RDA 10th Plenary Montreal, Canada
BOF open survey session 19 September 2017

Topic: An Open Survey Questionnaire on Open Data

Practical information
Slides of the short introduction into the topic will be made available.
Names and e-mail addresses of participants are collected to follow up on the discussion, and send them the minutes.

Introduction

In general, the idea is that there is a multiplicity of surveys taking place. Convergence, coherence and collaboration is what seems to be lacking now. This is motivating the effort we are proposing today: 1. developing/designing a survey; and/or 2. collecting additional data; to follow up on changes in open data practice.
We envision that it should be implemented through community involvement including stakeholders globally, so that the survey is broad enough and general enough.

The rationale is that policies are aiming for full blown open data. To see change take place, it seems there are those who push and those that pull when it comes to sharing research data. The ones that pull seem to be far in the minority, which means there is a world to win.

Suggestion was made to use the word 'sharing' data instead of 'open data'. Open might put off people while sharing is a softer term. What it actually means is providing access under well-defined conditions.

The **discussion** covered the following aspects:

1) **Developing/designing a survey**
The suggestion is to collect questions that 1. can be used by everyone; be it generic questions, and 2. questions that can be specifically adapted. Suggestion was made to develop modules around specific themes/topics. Questions need to be stakeholder specific. So it is important to think about who will use it. This is not an insignificant effort. Over time there will be a set of questions that will remain, while new ones or modified ones will be added. The next suggestion is to develop this as a question bank for others to pick and choose from, rather than developing it as a questionnaire (ie. Eurobarometer question bank). By turning it into a question bank rather than a survey, we avoid pretending that this will be the mother of all questionnaires.

With regard to the questions, formulation is important, and the document should include the considerations on question phrasing. This can relate to cultural differences in language. Despite the pitfall of translation, it would be good to translate questions into national languages. This is for example what Cochrane centers do translating locally, and to always have it bilingual (English + local language). Another example was how the Canadian Research council for social sciences supported the Canadian university surveys in with translation.

Based upon current existing questionnaires, are there still burning issues that are not yet been asked? This can be explored by asking a smaller sample of people, e.g. as a pilot in research communities to get consensus on specific issues/topics by asking open end questions. It is also important to keep in mind the

technical challenges around data sharing: data is a nebulous concept, and we need to be aware of the definition of terms. In such cases a survey may affect the people who answer, so an alternative approach could be to carry out case studies to support the open end questions. This is what has been done in other previous surveys as well.

Regarding the type of research communities, there are many that do not have clear ideas about data sharing, and there is sometimes a lot of skepticism. Also in professional communities and in industry the data sharing is perceived in many directions. It is important to understand the interactions.

2) **Benchmarking survey/data collection**
The next step would be to collect more data on the researcher perspective to follow up on developments. The first suggestion is a technical point and concerns to at least have the data collection and archiving being interoperable, carrying the whole process under the FAIR principles.

Further, the question is whether the RDA should run the whole survey itself. The common feeling was not because of the large cost involved. In order to increase response rate, one could use students to call the respondents. That would be a heavy investment, but it would give a more significant picture. Currently, we hardly know the views of non-respondents so it would be worthwhile to get better insights. Also, the survey should be run in a timely fashion.

An important issue is to have clear ideas on the sourcing of the respondents. In an open survey on research data it is easy to miss out on contextual effects: research communities respond differently, and social effects can be strong. For example , if someone in your work environment is a key person who is doing data sharing that is a strong driver; nobody wants to be first. If researchers have difficulty in creating their own data pipeline, we need to know where this is taking place but we might lack the tools to identify it.

Surveys could be easily carried out by funders or other stakeholders (e.g. societies). Proper stakeholder mapping will be required. The question bank could be a useful tool for them since they are key stakeholders able to instigate change (e.g. by requiring researchers to devote 5% of their time to making data open). And from the perspective of teaching the younger ones: the survey could be useful for graduate student associations (e.g. EURODOC).
From a wider perspective, users of the survey could also be for example patient groups, or researchers in industry who are hard to reach. This would also put extra pressure on the phrasing.

3) **Analysis of results: measure policy effects**
The main purpose to run several times over time such surveys, is to measure the effect of policies that aim to increase the socio-economic benefits of sharing data. If policy interventions work to get researchers more responsive toward open data, that would be very helpful from a policy perspective. It is setting policies in the dark if we cannot measure their effectiveness. There is no real alternative, so benchmarking surveys could provide relevant metrics, bringing clarity to policy implementation effect. Some doubted whether the survey would measure the ultimate goal, namely effective data reuse.

4) **Is this a needed activity?**
The discussion also raised some doubts. Not everyone was fully convinced yet. Questions raised were: Is a single survey the appropriate methodology; Will this be the instrument to grasp the views of researchers over time, as ideas, activities and perceptions will change over time as well. There were also questions about actual feasibility. As for any survey: will it be representative enough as there are 4

important dimensions for surveys: discipline, location, time, and stakeholders.

5) **How to organize the follow up**
In RDA, a WG is typically ready in less than 18 months. What we discuss here should be available for a longer time than to wrap up in 18 month. Alternative approach could be to start an Interest Group, in which different strands could start at the same time or follow up shortly after each other.