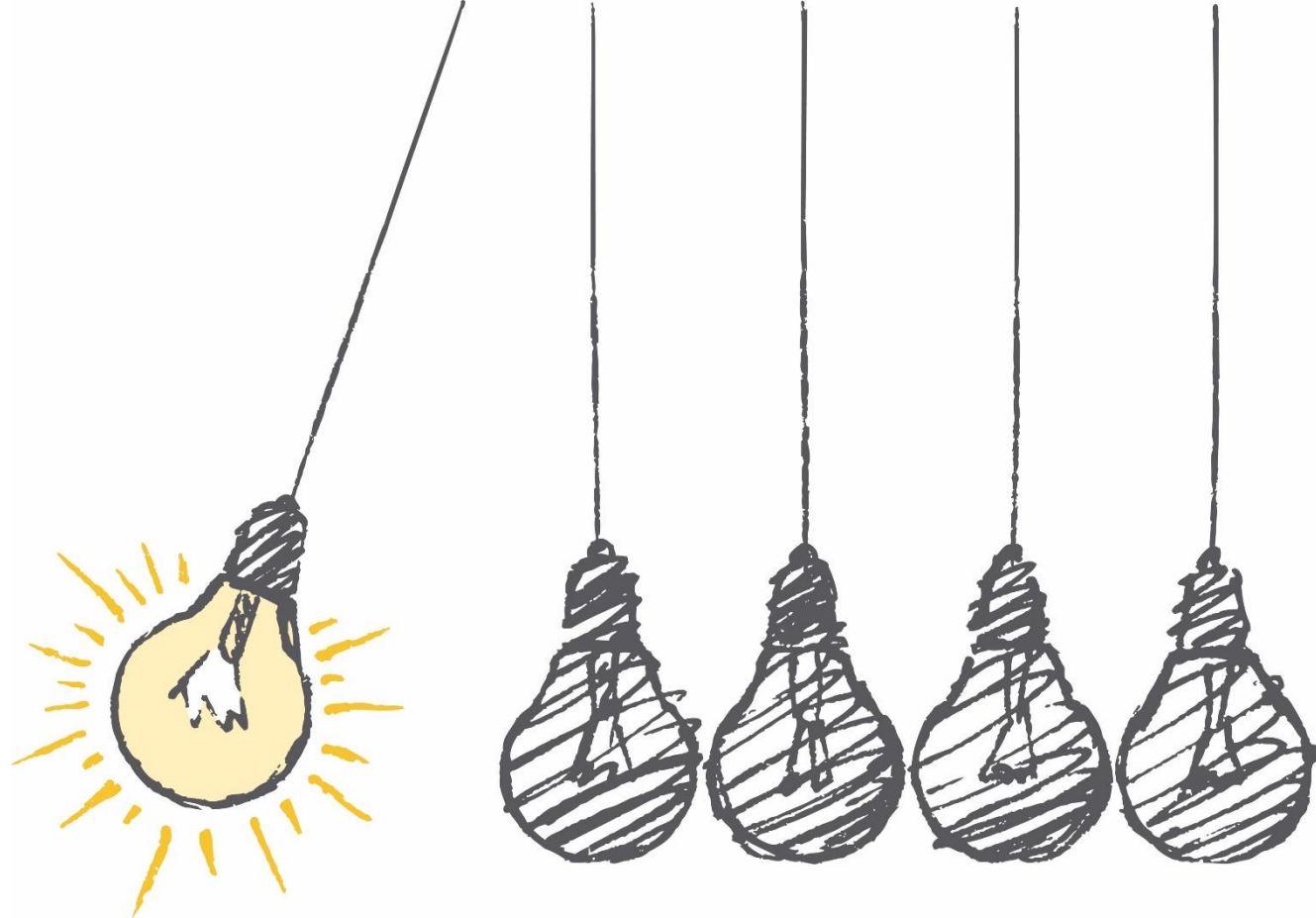


Data Publication: Microdata Catalog

Prepared by **DIME Analytics**
dimeanalytics@worldbank.org

Presented by **Roshni Khincha**
rkhincha@worldbank.org



WORLD BANK GROUP



Norad

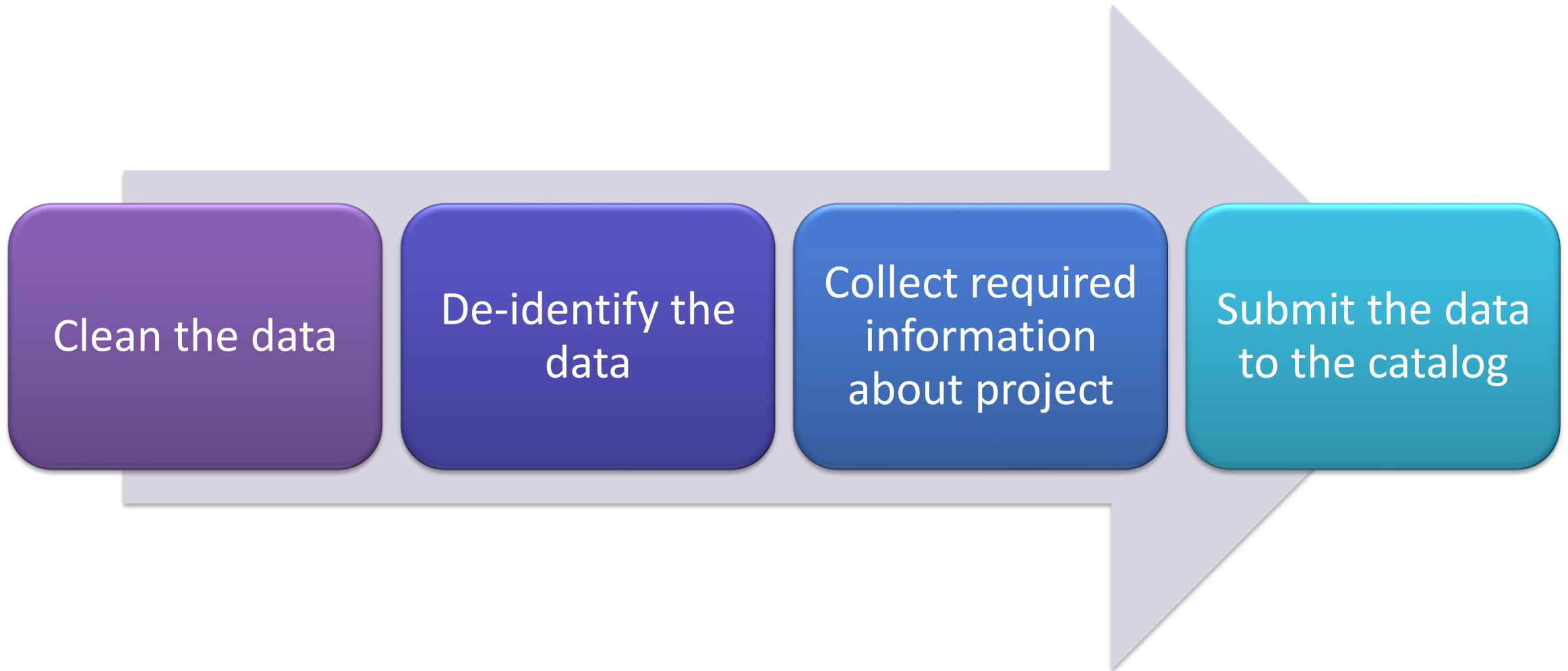
Microdata Catalog

- An online platform offering free access to microdata produced by World bank and other organizations
- Recommended to publish data collected on IEs on the catalog as
 - Transparency
 - It creates a repository of the data
 - Helps ensure your dataset has no identifiers
 - The Microdata team has a through eye and ensures all PII are removed before publication
- Mainly for survey/primary data

Some specific cases

- Data collection was paid for by the government
 - Discuss data use agreement with the counterpart
 - Check if analysis dataset can be published
- Secondary data
 - Secondary data can be published depending on the agreement with the counterpart
 - Check if analysis dataset can be published
 - At least upload the metadata on Development Data Hub
- Web-scraped data?
 - Provide open access to the scraping code (eg: via GitHub)

Preparing data for publication



Cleaning the data

- Ensure each observation has a unique ID
 - i.e no duplicates
- Label all variables and values
- Code missing values correctly
- Ensure variable names match the questionnaire
- DO NOT construct any variables in case of primary data

De-identifying data

- Drop identifying information
- Encode potential identifiers and drop labels
- Mask GPS coordinates

Make use of resources mentioned earlier

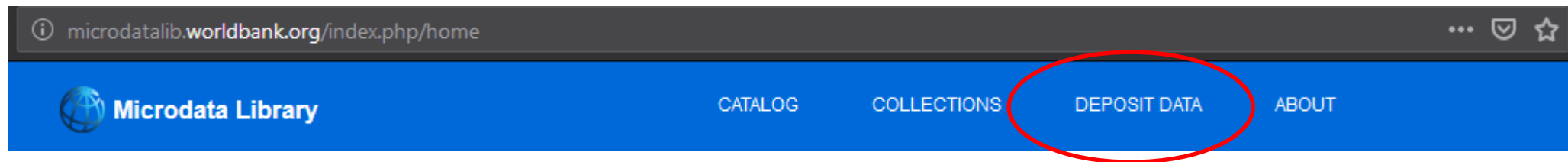
- *PII Scan*
- *iecodebook*
- *SdcMicro*

Collect required information and documents

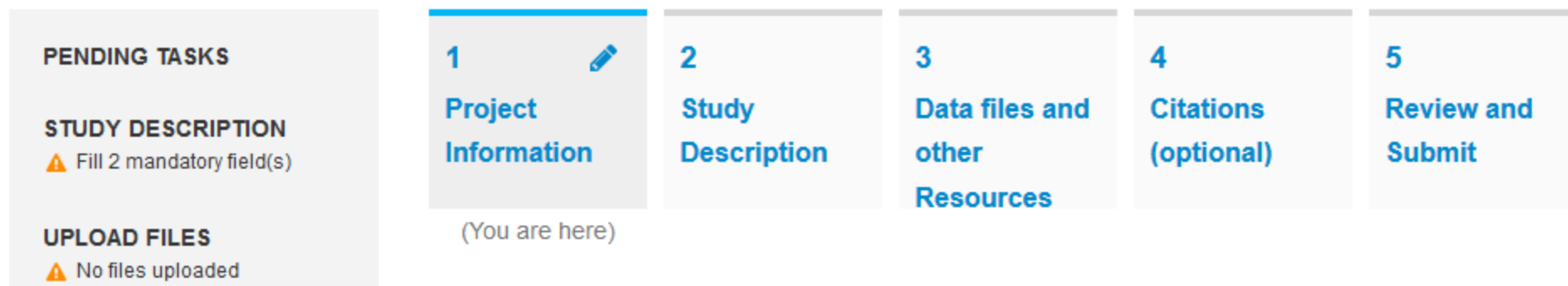
- Mandatory information
 - Title; Country; Dates of data collection; Access policy; Catalog to publish in
- Recommended information
 - Abstract; Geographic Coverage; Primary Investigators; Funding; Sampling Procedure; Weighting
- Documents
 - Questionnaires (paper format equivalent is better than CAPI form)
 - Enumerator manuals
 - Data cleaning documentation
 - Outputs such as reports, presentations, publications and papers

Your data is ready to be published. Now what?

- Go to microdatalib.worldbank.org and select deposit data
- Select deposit data



- Log in to the portal (requires RSA token to log in)
- Select 'Create New Project' and fill in the sections shown in the screenshot below



Some additional pointers

- Best practice to publish data soon after data collection is complete
- Use the [checklist](#) to ensure all steps are completed
- Share as much information about the project as possible
- [Type of access](#) suggested: *Limited/Licensed*
 - Anyone who wants to access the data has to reach out to the research team and sign a data-use agreement
- Access can be *embargoed* for given time period
 - In case your project doesn't want the data to be published immediately

Thank you!

