

Álvaro Rocha
Carlos Ferrás
Carlos Enrique Montenegro Marin
Víctor Hugo Medina García *Editors*

Information Technology and Systems

Proceedings of ICITS 2020

Álvaro Rocha · Carlos Ferrás ·
Carlos Enrique Montenegro Marin ·
Víctor Hugo Medina García
Editors

Information Technology and Systems

Proceedings of ICITS 2020

Editors

Álvaro Rocha
DEI/FCT
University of Coimbra
Coimbra, Portugal

Carlos Enrique Montenegro Marin
Facultad de Ingeniería
Universidad Distrital Francisco José de
Bogota, Colombia

Carlos Ferrás
Facultad de Geografía
University of Santiago de Compostela
Santiago de Compostela, Spain

Víctor Hugo Medina García
Facultad de Ingeniería
Universidad Distrital Francisco José de
Bogota, Colombia

ISSN 2194-5357

ISSN 2194-5365 (electronic)

Advances in Intelligent Systems and Computing

ISBN 978-3-030-40689-9

ISBN 978-3-030-40690-5 (eBook)

<https://doi.org/10.1007/978-3-030-40690-5>

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland



Building a Web Tracking Browser Information System: The Online Panel as a Research Method in Internet Studies

Filipe Montargil¹(✉), Branco Di Fátima², and Cristian Ruiz¹

¹ ESCS (School of Communication and Media Studies), Lisbon, Portugal
fmontargil@escs.ipl.pt

² CIES/IUL (Centro de Investigação e Estudos de Sociologia), Lisbon, Portugal

Abstract. Internet research requires internet-based research methods and tools. Since the commercial adoption and the rapid expansion of internet use in civil society, social scientists have tried to understand the impact and contribution of this technology in the shaping of contemporary human existence. However, academic research has been based mainly in mature and traditional data collection and analysis methods, previously established, like survey research or social network analysis. Some authors suggest that this trend represents a challenge for the social sciences' ability to champion innovative methodological resources, underlining the need to create also native digital methods. The authors developed, considering this challenge, an online panel of internet users – a method already used in the market research industry to monitor and characterize audiences and consumer behaviour, but not usually explored in social science academic research. Following a case study approach, the current paper presents the information system and technological infrastructure developed to support this online panel.

Keywords: Information system · Online panel · Internet Studies · Internet use · Native digital methods

1 Introduction

The internet is today an essential part of contemporary life, mostly in developed and middle-income countries. We use it now, in our everyday life, for almost everything, from looking up for information, entertainment, communicating with family, friends or people with common interests, to buying products and services.

With an amazing amount of information produced every second it is easy to assume that, one way or the other, this technology is changing human habits and social practices. This called the attention of social scientists who, since the very beginning of commercial use of the internet, began to study its different manifestations in society, shaping an area that many refer now to as *Internet Studies*.

With the growing interest of several scientific areas, quickly emerged the need to find adequate data collection and analysis methods for this field. Two data collection methods have been widely used and are probably becoming central, in the field: on one side, we can find data collected through self-report methods, relying on the individual's

own report of their opinion or behaviour (through surveys, individual qualitative interviews, focus groups or diary studies, for instance) and, on the other side, data obtained online and treated through social network analysis. Both have their roots in more traditional and conventional research areas, in the social sciences.

Some authors suggest, however, that the proliferation of other sources of data ('social' transactional data) and resources for its analysis is challenging the social sciences' ability to champion innovative methodological resources [1, 2].

The LLMCP (Living Lab for Media Content and Platforms), a research project from ESCS (School of Communication and Media Studies), based in Lisbon (Portugal), intends to contribute to this debate and to develop a native digital method using an online panel of internet users, which is already used in the industry, to monitor and characterize audiences and consumer behaviour [3], but not usually explored in academic research.

Following a case study approach, the current paper aims to present the information system and technological infrastructure developed to support this online panel, in order to describe and revisit a year of development in our exploratory research (September 2018–September 2019).

2 Internet Studies and Online Panels

Many have been the initiatives to develop methods for the Internet Studies in Social Sciences, with different areas working in new concepts and perspectives. In media studies, for example, several authors have carried out efforts to develop methods adapted to the analysis of Computer-Mediated Communication (CMC) [4–6].

Rogers [5, 6] identifies the web, websites, hyperlinks, search engines, wikis and social network sites as entirely digital objects, with its specific ontology, and proposes the expression *native digital methods* to identify methodological resources that are originally conceived and developed in the digital context, as opposed to existing methods that are translated to online spaces.

Another reality can be found in the industry, with profit-oriented research, not aiming to understand, discuss and explore social phenomena, but instead to describe and predict consumer behaviour, in order to make more efficient and profitable business decisions.

The online panel methodology was developed in the internet business landscape, as a consequence of the convergence of the audience measurement industry with the new digital era. According to Kent, "In market research, a panel is a representative sample of individuals, households or organizations that have agreed to record, or permit the recording of, their activities or opinions in respect of an agreed range of products, services or media-use behaviours on a continuous or regular basis" [7] (p. 10). While it is true that this modality already existed before the study of media audiences, like television and radio, and that there are digital panels that follow the same classical

model, it is also true that there is an entirely digital approach in which a group of users accept to be monitored, over a period of time¹.

Our goal, with the LLMCP – Living Lab on Media Content and Platforms project, is to bring this method and the corresponding resources to academic research.

3 The LLMCP Information System

By the end of 2018, the team developed an information system as an organizational structure for an online panel of Internet users.

The project consortium was led by ESCS - School of Communication and Media Studies, from the Lisbon Polytechnic Institute, and included also the Aveiro University, the Leiria Polytechnic Institute, the Santarém Polytechnic Institute and Innovation Makers, a private company operating in the IT sector. The project received public funding through FCT - Fundação para a Ciência e a Tecnologia, the Portuguese national science foundation.

According to the project's cost-benefit assessment, a Google Chrome extension was the most efficient solution to collect systematically data from the web browsing history, in this exploratory stage [8].

Despite of its limitations, that must be considered natural, mostly in a period of development of native digital methods, Chrome is the most used browser by the Portuguese and has the largest market share in the global browser industry [9].

In the first stage of the project, the team looked for similar experiences that could support the panel's methodological decisions. The development of the system was planned in three major stages: (i) project design; (ii) development of user form and; (iii) computer programming of the extension. The development of the system took about two months, which also featured information collection tests [10]. After several tests, the system was launched in October 2018, initiating recruitment of members in November and regular and ongoing data collection in January 2019.

The recruitment process works in the following sequence: the candidate to participate in the panel applies through a form in a HTML page, with JavaScript and CSS components, divided into four main sections, (i) personal information, (ii) labour situation, (iii) issues related with internet access and (iv) account authentication data (See Fig. 1). The registration creates a REST (Representational State Transfer) to the server developed in Java, that at the same time processes information and communicate with a database in MySQL and Linux.

¹ See, for example:

Nielsen: <https://computermobilepanel.nielsen.com/cmp/landingeng.jsp>.

ComScore: <https://www.comscore.com/comScore-Panel-Services>.

Marktest: <http://netpanel.marktest.pt/>.

This protocol guarantees adequate storage of information and allows researchers to access the platform through a back office. Then, after a login operation with credentials (user/password) a new REST call is made, that validates information of the admin. If the procedure runs correctly, the system generates a token that gives access, during a limited period of time, to the back office. To guarantee the sample control, essential in social sciences [11], every application must be analysed by members of the research team.

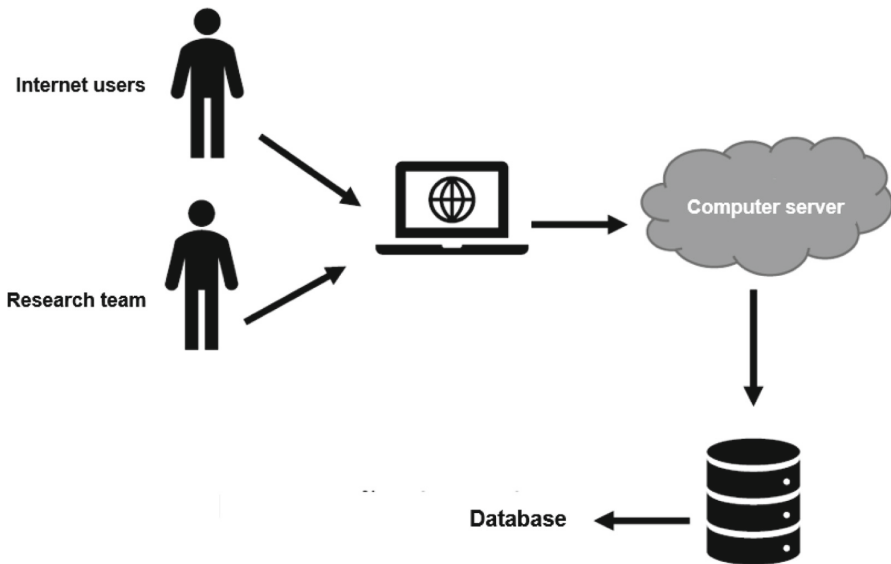


Fig. 1. LLMCP's information system

After login into the system, it is possible for the admin to transfer stored information through an .csv output file, containing data resulting directly from real time monitoring of web navigation actions performed through Google Chrome.

The system also allows access to the participant list and to the creation of subsets by gender, age, civil status or other available variables. This list is located in the back office, developed in HTML, JavaScript and CSS (Fig. 2) and also available for download, through an .xls output file.

In case of approval, the candidate receives an email with a welcome text in Portuguese (“*Obrigado pelo vosso registo. Foi aceite para participar do painel de utilizadores da Internet. Por favor, instale a extensão para Google Chrome, autentifique o email e password para fazer parte da nossa investigação. Em caso de dúvida, não hesite em contactar-nos*”) and a link to download the extension, from the Chrome Web Store.

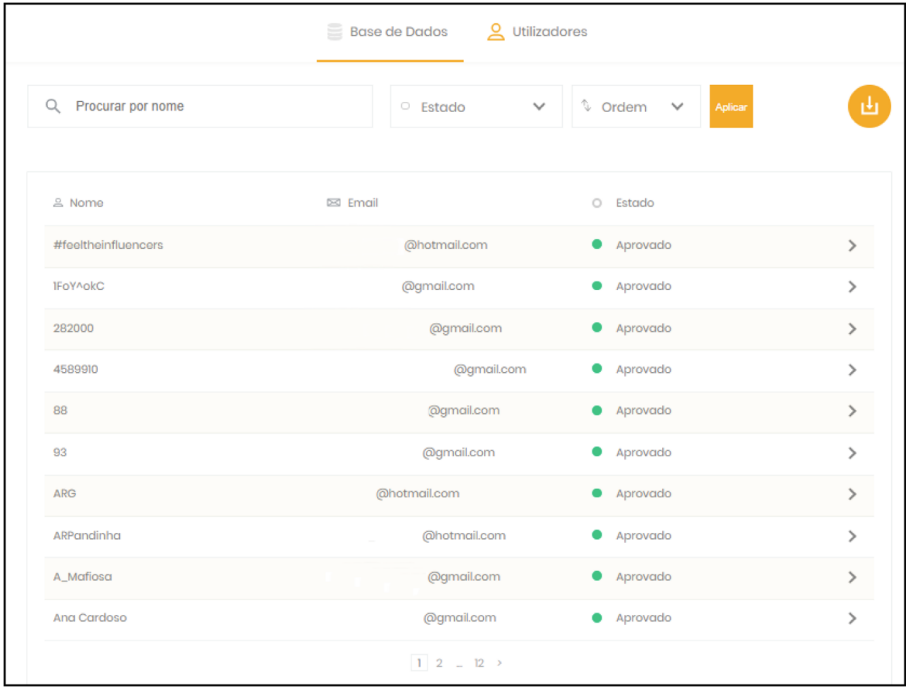


Fig. 2. Back-office (user management area)

In case of rejection, the research team will prepare a personal email, explaining the reasons leading to the decision.

The user is required to login, through the Google Chrome extension, with the chosen username and password, after installing the extension. When login is conducted successfully, as in the above-mentioned case of the admin, the system produces a token that will work until the user logouts.

Whenever active, the extension will connect with the API (Application Programming Interface) of Google Chrome. From this point, with a JavaScript communication, it will track, extract and save information regarding web navigation actions in the database. Otherwise, if deactivated, there is no way to collect data (Fig. 3). Web tracking can be suspended by the panel member at any moment, with a simple click in the user app interface. The system, including the extension and the application form, comply with the General Data Protection Regulation (GDPR), in operation since May 25, 2018, in the European Union.

The data transfer service sends a REST request to the server, that performs a search in the MySQL database. This procedure, as mentioned above, creates an output file in. csv (Comma Separated Values). Every archive stores 50.000 web navigation actions, in chronological order, with approximately 15 mb.

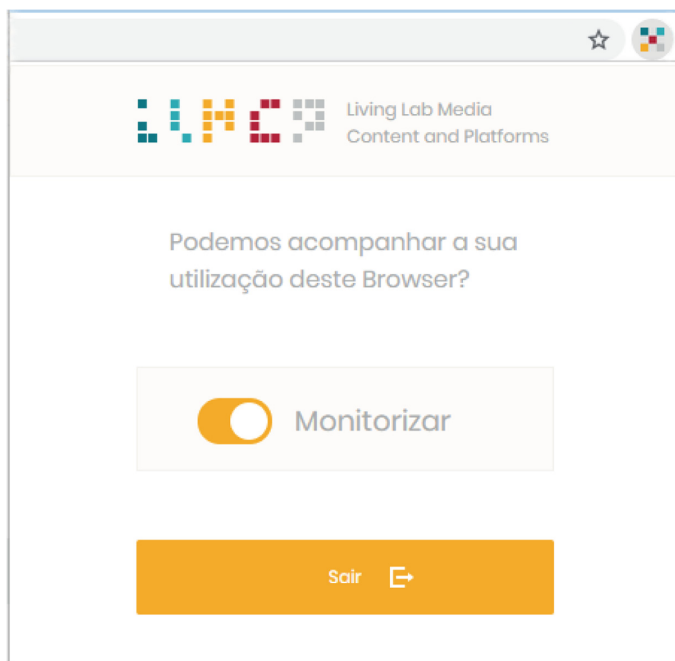


Fig. 3. Web tracking suspension user interface

In the database, every navigation action represents a website, with an URL, tab number, page activity, content date and time of access. To reduce communication time with the server and minimize timeout errors, usually present in large database extractions [12], files are zipped and sent by email. Under normal connection to broadband Internet, the extraction process takes no more than 5 min.

By the end of September 2019, the panel had 165 registered users, the majority being students from ESCS. From these registered users, 90 have been tracked and monitored at least once. The main database, storing 780.000 web navigation actions, occupies approximately 212 mb, in .csv file format.

4 Conclusions

The project team developed an information system to support the deployment and operation of an online panel of internet users, presented here. This effort brings methods from panel research, in the market research industry, to social science research and has the major goal to contribute to the creation of native digital methods.

Considering the accumulated know-how, experience and knowledge, as a consequence of the development of the information system presented here, the team considers there is space and potential to develop new methods to further explore and understand our online behaviour.

Our knowledge about internet use doesn't need to be reduced to information collected exclusively through self-report methods, on one side, or based only in global aggregate data, on the other side. There is room to explore new methods, providing the scientific community with information that can be used to analyse effective online behaviour and web navigation actions performed by individuals. And these methods can, obviously, be combined with self-report methods, in order to further explore representations, opinions or individual motivations.

As next steps, the team expects to explore the available data and to develop the existing information system, improving some of the existing limitations and shortcomings.

References

1. Savage, M., Burrows, R.: The coming crisis of empirical sociology. *Sociology* **41**(5), 885–899 (2007)
2. Savage, M., Burrows, R.: Some further reflections on the coming crisis of empirical sociology. *Sociology* **43**(4), 762–772 (2009)
3. Kumar, R., Tomkins, A.: A characterization of online browsing behavior. In: *The 19th International Conference on World Wide Web*, pp. 561–570. ACM, Raleigh (2010)
4. Jones, S.G. (ed.): *Doing Internet Research Critical Issues and Methods for Examining the Net*. Sage, London (1999)
5. Rogers, R.: *Digital Methods*. MIT Press, Cambridge (2013)
6. Rogers, R.: *Doing Digital Methods*. SAGE, London (2019)
7. Kent, R. (ed.): *Measuring Media Audience*. Routledge, New York (1994)
8. Montargil, F., Miranda, S., Di Fátima, B.: Medir a sociedade de informação: sistema para um painel online de utilizadores da Internet. In: *Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–4. IEEE Xplore Digital Library, Coimbra (2019)
9. Statcounter: Top desktop, tablet, and console browsers per country. Global Stats Tool, New York (2018)
10. Montargil, F., Rodrigues, V., Di Fatima, B.: Sistema de informação para um método digital em ciências sociais. In: *Lisboa, 19.^a Conferência da Associação Portuguesa de Sistemas de Informação –CAPSI'2019* (2019)
11. Ander-Egg, E.: *Introducción a las técnicas de investigación social*. Editorial Humanitas. Buenos Aires (1978)
12. Dai, T., et al.: Understanding real-world timeout problems in cloud server systems. In: *Conference on Cloud Engineering (IC2E)*. IEEE Xplore Digital Library, Orlando, pp. 1–11 (2018)