NIST Big Data Working Group

May 2018

Big Data Technology Use Cases in 2020 Census

Quyen Nguyen
Enterprise Solution Architect
US Census Bureau
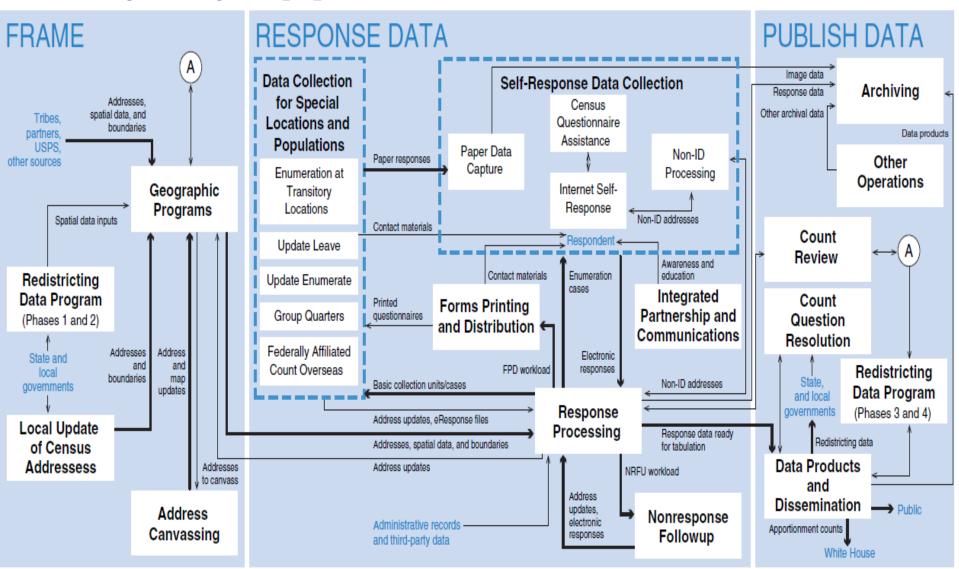
United States™
Census
Bureau

Agenda

- Overview
- Adaptive modeling
- Disclosure Avoidance
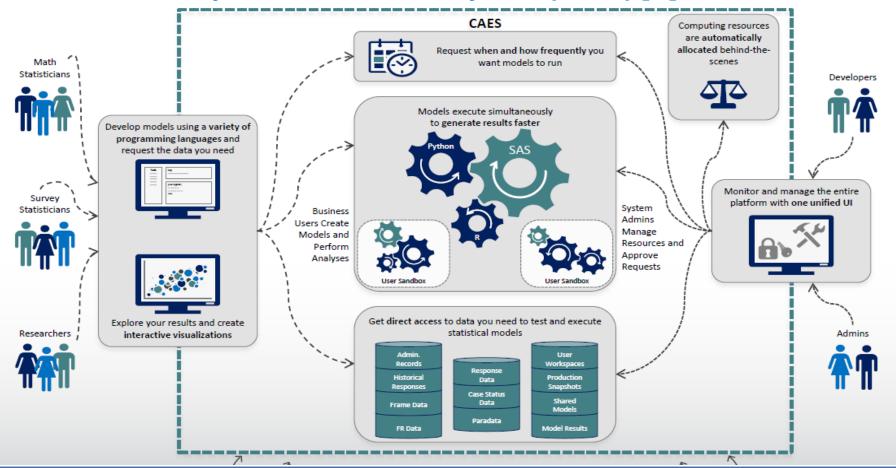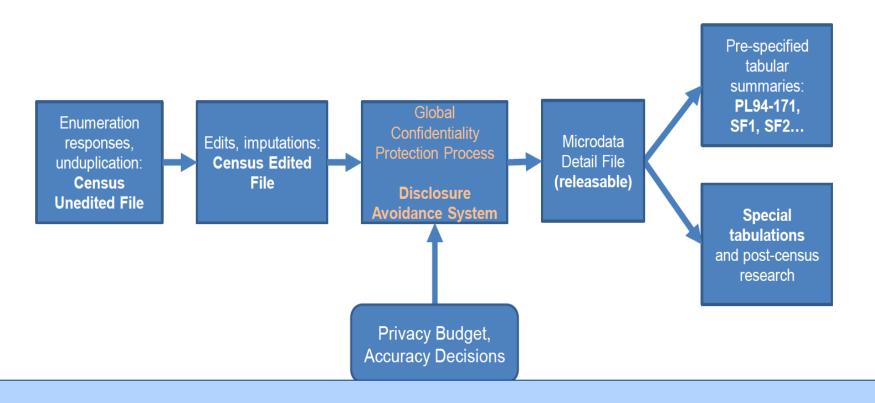- Census Data Lake

Survey Lifecycle [1]

Concurrent Analysis and Estimation System (CAES) [2]

- Analyze third-party data from various sources
- Generate rules for NRFU (Non-Response Follow-Up) cases
  - Remove vacant and non-responding occupied housing units
  - Optimize contact: best time, maximum number of contact attempts

Disclosure Avoidance System [3]

- Critical component of the response processing phase
- Goal: Privacy guarantee does not depend on external data
- Use parallel processing to face the challenge of processing entire country at once
- No trace back to confidential data
- Top down algorithm from nation level to lower geographic levels: state, county, etc.

Census Data Lake

Data Ingest Services

- Ingest data from various data sources
- Ingest Quality Attributes to support Volume, Velocity and Variety attributes

Data Management Services

- Data Transformation
- Data Maintenance, Retention
- Multi-tenancy support
- Audit, Logging, Backup & Recovery

Data Access Services

- Access to data for systems across the Bureau
- Adherence to enterprise standards for dissemination and access

Enterprise Survey Data and Paradata Repository

- Scalable platform for un/ semi -structured data
- Eliminate data regret with ELT
- Enterprise paradata and response data archive

- NBDRA: Big Data Framework Provider and Application Provider [4]

- Support analytics applications

  - Reporting

  - Data review and anomaly detection

  - Data visualization

References

[1] 2020 Census Operational Plan v3.0, September 2017. Online: https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan3.pdf

[2] Michael Thieme and Ben Reist. "Center for Adaptive Design Update". 2017 CSAC Spring Meeting, Mar 2017.

[3] Simson Garfinkel. "Modernizing Disclosure Avoidance: Report on the 2020 Disclosure Avoidance Subsystem as Implemented for the 2018 End-to-End Test". 2017 CSAC Fall Meeting, Sep 2017.

United States™
Census
Bureau