Evaluation of the transparency in the plenums minutes of the Water Basins in the Sao Paulo Macrometropolis (Brazil)

Beatriz Milz , Diego Braga , Pedro Roberto Jacobi

Palabras clave: Open Data - Web scraping - Integrated Water Resources Management

Abstract[1]

The Sao Paulo Macrometropolis (SPMM) is the most important urban agglomeration in Brazil, counting more than 170 municipalities (including the Sao Paulo Metropolitan Region) and more than 33.6 million inhabitants (EMPLASA 2017). Managing the water resources within this region is considered a quite complex issue (DAEE 2013).

The Brazilian National Water Law was created in 1997 (BRASIL 1997), incorporating modern water resources management principles and institutionalizing "Water Basin Committees" (WBC), which includes civil society actors on decision making. Also, the State of Sao Paulo Water Law (Sao Paulo State 1991) separates the state in 22 Water Resource Management Units (WRMU, or UGRHI in the original). Eight WRMU are completely or partially contained in the area of the SPMM: Paraíba do Sul, Litoral Norte, Piracicaba/Capivari/Jundiaí, Alto Tietê, Baixada Santista, Mogi-Guaçu, Tietê/Sorocaba, and Ribeira de Iguape/Litoral Sul.

Each WRMU have a corresponding WBC, and the WBC discuss and make decisions in plenums. The Sao Paulo State Water Law determined that these WBC plenums must be public (Sao Paulo State 1991). Therefore, the minutes of the plenums should be available to the general public. The Integrated Water Resources Management System (IWRMS, or SigRH in the original), makes available information about the Water Resources Management in the State of São Paulo publicly available through its website, including information about each WRMU and corresponding WBC.

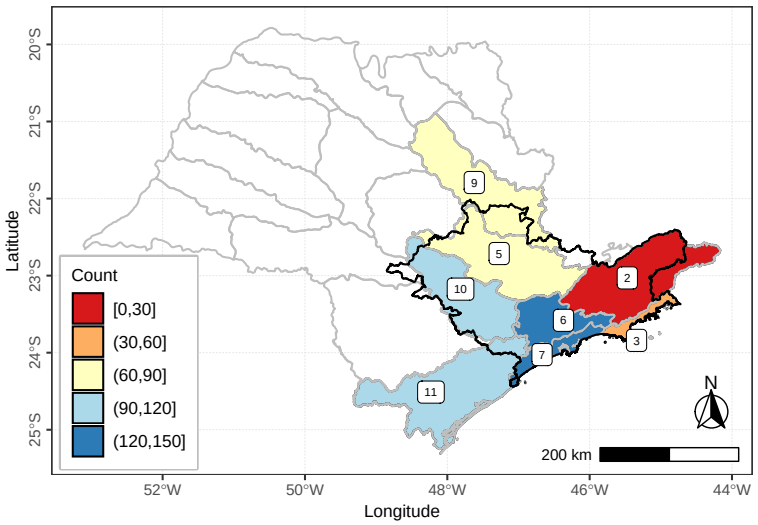
The book Principles of Open Government Data (Tauberer 2014) presents criterias that can be used to evaluate if the public data can be considered "opened". Taking the importance of the transparency of information on water resources management into account, the aim of this research is to evaluate if the plenums minutes of the WBC within the SPMM are made publicly available on the website of the SigRH[2] (Sao Paulo State 2020).

In order to evaluate whether the plenums minutes were available on the website, a technique called Web Scraping was used to collect data from websites, using the programming language `R` (2019). The packages used in all stages of this research was: `rvest`, `purrr`, `dplyr`, `tibble`, `stringr`, `magrittr`, `glue`, `ggplot2` (which are part of the `tidyverse` (Wickham et al. 2019)), `httr`, `sf`, `ggspatial`. The authors developed functions in order to collect data from the webpages of the each WRMU. With these functions, a `tibble` was created with the following information: (a) date of data collection, (b) information about the plenum (which WBC, year, name and date) and (c) information about the plenums minute (url of the link, format of file available and status code). The data presented was collected in September 2020.

Figure 1 presents a map with the counting of plenums minutes accessible through the website for each WRMU. The WRMU Paraíba do Sul did not made available any of the plenums minutes, what represents an alert of a lack of transparency. The file format in which a plenums minute is made available is important to evaluate the "Non-proprietary" and "Machine Processable" principles (Tauberer 2014). The most common file formats used were `PDF` (39.9 %), followed by Microsoft Word Document (`.doc` and `.docx`) (28.3 %), webpages (`.html` and `.htm`) (17.2%), images (`.jpg`) (13.4 %) and other file formats (1.2 %). Scanned image files are not

[1]Acknowledgments: This research is funded by **São Paulo Research Foundation** (FAPESP) (process n. 2018/23771-6). This work is part of the activities of the thematic project "Environmental governance in the Sao Paulo Macrometropolis in the context of climate variability" (process n. 2015/03804-9), financed by FAPESP and linked to the FAPESP Research Program on Global Climate Change.

[2]Portal SigRH - http://www.sigrh.sp.gov.br/



WRMU: 2 − Paraíba do Sul; 3 − Litoral Norte; 5 − Piracicaba/Capivari/Jundiaí; 6 − Alto Tietê;
7 − Baixada Santista; 9 − Mogi−Guaçu; 10 − Tietê/Sorocaba; 11 − Ribeira de Iguape/Litoral Sul

Figure 1: Map of the State of Sao Paulo and the WRMU that are part of the SPMM. Source: Elaborated by the authors based on the delimitation of the SPMM by DAEE and the packages ggplot2, message=FALSE, warning=FALSE, ggspatial, and sf. Shapefiles of the WRMU by DATAGEO. Shapefiles of the SPMM by LaPlan-UFABC, 2018.

considered a very suited file format to be processable by machines (Tauberer 2014). Also, 3.8% of the links to the files of the plenums minutes were broken (not available).

The next step is to evaluate more deeply these data based on the Principles of Open Government Data (Tauberer 2014). Moreover, it is important to perform the scrape periodically and to monitor the resulting data over time. Also, another future aim is to expand the data scraping to contemplate other possibilities, such as the information on the composition of the WBC.

References

BRASIL. 1997. "LEI N. 9.433, DE 8 DE JANEIRO DE 1997." http://www.planalto.gov.br/ccivil_03/Leis/L9433.htm.

DAEE. 2013. "Macrometrópole - Sumário Executivo - Plano Diretor de Aproveitamento de Recursos Hídricos Para a Macrometrópole Paulista." http://www.daee.sp.gov.br/index.php?option=com_content&view=article&id=1112:plano-diretor-de-aproveitamento-dos-recursos-hidricos-para-a-macrometropole-paulista.

EMPLASA. 2017. "Macrometrópole Paulista -." *EMPLASA*. https://www.emplasa.sp.gov.br/MMP.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sao Paulo State. 2020. "Portal SIGRH - Sistema Integrado de Gerenciamento de Recursos Hídricos Do Estado de São Paulo." http://www.sigrh.sp.gov.br/.

———. 1991. "Lei N. 7.663, de 30 de Dezembro de 1991." http://www.al.sp.gov.br/leis/legislacao-do-estado/.

Tauberer, Joshua. 2014. *Open Government Data: The Book*. 2nd ed. https://opengovdata.io/2014/8-principles/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.