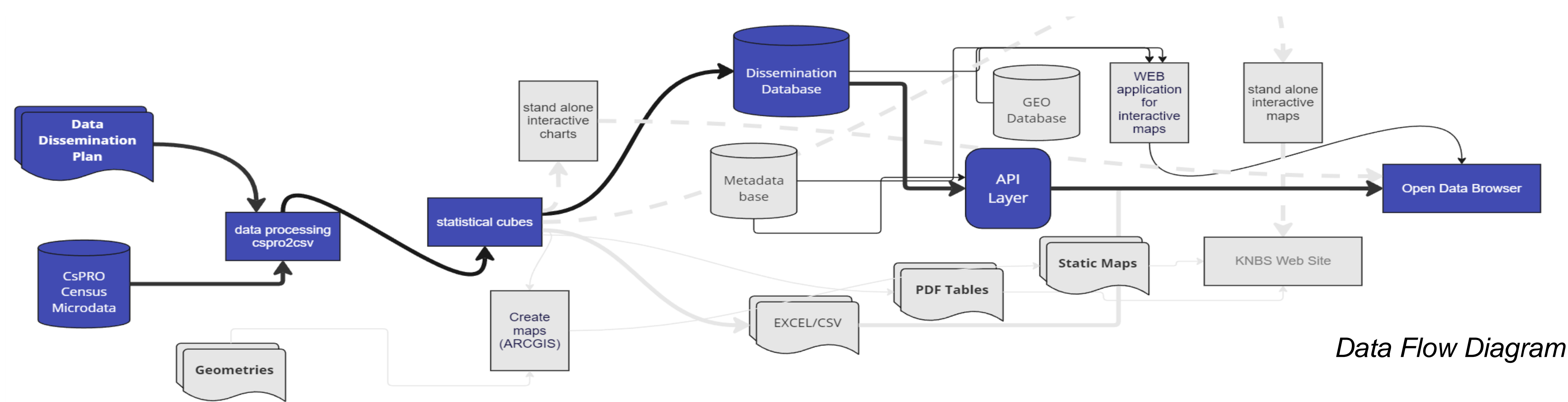


AN OPEN-SOURCE DATA PIPELINE FOR THE DISSEMINATION OF STATISTICAL DATA

MAURO BRUNO - Istat mauro.bruno@istat.it | PAOLO GIACOMI - Istat paolo.giacomi@istat.it | MILENA GRASSIA - Istat grassia@istat.it | VINCENZO PATRUNO - Istat patruno@istat.it
CLAUDIO QUARESIMA - Istat quaresima@istat.it | PAUL WAWERU NGUGI - KNBS pngugi@knbs.or.ke | MAURICE KAMAU - KNBS mkamau@knbs.or.ke

In today's world, statistical institutes of various countries and international organizations often use their institutional websites as the main channel for disseminating data. They do this by adopting simple solutions, such as publishing data tables for download on a specific section of the website, and more advanced solutions, such as adopting and using dedicated software dissemination systems. These solutions are typically developed using proprietary technologies and approaches often far from the open-source logic that should be adopted, especially in international cooperation projects involving many countries in the Global South.

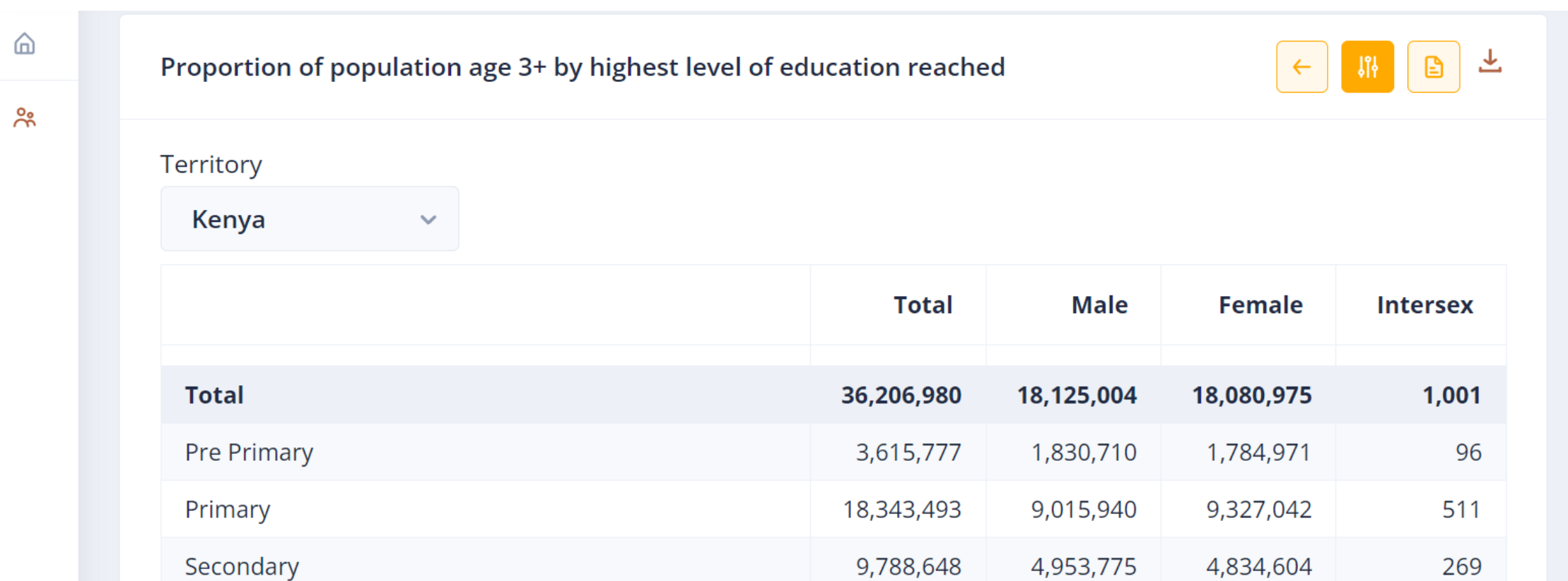
The cooperation project conducted over the past two year at the Kenyan National Bureau of Statistics (KNBS) by Istat with funding from Italian Agency for Development Cooperation (AICS), underscored the adaptability of our approach. It highlighted the need to manage the entire dissemination process in a simple and effective approach without the need to adapt all or part of the technological infrastructure and utilize existing resources and expertise.



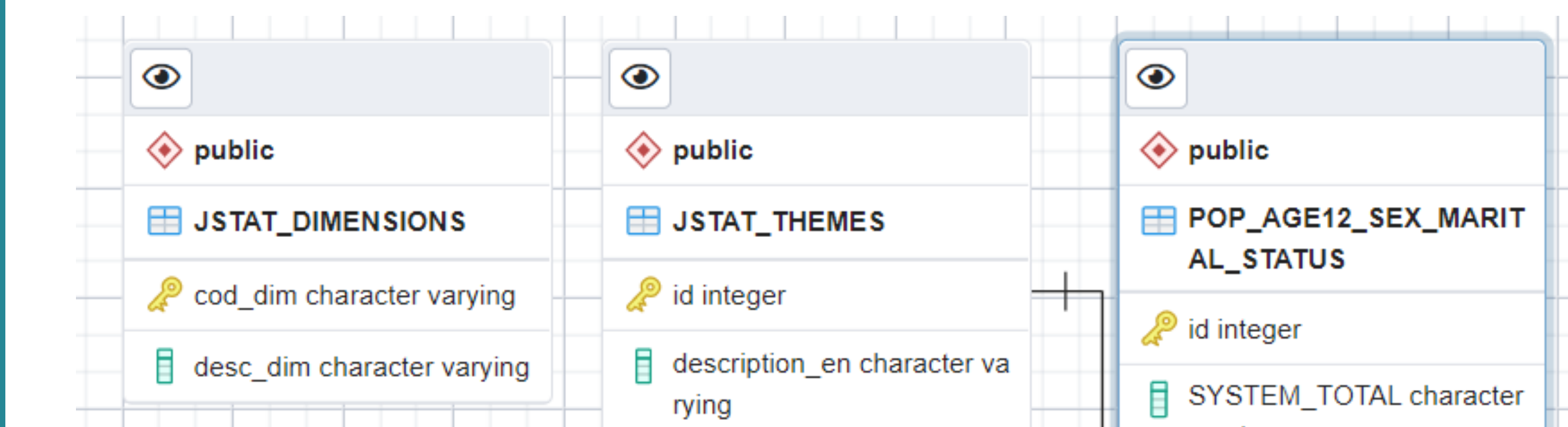
The project therefore proposes a comprehensive and Open approach to the dissemination of statistical data. It is a system that can be easily installed in minutes, scalable, and provides a clear solution to the entire data dissemination process. The project consists of four parts:

1. An accessible and responsive **Open Data Browser** (developed with the Open Source Angular framework) that serves as the graphical interface for navigating statistical cubes

<http://data.knbs.or.ke/>

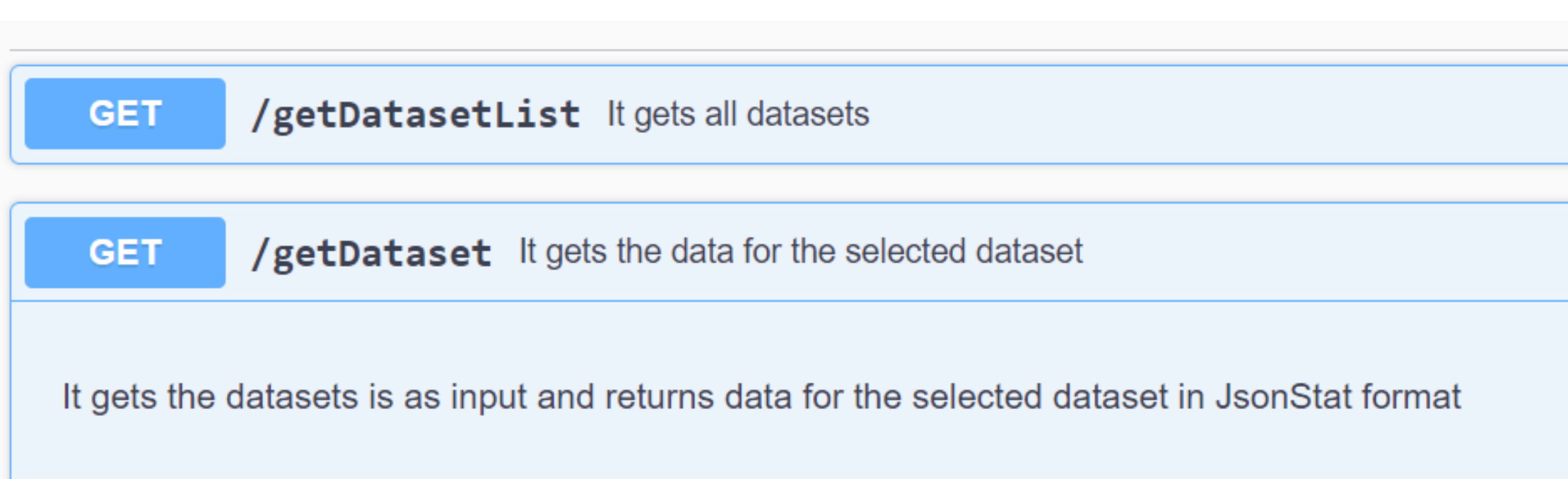


Territory	Total	Male	Female	Intersex
Kenya				
Total	36,206,980	18,125,004	18,080,975	1,001
Pre Primary	3,615,777	1,830,710	1,784,971	96
Primary	18,343,493	9,015,940	9,327,042	511
Secondary	9,788,648	4,953,775	4,834,604	269



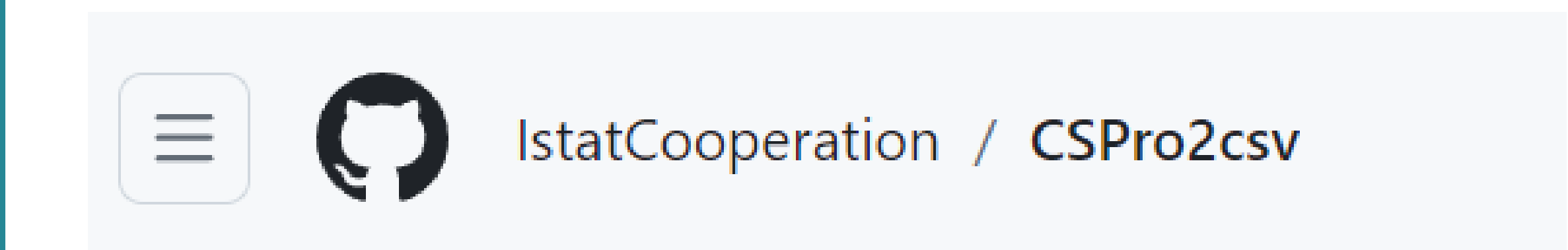
3. A **Python-based API layer** that allows data retrieval from the Dissemination Database. It currently supports the Json-Stat standard format (<https://json-stat.org/>) for data navigation within the Open Data Browser and the CSV format for bulk data downloads.

<http://92.205.128.96:81/knbs/>



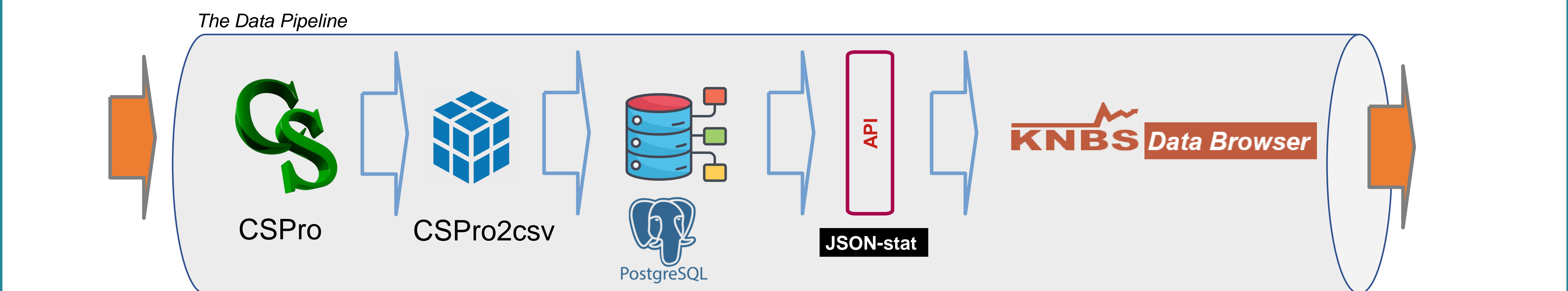
Method	Endpoint	Description
GET	/getDatasetList	It gets all datasets
GET	/getDataset	It gets the data for the selected dataset

It gets the datasets as input and returns data for the selected dataset in JsonStat format



4. A **Python Notebook (CSPro2csv)** that easily transforms tables created using the "tabulate" function specific to Census and Survey Processing System (CSPro - <https://www.census.gov/data/software/cspro.html>) into a multidimensional dataset ready to be inserted into the database.

<https://github.com/IstatCooperation/CSPro2csv>



Therefore, the system seamlessly transitions from data managed within CSPro, a worldwide used data collection tool, to browsable statistical cubes through a web interface.