Appendix 1

**Methods**

*Data collection*

A questionnaire was sent to all faculty members in biology departments at the 20 highest-ranked Canadian universities based on the 2019 Times Higher Education World University Rankings and who were self-described researchers in the fields of ecology and evolution. The biology department website at each of these universities were reviewed in a random order to identify all researchers primarily conducting research in the fields of ecology and/or evolution (E&E) with a rank of assistant, associate or full professor. Adjunct professors and researchers who primarily focus on molecular biology, genetics, genomics, bioinformatics, theoretical biology, comparative physiology and paleontology were excluded.

Researchers were contacted via their institutional email and were informed about the aim of the study before requesting their participation; those who partook in the questionnaire comprising of thirteen questions were queried about positive and negative outcomes resulting from publicly sharing their data. Positive outcomes may include instances of new collaborations, co-authorship, promotions, awards and/or accolades commending respectable sharing practices. An especially useful benefit would be an increase in efficiency from adopting better data management practices (e.g., when having to revise analyses, send data to collaborators, or share data as a condition of publication). Negative outcomes include being ‘scooped’ on a paper that was planned on being published, not being offered co-authorship on a new study reusing one’s data, known data misinterpretations or misuses, significant time loss, and shaming or reputational damage from identified errors as a result of open data. In addition, we inquired about the time investment associated with data management and preparation for sharing. Specifically, we asked whether efforts needed to share data have decreased over time and whether they and their group’s research is more efficient as a result of new data management and sharing practices. For each question, participants had the option of leaving comments or expanding on their answers. These notes were used to emphasize commonly experienced concerns relating to data sharing as well as stimulate discussion on other topics worth considering for informed policy making.

Personalized emails as well as two reminders were sent at 8-day intervals to maximize the number of respondents participating in the study (the questionnaire and emails are provided in Appendix 2). For details on ethical considerations, please see the transparency declaration below.

The questionnaire was developed using the formr survey software (<https://formr.org/>) which integrates with the R statistical software for data analysis (Arslan 2017, Arslan et al 2018). Initial invitations were sent on January 20th, 2020; respondents were given until the end of the survey period on February 29th, 2020 to answer, modify or withdraw their answers if needed. We estimated the total number of potential participants (i.e. people to whom we sent a questionnaire) at 350 and anticipated a response rate of approximately 30%, hence a target sample size of approximately 100 respondents.

Upon implementing the questionnaire, we realized that some of formr’s documentation could be more straightforward; we therefore created a manual entitled “How to use formr – A beginner’s guide” (see Appendix 3) to ease the introduction to a new user to the vocabulary and logic to efficiently use the software.

*Statistical analysis*

We used descriptive statistics to quantify actual costs and benefits of open data practices across a reasonably large and representative sample of Canadian researchers (N=140), as most questions comprised of binary or categorical answers. The investigated topics pertained to trends and impacts of data sharing in academic research, personal opinions and experiences as well as general participation tendencies and support of mandatory policies.

We also used two generalized linear models (GLM) with a binomial distribution of error terms to examine the effect of gender and career stage (year PhD obtained) on whether respondents had experienced costs and benefits from sharing open data (N=138). We used a multinomial logistic regression (multinom function in the R package nnet; Ripley et al, 2016) to examine the effect of gender and career stage on the respondent’s level of support for mandatory open data policies (N=138). We treated the predictor variables PhD completion year and gender as continuous and binary, respectively. GLMs were verified with diagnostic tests in the R package DHARMa (Hartig 2017) – see archived scripts for details. We used likelihood ratio tests and type III sum of squares test for significance in the multinomial model. Non-significant interaction terms were removed from the models when extracting summary outputs for main effects. All analyses were carried out in Rv3.6.3 (R Core Team 2020).