

Automating Production of Database Metadata

U.S. Census Bureau | Demographic Statistical Methods Division

Michael Jiang, Data Engineering Fellow | *UCLA, Computer Science and Linguistics*

Jenna Braun, Data Engineering Fellow | *UC Berkeley, Data Science*

Keywords:

Source Management, Databases, Automation via Python, PostgreSQL,

Summary:

We developed an automated source management system for the U.S. Census Bureau's Demographic Frame, aiming to centralize and streamline the management of data sources. Using **Python, PostgreSQL, and pgAdmin4**, we tracked source statuses and minimized dependencies between source management and data loading to ensure future adaptability. This system enhances the reliability and efficiency of data integration, reducing errors and supporting scalable, long-term improvements in demographic data processing.

coding it forward >

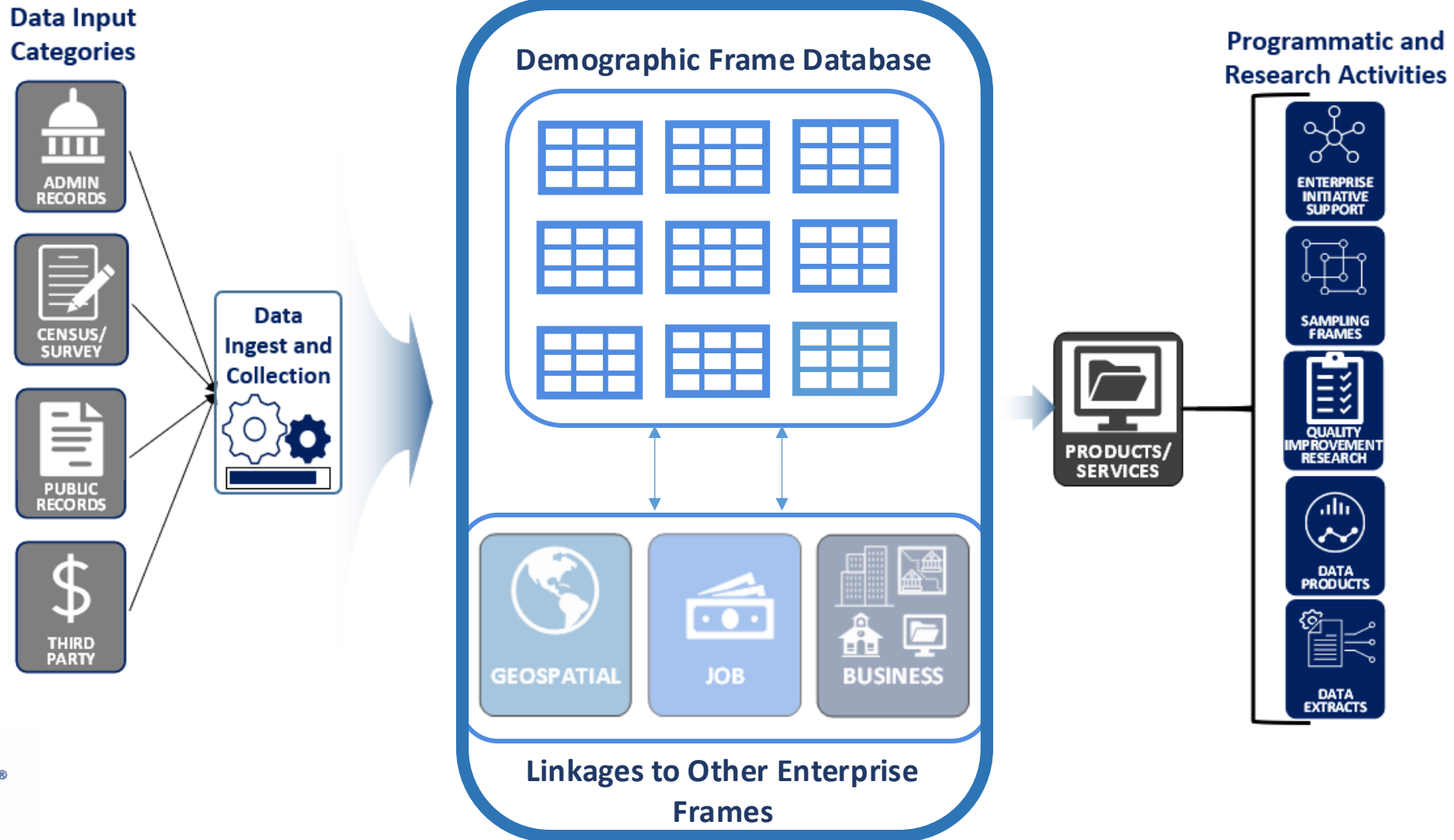
Automating Production of Database Metadata

Jenna Braun and Michael Jiang

Supervised by Jennifer Ortman & Andy Chen

Demographic Statistical Methods Division – Demographic FRAMES Program

The Demographic Frame



Project Overview

- **Issues:**

- The Demographic Frame uses dozens of sources across hundreds of different files
- Currently, there's no automated, centralized source management system

- **Solutions:**

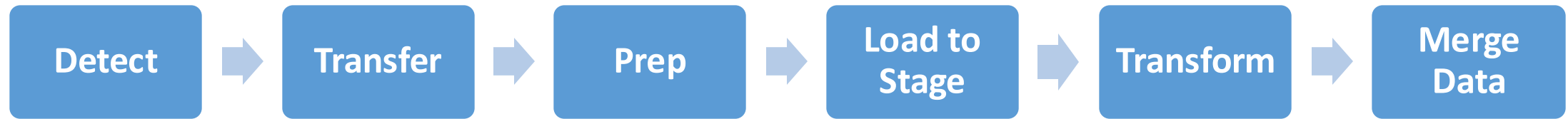
- Develop a source management system that tracks sources and their statuses using a modular approach to ensure stability and future adaptability.
- Minimize interaction and dependency between the source management and data load process, to ensure adaptability.

A standardized computing environment is key to continued development of the Demographic Frame.

Project Contributions

- **Functional Demo Frame Prototype**

- Code to populate source data into source management tables



- **Tested and verified table entries**

- Duplicates were previously able to show up in the tables
 - Added uniqueness constraint on the tables

Project Final Deliverables

- **Porting Source Tables to Source Management Tables**

- Map relevant columns in the old source tables to the new source management tables.
- Identify missing information during the porting process.
- Develop a data transfer program that transfers column information and fills in any gaps.

- **Data load functions**

- Automatic, independent functions that assist with the data load process
- These can interact with the data in
 - S3 (File storage system on Amazon's cloud)
 - EC2 (Cloud Compute aka virtual computers)
 - DB (Database)
 - Or as a whole

Next Steps

- ➡ Creating audit log tables to track changes
- ➡ Automate other components of the data load process
- ➡ Test all functionality on a test server before deploying to production

Thanks!



Email:

dsmd.demographic.frame.team.list@census.gov

michael.c.jiang@census.gov

jenna.m.braun@census.gov



General Information:

[Census.gov/Frames](https://www.census.gov/frames)