[20] Wimmer, M.A., Scherer, S., Appel, M.: The Role of Trust in E-Participation: Predictors, Consequences, and Design. In: Electronic Government and Electronic Participation: Joint Proceedings of Ongoing Research, PhD Papers, Posters and Workshops of IFIP EGOV and EPart 2015. vol. 22, p. 3. IOS Press (2015)