**ReadMe << Name of dataset/project >>**

|  |
| --- |
| 1. ***Short description of dataset*** |
| Project abstract  […]  Data generation  […]  Content of dataset  […]  For a detailed overview over the variables in the dataset see section 7 of this document. |
| Unit of analysis: **… (e.g. patent level data)**  Data sources: **…**  # of observations: **…**  # of variables: **…**  Size: **… kB/mB** |

|  |
| --- |
| 1. ***Keywords*** |
| … |

|  |
| --- |
| 1. ***Contributors*** |
| surname, first name (e-mail)  …, …  …, … |

|  |
| --- |
| 1. ***Associated publications and working papers*** |
| **XXXXXXX, X. (YEAR): *„…..“*, ….** |
| … |
| … |
| ***Note: Please cite publication in bold when publishing work based on this dataset*** |

|  |
| --- |
| 1. ***Acknowledgement*** |
| Users of the dataset should acknowledge data access with the following sentence:  ***We thank the Max Planck Institute for Innovation and Competition (Munich, Germany) for providing access to their data.*** |

|  |
| --- |
| 1. ***Computational requirements*** |
| INSTRUCTIONS: In general, the specific computer code used to generate the results in the article will be within the repository that also contains this README. However, other computational requirements - shared libraries or code packages, required software, specific computing hardware - may be important, and is always useful, for the goal of replication.  **Software**  INSTRUCTIONS: List all of the software requirements, up to and including any operating system requirements, for the entire set of code. It is suggested to distribute most dependencies together with the replication package if allowed, in particular if sourced from unversioned code repositories, Github repos, and personal webpages. In all cases, list the version you used.   * OS + version * Software versions * Libraries used   **Hardware requirements**   * Processor * Memory * GPU |

|  |
| --- |
| 1. ***Running time*** |
| INSTRUCTIONS: Memory and compute-time requirements may also be relevant or even critical. Some example text follows. It may be useful to break this out by Table/Figure/section of processing. For instance, some estimation routines might run for weeks, but data prep and creating figures might only take a few minutes. **Summary** Approximate time needed to reproduce the analyses on a standard (CURRENT YEAR) desktop machine:   * <10 minutes * 10-60 minutes * 1-8 hours * 8-24 hours * 1-3 days * 3-14 days * >14 days * Not feasible to run on a desktop machine, as described below. |

|  |
| --- |
| 1. ***Instructions to run the replications*** |
| INSTRUCTIONS: The first two sections ensure that the data and software necessary to conduct the replication have been collected. This section then describes a human-readable instruction to conduct the replication. This may be simple, or may involve many complicated steps. It should be a simple list, no excess prose. Strict linear sequence. If more than 4-5 manual steps, please wrap a master program/Makefile around them, in logical sequences.  Instructions could be:   * Edit 02\_scripts/00\_main.do to adjust the default path * Run 02\_scripts/00\_main.do to run all steps in sequence. * **…** |

|  |
| --- |
| 1. ***Folder directory*** |
| |  |  |  | | --- | --- | --- | | **Folder** | | **Content** | | 01\_data | 01\_orig | Input data for processing datasets | | 02\_external | … | | 03\_intermediate |  | | 03\_final | … | | 02\_scripts | 00\_main.do | Master-Do-File (running this activates the whole process) | | 01\_... | Do-files for processing datasets | | 02\_... | Do-files for analyzing datasets | | 03\_... | ... | | 04\_... | Do file for making tables and graphs | |  |  |  | |  |  |  | |  |  |  | | 03\_results | 00\_log |  | |  | 01\_tables | Final tables for publication | |  | 02\_figures | Final figures for publication | | 04\_publication |  | Associated publications and working papers | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  |   *Ideally, the project should be all-inclusive and runnable on its own. If this is not the case, indicate data source that are coming from somewhere else.* |

|  |
| --- |
| 1. ***List of variables (e.g. from Stata)*** |
| 1. … data  |  |  | | --- | --- | | **Variable name** | **Description** | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  |   *What information is needed for an external researcher (or the future you) to understand the contents of the data? If standard data set are used, a reference to documentation is of course OK.* |
|  |
|  |
|  |