Database Access and Data Sharing Policy

Nith Kosal

Future Forum

April 1, 2021

What is the Data Hub?

- ▶ New initiative by Future Forum in 2021
- ▶ The propose is to provide the bibliographic for Cambodia's database dedicated to public policy, economics, development, and other social science
- ▶ Available freely for Future Forum staff and its research community
- ▶ Unfortunately, more dataset that we have can not share you because of the regulation of data producers

Who Can Access to database at Data Hub?

- ▶ All Future Forum staffs, young researchers and membesr of Future Forum research community can access Data Hub repository

- ▶ With *DataHubREADME* will guide you through a summary and list all the microdata that can found in the Data Hub repository.

- ▶ All requesters should follow all terms of data condition

- ▶ Contact me directly via `nithkosal@futureforum.asia` or by telegram @NithKosal (+855) 12 340 226

- ▶ I and Future Forum management team have a right to evaluate your application and send you back at last 5 business days

- ▶ I will take a copy for you through a USB or send you by email.

Terms of Dataset Use

Terms of Use

1. Open Data Access
   - Free to copy, distribute, adapt, display or include the data in other products for commercial and noncommercial purposes.
2. Direct Data Access
   - You can download free without any application form to request the dataset but cannot share it.
3. Public Use Files
   - Request the dataset with a strong application. Should go direct to the primarily producer or the institute have a right to share.
4. Licensed Dataset
   - Request with a strong application of research project and expected output to Data Hub or go directly to the producer.
5. Extental Dataset
6. No Data Access

Direct Data Access and Public Use Files

- ▶ Generally, the data producer and the secondary parties that have the right to share datasets have the role of managing the use of data by users

- ▶ They want to facilitate and manage all tasks related to their data

- ▶ Future Forum does not have the right to share and does not assume any responsibility you do not respect their policies

- ▶ To do this, you must create a user account to request the dataset from their repository.
  - ■ Nada NIS, The Bank Microdata, The Bank Enterprise Survey, IPUMS, DH Survey, ICPSR, Harvard-MIT Dataverse, AEA Journals, ILO Survey, researcher website, and others

Application for Access Dataset

You can download the application form here!

- ▶ First Name and Last Name

- ▶ Organization

- ▶ Email

- ▶ Dataset requested

- ▶ Intended use of the data: Project question, objectives, methods, expected outputs, and partners

- ▶ List of expected output(s) and dissemination policy

- ▶ Expected completion date

- ▶ Research team members

- ▶ Agree with the conditions

Terms of Licensed Data Use

▶ Data and other material will be used solely by the requester, and shall not be redistributed

▶ Any data created on the basis of the original data will not be copied or made available to anyone other than those mentioned

▶ The data will only be processed for the stated statistical and research purpose

▶ No attempt will be made to identify respondents or microdata providers

▶ No attempt will be made to create links between datasets provided by the Data Hub, or between and other datasets that could identify individuals or organizations.

▶ Any books, articles, or other publications that employ data obtained from the Data Hub will cite the source of data

Terms of Licensed Data Use

▶ An electronic copy of all reports and publications based on the requested data will be sent to the Data Hub

▶ Data Hub bears no responsibility for the data's use or for interpretation or inferences based upon it

▶ Any violation of this Terms of Use will be considered to have occurred on behalf of the user's organization, and we will take appropriate measures to sanction such misconduct, which may include denying any further and future access by any user from that organization

▶ The intended use of the data, including a list of expected outputs and the organization's dissemination policy, must be identified and no different uses shall be permitted without prior written consent from the Future Forum

Introduction to the Statistical Software

▶ **Excel:** A wide variety of tools for data visualization and simple statistics.

▶ **SPSS:** Ability to build visualizations and reports with simple point and click actions without coding. Quick identification of invalid and missing data.

▶ **Stata:** Thousands of built-in Bayesian model, regression, generalized and choice models.

▶ **EViews:** Use for statistical, forecasting, and modeling tools. The best for time-series dataset.

▶ **R:** Free statistical package that is widely used for social science research and other fields.

Introduction to the Statistical Software

- ▶ **MATLAB:** For mathematics, statistics and machine leaning. Many macroeconomists like it and useful in plot data.

- ▶ **Wolfram Mathematica:** For image processing, geometry, visualization, machine learning, data mining. Provide the mathematical formula.

- ▶ **Python:** High-level and general purpose programming language.

- ▶ **SAS:** For big data analysis, data management, advanced analytic, multivariate analysis, business intelligence.

Introduction to RStudio

Introduction to RStudio

▶ R traning docs by CDRI: `https://drive.google.com/drive/folders/1rEcYfpLMp51hQtBRIBDqsczveQqceRnV?usp=sharing`

▶ Intermediate level: R for Data Science by Hadley Wickham and Garrett Grolemund. Available online: `https://r4ds.had.co.nz/`

▶ Advanced R by Hadley Wickham. Also available here: `https://adv-r.hadley.nz/`

▶ R Blog: `https://www.r-bloggers.com/`

▶ Share your knowledge and ask questions: `https://stackoverflow.com/questions/tagged/r`