

PROBLEM STATEMENT

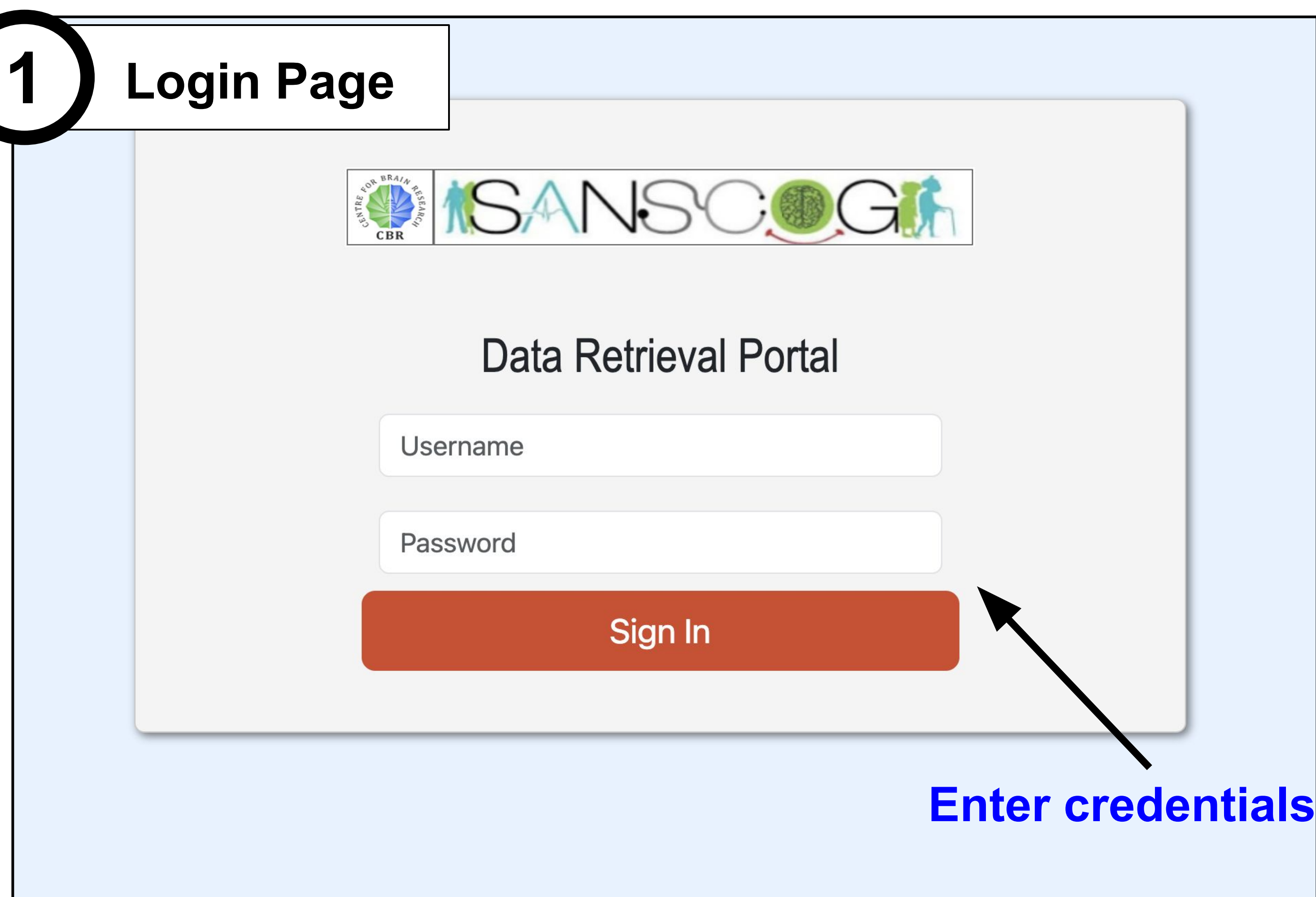
As of May 2023, the SANSCOG Dataset comprises of different categories of data including **Socio-Demographic, Clinical, Nursing, Cognitive, Blood Biochemistry, MRI and ApoE**. Combined, they account for **4241 columns**. This makes data extraction a tedious task for the researchers who want to leverage this huge corpus. Hence, there is a burning need for an extraction procedure which is easy to use, intuitive and saves hours of manual work.

METHODOLOGY

We create a web based application based on Flask, which is an open source Python package for backend development. The frontend was created using HTML/CSS and Bootstrap. The web app loads the source dataset in a dataframe and dispenses the required columns as requested by the user.

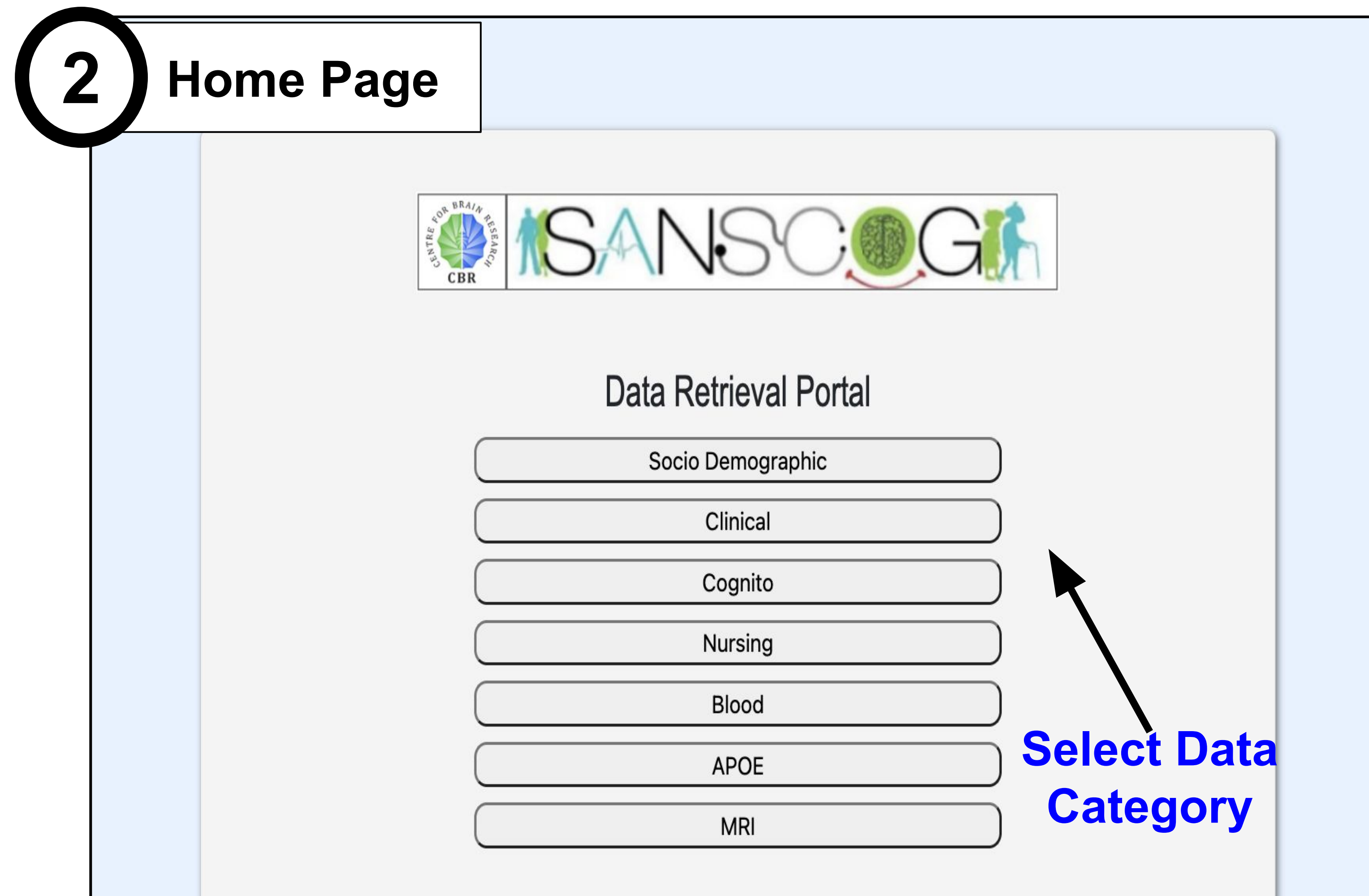


1 Login Page



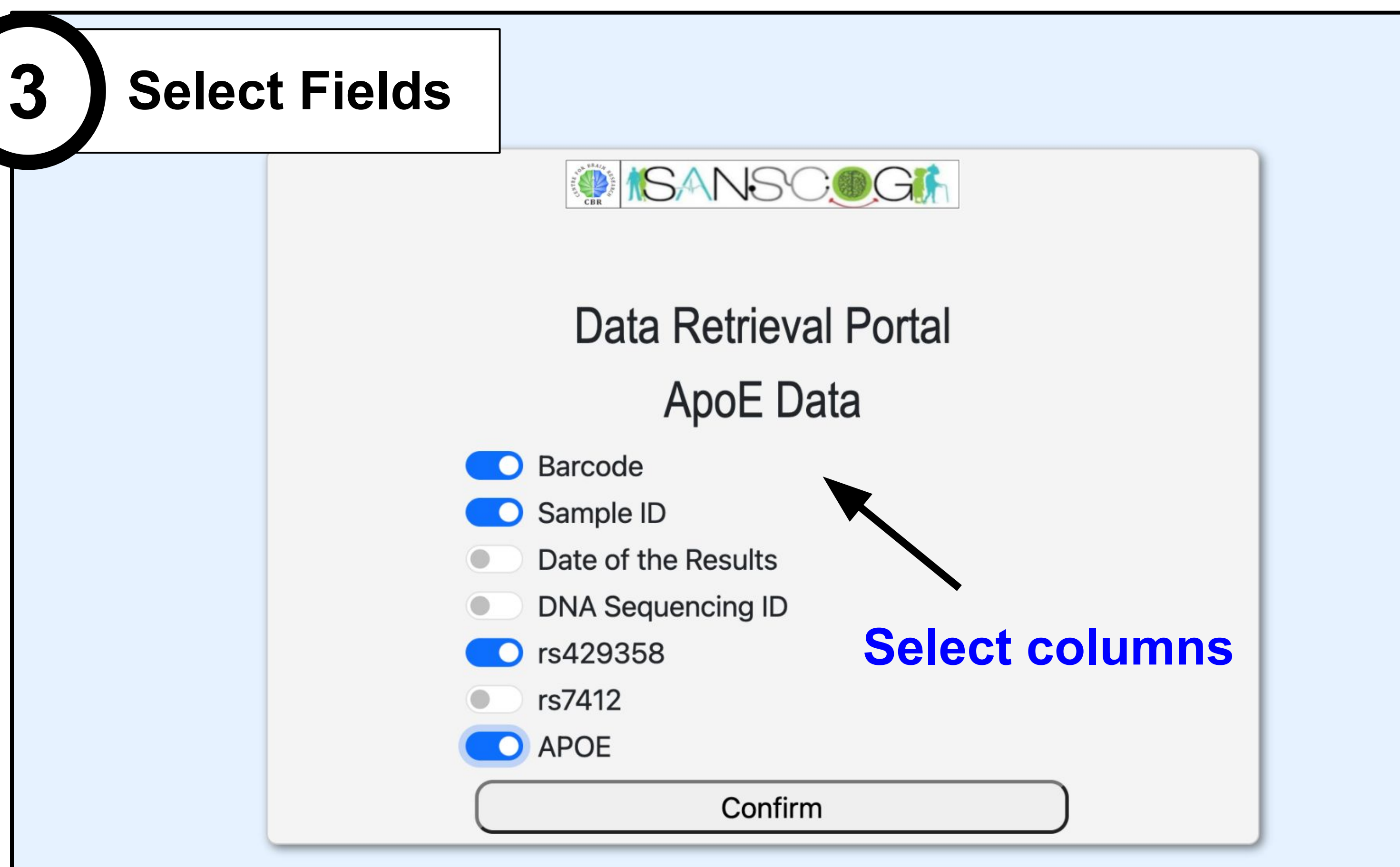
The Login Page shows the 'Data Retrieval Portal' header. Below it are input fields for 'Username' and 'Password', followed by a red 'Sign In' button. An arrow points to the 'Sign In' button with the text 'Enter credentials'.

2 Home Page



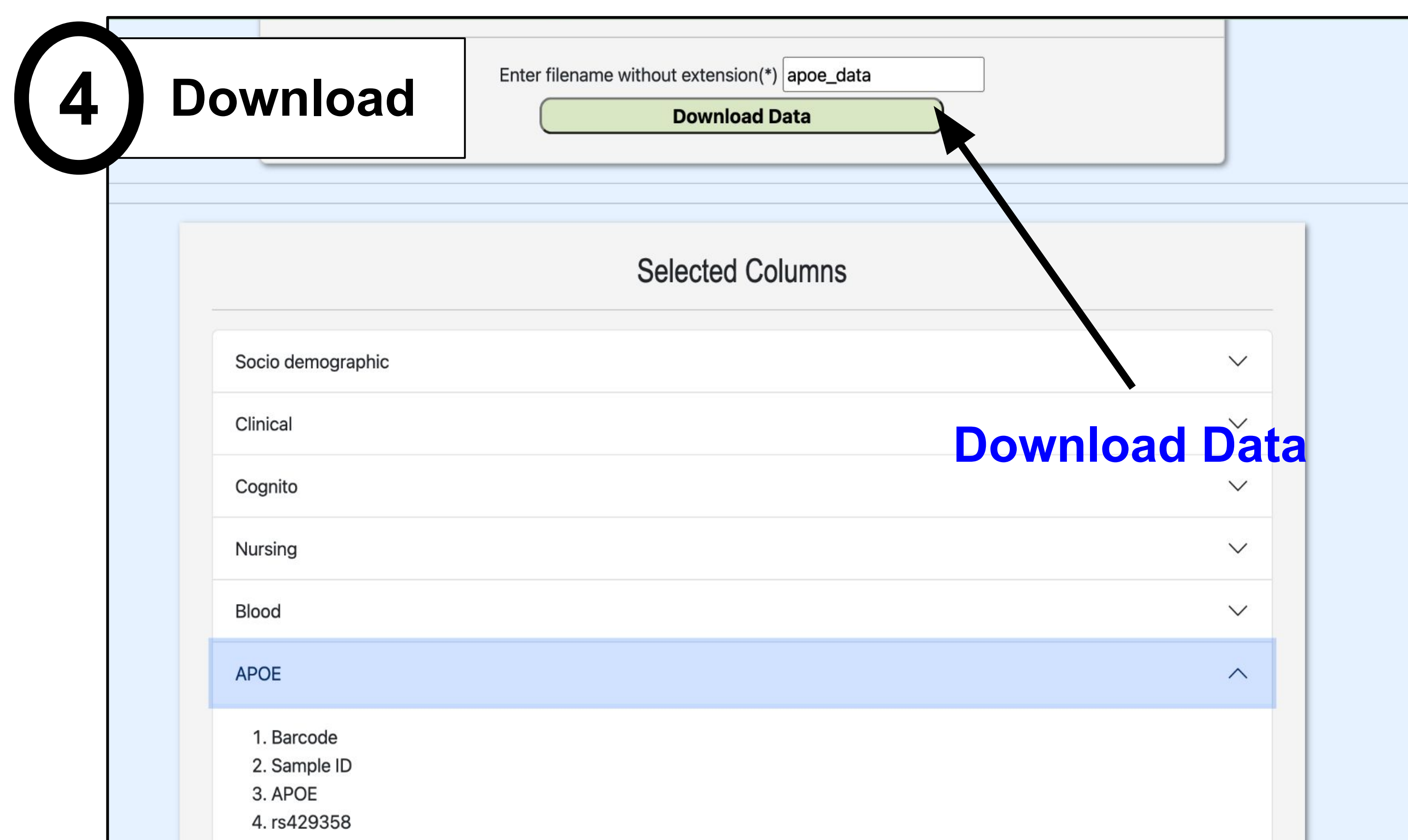
The Home Page shows the 'Data Retrieval Portal' header. Below it is a list of data categories: 'Socio Demographic', 'Clinical', 'Cognito', 'Nursing', 'Blood', 'APOE', and 'MRI'. An arrow points to this list with the text 'Select Data Category'.

3 Select Fields



The Select Fields page shows the 'Data Retrieval Portal' header. Below it is the 'ApoE Data' section with a list of fields: 'Barcode', 'Sample ID', 'Date of the Results', 'DNA Sequencing ID', 'rs429358', 'rs7412', and 'APOE'. Each field has a toggle switch. An arrow points to the 'APOE' toggle with the text 'Select columns'. At the bottom is a 'Confirm' button.

4 Download



The Download page shows a 'Download Data' button at the top. Below it is a 'Selected Columns' section with a list of columns: 'Socio demographic', 'Clinical', 'Cognito', 'Nursing', 'Blood', and 'APOE'. The 'APOE' column is selected. An arrow points to the 'Download Data' button with the text 'Download Data'.

USER INTERFACE

- (1) The landing page of the application is a login page(which ensures data security and only approved researchers to use the facility).
- (2) Once logged in, the home page opens which contains the different categories of data to select from. It also consists of a dropped down menu which consists of the currently selected columns for download (empty by default).
- (3) The user must click on the required category, which further opens a new page consisting of the column subcategories and names which can be selected using a slider button as required.
- (4) Once the user confirms the selected columns, they can be downloaded in .xlsx format with the click of a single button after entering the filename.

FUTURE DIRECTIONS

- (1) We aim to further expand this to build an unified application encompassing data collection as well.
- (2) The data corpus must be hosted in an SQL based database for faster retrieval, maintenance and scalability, and synchronised function of Srinivasapura on-site data collection centres and CBR Researchers.