# data management plan

# Strengthening the Capacity of the Economic Committee of the National Assembly in Integrating Inclusive Growth in Macroeconomic Policy Making and Oversight in Vietnam

### Admin Details

**Project Name:** My plan (Portage Template)

**Principal Investigator / Researcher:** Cameron Neylon

**Institution:** Portage

### Data Collection

#### What types of data will you collect, create, link to, acquire and/or record?

The project collects primary data from interviews with 760 SMEs in 12 provinces of Viet Nam. The sample was drawn from a stratified random procedure.

Throughout the 3-year period of the projects 4 rounds of the surveys will be conducted. The baseline surveys were conducted in the form of face-to-face interviews, with data entries being conducted with the rtSmartSurvey CAPI platform of Real-Time Analytics (RTA). The follow-up rounds are done on phones with data being entered to the electronic questionnaire while an interview is being conducted.

In short, the project generates data from primary data collections over 760 SMES. The data collections are rolled out into 4 rounds over the course of 3 years

#### What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

Data are first stored in MySQL Database system of rtSmartSurvey platform. From there, the data are exported to Stata and CSV formats for analysis purposes.

For reuse, sharing and long-term access, Stata and CSV formats are highly relevant.

#### What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

For activities under the project, we mostly work with data in Stata format. Stata data files has an embedded meta-data system, that includes variable labels and data set labels. The labels help explain what the variables are, when the dataset was created and (last) modified.

To help data analysts even further, we name the variables following the question numbering in the questionnaire.

### Documentation and Metadata

#### What documentation will be needed for the data to be read and interpreted correctly in the future?

The most useful documentation is the codebooks of the datasets. The codebooks provide lots of meta data such as the total numbers of variables and observations; for each variable, it gives fundamental summary statistics and examples of data values.

The second documentation is technical notes such as sampling design note, data entry system and fieldwork plans.

The third documentation is a collection of descriptive analyses of the data sets.

#### How will you make sure that documentation is created or captured consistently throughout your project?

There are three major mechanisms for ensuring consistency of the documentations throughout the project, including:

* Automating the creating of documentation whenever it’s possible: we rely on rtSmartSurvey to generate the codebooks automatically
* Thorough reviewing of written reports and documentations. By being thorough I mean we get the papers reviewed by multiple staff in several rounds.
* Continuous revising: whenever there is a comment over the data that needs a revision, we go back to the dataset and conduct it. Then, an update would be added to the documentations.

#### If you are using a metadata standard and/or tools to document and describe your data, please list here.

For the codebooks of the data sets we use the Stata in-built tools to translate from log files to PDF.

For other documentations, they are just reports/papers written in MS Office formats.

### Storage and Backup

#### What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

Since most data are in text format, all data storage is expected to be less than 1 Gigabyte.

We envision that we will store it for at least 10 years.

#### How and where will your data be stored and backed up during your research project?

We store the data sets in our staff’s computers and a server operated by Real-Time Analytics.

#### How will the research team and other collaborators access, modify, and contribute data throughout the project?

Currently, we share the data sets with the research team and other collaborator via dropbox sharing and emails.

### Preservation

#### Where will you deposit your data for long-term preservation and access at the end of your research project?

We are considering developing an API (Application Program Interface) for accessing the data and hosting it at RTA’s website at: www.rta.vn.

#### Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

As described above, we think API-based protocol is best for this requirement.

### Sharing and Reuse

#### What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

We will share the final data sets as open access (with registration required) and the raw data upon request.

#### Have you considered what type of end-user license to include with your data?

Not yet. We would be keen on learning relevant options and choose one that fits the best.

#### What steps will be taken to help the research community know that your data exists?

There are several steps we can do, including:

* Mentioning the free access policy in the project’s papers and reports
* Advertising the policy during workshop occasions
* Advertising the policy on websites of RTA and other project partners (CAF and ECNA)

### Responsibilities and Resources

#### Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

It shall be Real-Time Analytics as the institution and RTA’s rtSolutions Team as staff who will be communicating on data access matters.

#### How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

Since we work as a team we have staff backup plan on a daily basis. I don’t expect major changes when a staff who is directly involved in the project leaves.

#### What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

The most resource consuming tasks are the development of the documentations, the development of API system and the continuous maintaining of staff who provides users with supports and feedback.

We will discuss with the team and develop a proposal for these tasks.

### Ethics and Legal Compliance

#### If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

We will trim all the personally(?) identifiable information/variables from the data sets before we release to public users. All observations and data points are anonymous.

#### If applicable, what strategies will you undertake to address secondary uses of sensitive data?

The use of sensitive data is strictly within Real-Time Analytics for the purpose of following up with the respondents.

#### How will you manage legal, ethical, and intellectual property issues?

We haven’t thought much about this line of issues yet.