

Big Data for use in Psychological Research

El uso del Big Data en la investigación en Psicología

Jorge Mauricio Cuartas Arias^{1*}

¹ Faculty of psychology, Universidad de San Buenaventura, Medellín, Colombia.

*Corresponding author: mauricio.cuartas@usbmed.edu.co

The treatment, organization, and manipulation of great repositories of data, or Big Data, make up a central element in the advancement of contemporary psychology. Big Data grows day by day, fed by a variety of sources including the internet, “open-source” databases, social networks and data acquired through georeferenciation or call centers, among others. Generally speaking, Big Data includes four dimensions: First, the volume or quantity of registered data (for example the 10 terabytes—10¹² information units—processed daily by Facebook). Second, the velocity, or speed of data acquisition (from web pages, social networks, e-mails, and digital applications). Third, the data’s variety. This dimension refers to both the great diversity of online and digital resources, as well as whether the data is structured (this can refer to the data’s presentation in lists and archives or if it is ordered or homogeneously defined) or unstructured (irregular data that is not necessarily stored in an ordered or relational form), such as web sites or emails. Finally, veracity has been recently included as a dimension of Big Data which evaluates both the trustworthiness of the source and its potential scope.

Big Data offers an evolutionary leap for its users following the development of Web 2.0, due not only to access to content, but also because of the potential to participate in social networks, virtual communities, blogs and online focus groups. User navigation has only been helped by the implementation of smartphones and other devices that have led to a significant increase in both the entry of and access to information. Big Data has been fortified by the elevated number of devices and interconnected mobile applications that produce an enormous quantity of data—data which, appropriately organized, can be analyzed to help benefit problem comprehension, decision-making, and the understanding of certain cultural patterns that trace the evolution of man in modernity (Chen & Wojcik, 2016).

According to eMarketer statistics, almost 64% of the population of the United States uses a Smartphone, and 12% of people who access the internet do so through a mobile device, numbers we can only expect to grow. EMarketer predicts that by 2018 the number of Smartphone users will exceed two

billion people worldwide (eMarketer, 2017).

Currently, a great number of research projects in the discipline are limited by participant sample size, particularly in the case of research on non-clinical populations or dealing with social phenomena in natural contexts or specific cultures. In this way, Big Data represents a relevant contribution to psychometrics by providing a significant amount of data for analysis. One example of this is the treatment of human adaptive conduct and behavior that underlie patterns of cultural evolution. The inclusion of individuals in digital culture and social networks is an emerging setting for the study of behavior patterns associated with the personality and mental health of populations. In this regard, the use of robust statistical models with sample sizes that can exceed 50,000 participants favors the sensitivity of approaches to clinical problems. One example of this was the research on depression carried out by Markowetz et al. using Big Data (Markowetz, Blaszkiewicz, Montag, Switala, & Schlaepfer, 2014). Other approaches are based on the use of social networks and how people exhibit behavior patterns that are related to personality characteristics. Take as an example research by Kosinski et al. that used data from 86220 people through the social network Facebook to obtain high predictive validity in both clinical matters such as the use of psychoactive substances as well as social group positions including political attitudes and physical health, among others (Kern et al., 2014; Kosinski, Wang, Lakkaraju, & Leskovec, 2016).

Many institutions are advancing alternatives to research everyday problems that affect physical and mental health. An excellent example is the data stores obtained by “Quantified Self,” an online movement that brings together thousands of people who register their data (data donors) and generate free-access electronic databases with everything from daily activity registers to biological parameters of health and illness, information on the consumption dynamics of goods and services, as well as data on nutrition, sleep, sex, mood, and other variables. QS allows for the quantification of its users’ day to day lives, and the identification of tendencies that promote their well-being.

While mobile devices weren't initially conceived of for psychological research, they currently provide an extraordinary supply of psychometric data that lets us investigate the evolution of cultural behavior patterns, adherence to psychopharmacological treatments and therapeutic groups, inclusion and social learning, and healthy lifestyles. In addition, mobile technology has enabled the use of cognitive recuperation and reserve strategies with the use of various applications. All of this takes advantage of an enormous data store for modern psychology.

References

- Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology [Journal Article]. *Psychol Methods*, 21(4), 458-474. doi: 10.1037/met0000111
- eMarketer. (2017). *Home to research and data on nearly every digital topic* (No. January 5th 2017). Retrieved from <https://www.emarketer.com/>
- Kern, M. L., Eichstaedt, J., Schwartz, H. A., Park, G., Ungar, L. H., Stillwell, D. J., ... Seligman, M. E. (2014). From "sooo excited!!!" to "so proud": using language to study development. *Dev Psychol*, 50(1), 178-88. doi: 10.1037/a0035048
- Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychol Methods*, 21(4), 493-506. doi: 10.1037/met0000105
- Markowitz, A., Blaszkiewicz, K., Montag, C., Switala, C., & Schlaepfer, T. E. (2014). Psycho-informatics: Big data shaping modern psychometrics. *Med Hypotheses*, 82(4), 405-411. doi: 016/j.mehy.2013.11.030