

## About the Canadian Hansard Dataset

The transcript of Parliamentary Debates ("Hansard") is a 150-year running record of Canadian political history. The richness of this record presents a needle-in-a-haystack problem of enormous magnitude. At a rate of a novel's worth of reading each day, it would take 66 years to read the transcripts of the House and the Senate. It would take a further 28 years to read what was added in the interim.

In 2013, a group of political scientists, computer scientists, and historians teamed up at the University of Toronto to solve this problem. With support from the SSHRC, the NSERC, the Digging into Data initiative, the Library of Parliament, Library and Archives Canada, Canadiana.org, and Michael Mulley at openparliament.ca, a key output of this collaboration is the first machine-readable and fully searchable historical Hansard. We have linked to these data to various biographical properties of parliamentarians, including their party and gender. This corpus underpins the search interface of this website.

We are continuously expanding this corpus and features on this website. We are adding the Debates of the Senate and the transcripts of parliamentary committees. We are also integrating more and more information about parliamentarians, including information about the demographic profiles and election outcomes in their constituencies. To request data or for information about how to get involved in this project, please contact us.

For citation purposes, a full description of our project is available in the following article:

Beelen, K., Thijm, T. A., Cochrane, C., Halvemaan, K., Hirst, G., Kimmins, M., Lijbrink, S., Marx, M., Naderi, N., Rheault, L., Polyanovsky, R., and Whyte, T. (2017). "Digitization of the Canadian Parliamentary Debates." *Canadian Journal of Political Science*, 50(3), 849–864. https://doi.org/10.1017/S0008423916001165.