Interest Group on Surveying Open Data Practices

DRAFT Reflections and Recommendations

For discussion at the RDA Session

Nov 8, 2020

Comments received before Nov 30 are welcomed. A final version will then be published.

Please send comments to: dobrien@idrc.ca

# Aim of the IG

The Interest Group on Surveying Open Data Practices draws attention to the growth of global and national-level surveys as a lens to characterize how researchers’ practices and perceptions toward open data are changing.

The open data landscape is transforming and we are beginning to understand the impact of policies and changes in researchers’ practice. As policies and practices co-evolve, survey data is critical to benchmarking and monitoring open data trends. It can serve to illuminate open data adoption (where, how and by whom) and help to identify the drivers of change.

Given the emergence of national and global surveys, the IG sought to:

* understand their findings and the potential of harnessing these insights to inform policy and practice; and,
* explore the potential contribution of a coordinated approach to improve the comparability and analysis of open data surveys.

# Main Messages

* Growth in surveys: Over the last five years, the number of global and national-level surveys has increased. Some of these pioneering examples have defined populations with several hundred responses to global surveys with thousands of responses. In some cases, longitudinal data are being collected (see Table 1 [here](https://www.rd-alliance.org/sites/default/files/Open%20Data%20Surveys%20IG%20Discussion%20Paper%20v1%20OBrien%20Gregoire%20Oct%202019_1.pdf)).
* Rationales vary: Publishers, academics, research networks and national agencies promoting open science have all taken an interest in surveying research practices. While they share an interest in characterizing such practices they are designed with different purposes which is an obstacle to coordination.
* Surveys are not comparable: Survey designers have not developed standardized questions and as a result, individual surveys create siloed datasets that cannot be aggregated or used for comparative purposes. (see presentation by [Hayashi](https://www.rd-alliance.org/sites/default/files/RDA%20Botswana%20Kaz%20Hayashi%20_Japan%20Survey.pptx))
* Harnessing the data for policy and practice: Surveys are being analyzed to inform the design of new products and services, and to develop open data infrastructure (see presentation by [Budroni](https://www.rd-alliance.org/sites/default/files/RDA%20Botswana%20P%20Budroni%20Austria_learning%20from%20surveys.pptx)). In Japan and some other countries, surveys are monitoring trends over time allowing sponsors to model the impact of their policies and services.
* Geographic coverage is uneven: We have an emerging picture of researcher practices but only in high income countries and in several large middle-income countries. There is a very poor understanding of daily researchers practices in most of Asia, Africa, the Middle East and Latin America. (see presentation by [O’Brien and Gregoire](https://www.rd-alliance.org/sites/default/files/Geography%20of%20Open%20Data%20Surveys%20RDA%20Helsinki%20Oct%202019%20.pptx))
* The RDA community, inclusive of researchers, practitioners, and decision-makers would benefit from a coordinated approach to track changes in practice and policy overtime.

# Insights from the IG’s Work

Aim 1: Mapping the Community

* There are multiple actors involved. Academic publishers (e.g., Elsevier, Wiley and SpringerNature) have made an important contribution to illuminating open data practices internationally. They led the first wave of surveys and re-issued them to generate time-series data. Their surveys have attracted the largest number of responses. In addition, there are a growing number of national-level surveys and these are led by different actors such as academies, university networks, research funders and gov’t agencies. Finally, there are a few academic-led surveys but these are less common.
* These global, national and sectoral efforts are generating a more comprehensive picture of research practices. The surveys in use all have unique questions even though they share a common focus. As such, the datasets cannot be meaningfully aggregated.
* The granularity, comparability and coverage of surveys needs addressing as open data practices vary widely between scientific disciplines and regions. Some of the national-level surveys have generated granular insights on research practices in individual countries and the large publisher surveys illustrate how practices differ among countries (see presentations by [Khodiyar](https://www.rd-alliance.org/sites/default/files/RDA%20Botswana%20V%20Khodiyar%20Findings%20from%20Springer%20nature%20Survey.pptx) and [Rosetta](https://www.rd-alliance.org/sites/default/files/RDA_Botswana_F%20Rosetta_Reflections%20on%20Elsevier%20CWTS%20Survey.pptx)).
* Their interests are divergent. The publishers tend to position their surveys as a public good - informing the research community of the prevalence of data-sharing across countries and disciplines. They are also being used to develop new services and policies. National level surveys tend to be more instrumental and explicit in their intent to design or evaluate national programs. They tend to provide a more detailed assessment of how researchers assess their environment and incentives.
* Research funders have steadily introduced research data management expectations and data-sharing requirements as grant conditions. Most agencies consult with the researchers they support and are interested in monitoring the adoption and impact of new grant conditions. A few national initiatives are utilizing surveys to monitor open data trends in their country and there appears to be interest from other jurisdictions.

Aim 2: Scope for a community-designed modular and interoperable open survey(s)

* A sample of 10 surveys with over 24000 responses was analyzed to understand their similarities and differences. This [analysis](https://www.rd-alliance.org/sites/default/files/Open%20Data%20Surveys%20IG%20Discussion%20Paper%20v1%20OBrien%20Gregoire%20Oct%202019_1.pdf) concluded there was no indication that recent surveys are adopting common questions. Comparable survey data remain elusive, even for the most foundational questions.
* Pilot Survey. As no emergent efforts were identified, the IG initiated a project to create a modular survey bank. This modular bank of questions is cross-referenced to existing survey questions from 17 surveys. The survey excludes open ended questions so as to enhance the cross-comparison of surveys.
	+ The questions are organized according to the components of Elinor Ostrom’s Institutional Analysis and Development Framework into the following sections: exogenous variables (biophysical, community attributes, rules in use), action arena (action situation and participants), interactions, evaluation criteria, and outcomes.
	+ Based on surveying need or qualitative research, the question bank can be used to build specific surveys for dissemination
	+ In addition to the question bank, the pilot also provides a reproducible website, template ethics form, analysis code and authorship tools to enhance the re-usability of the survey tool.
	+ A pilot of the survey tool was created for dissemination in 8 African countries (Oct-Dec. 2020). Questions were selected based on prior qualitative research and in consultation with potential users and key experts.
	+ The survey has been translated in French and Portugese to get a better reach and response.

Aim 3: Determine how such open survey(s) could be implemented and results analyzed globally

* Our analysis of ten surveys found that the datasets are findable and accessible. To compare the results of two or more surveys, extensive recoding of variables is required. This is time consuming and places limits on the comparability of survey responses.
* There is no common platform being used to bring individual surveys together to support secondary use of survey data. There is a data visualization platform called SuAVE that currently hosts [four open data surveys](http://suave.sdsc.edu/suave-for-questionnaire-surveys-in-it-and-other-fields/). SuAVE allows the user to examine and visualize responses.
* The pilot survey and website provide tools to analyze the data consistently. This approach also seeks to promote reliability in the analysis of surveys and comparability of findings, which should allow for better benchmarking.
* The IG is currently investigating options for long-term hosting of the pilot tools.

# Recommendations

## For Survey Designers

* Survey designers who are re-administering surveys are encouraged to promote comparability with other surveys.
* Designers are encouraged to publish metadata, supporting material to help understand published datasets. This might include codebooks, background information on the aim of the survey, whether the survey structure adheres to a conceptual framework and guidance on how to analyze the data. Such material would support replicability and re-use.
* Survey designers creating new surveys are urged to consult and utilize the questionnaire databank and the pilot survey. Consider building on the core questions for future surveys. If new questions are created, designers are encouraged to contribute to the question bank.

# For Survey Sponsors

* Academies, research funding agencies, university networks, and government agencies are encouraged to support efforts to better understand the open data perceptions and practices of researchers they seek to support.
* In doing so, survey sponsors are encouraged to consult the (the pilot survey and the survey question bank) and existing survey instruments. If the research community is surveyed, survey questions should be comparable.
* New surveys should follow current practice in making the datasets available for re-use.
* Agencies that have established national strategies and a monitoring and evaluation plan to assess the influence of policies and programs are encouraged to openly publish the full set of tools to support others interested in following a similar path.

# For RDA

* A uniting interest of the RDA community is to improve policy and infrastructure to support data-sharing and data-driven research. Given this interest, and the significance of activities the RDA community is involved in, RDA should invite reflection and profile efforts that seek to identify open data trends.
* This IG was not focused on how to assess the impact of open data initiatives but survey data is critical for such purposes. RDA should encourage members who have a role in supporting significant open data interventions to undertake impact assessments and make the underlying data available.
* Alongside RDA adoption cases, the RDA should explore how its members could build the evidence base on the impact of open data policies or programs. RDA experts could be identified to support agencies assess the influence of their policies and programs.
* Within the RDA, IGs and WGs have implemented surveys on open data practices. There is an opportunity for RDA to collectively organize this survey data and make it visible to members.
* For external surveys, RDA should consider establishing a standing committee responsible for disseminating new survey research findings to RDA members and promote metadata standards that would allow country- or theme-specific surveys to be findable. Such coordination may improve awareness of trends by the RDA community.