
# US Census Bureau Geographic Support System Initiative Partnership Program

United States™ Census Bureau

## OVERVIEW

- Centrally managed, authoritative database (MAF/TIGER)
- Centrally managed Partnership Program SDE geodatabase
- Supports all censuses and surveys

## CHALLENGES / OPPORTUNITIES

- Need for centralized repository for data sets, security for mission critical data
- Need to support access by various COTS
- Increased level of automation for update processes
- Maintaining persistent topology

## SOLUTIONS

- Oracle Database 11g Enterprise Edition 11.2.0.3
  - Spatial Option with Spatial Topology Data Model
  - Partitioning
- Oracle Web Logic 10.3.6
  - MapViewer 11.1.6, including Oracle Maps
- Oracle Business Intelligence Enterprise Edition
- Oracle Internet Directory

## SOLUTIONS – Cont'd

- Oracle APEX (Application Express)
- Oracle Exadata

## RESULTS

- Data is secured with FISMA compliant mechanisms (using Oracle OID)
- Single Sign On implemented (using Oracle OID)
- Successful processing of 450+ partner data sets, by 50+ business analysts, with thousands more to come
- Consolidation of partner road-centerline and address data in 27 GB sized (and growing) central repository
- Successful integration with FME (for ETL), ArcGIS SDE (for browsing and analysis), 1Spatial Radius Studio (for conflation automation tasks), and Acquis Data Editor (for editing)
- Used Oracle Spatial Topology Data Model to successfully maintain persistent topology
- Exadata has enabled faster processing – timely completion of previously time-intense tasks
- Oracle Exadata compression helps manage large datasets with ease (MAF/TIGER Data sets total approx. 17 TB)

May 2014

# Oracle Spatial Summit

ORACLE


May 21, 2014
Walter E. Washington Convention Center
Washington, DC  USA


# Jay E. Spurlin

Special Assistant to the Assistant Division Chief for Spatial Data Systems and Database Management


# GSSI Partnership Program


# Program Agenda

- Overview
- MAF/TIGER Database
- Geographic Partnership Program
- Oracle Exadata

United States™ Census Bureau


# Overview

## US Census Bureau - Geography Division

- The Geography Division handles geographic and cartographic activities in support of the Census Bureau's statistical programs.
- We continuously update features, boundaries, geographic entities, and address information
  - TIGER (Topologically Integrated Geographic Encoding and Reference)
  - MAF (Master Address File).
- We research geographic concepts, methods, and standards needed to facilitate data collection and dissemination.

United States™ Census Bureau


# Overview

## Geographic Support System Initiative

- In support of the 2020 Decennial Census, the Census Bureau is evaluating what areas should be targeted for a traditional, on-the-ground address canvassing operation and areas in which it is not necessary.
- How will we decide which areas should be considered for targeting?
  - GEO has evaluated the MAF/TIGER database and assigned quality indicators to each of the census tracts
  - The MAF/TIGER is being updated through the Geographic Partnership Program
  - A Targeted Address Canvassing strategy has been developed that contains an inventory of criteria for evaluation

United States™ Census Bureau


# Overview

## Geographic Support System Initiative

- The Geographic Partnership Program is now underway.
  - GEO is receiving both address and spatial data from invited partners
  - The data are being evaluated and integrated with the MAF/TIGER database.
  - The next step is to determine what level of feedback we can give to the partners about their data.
- The combination of the evaluation of the current MAF/TIGER database, the partner data, and predictive modeling will contribute to the recommendation on which areas of the country should be considered for targeting.

United States™ Census Bureau


# MAF/TIGER Database

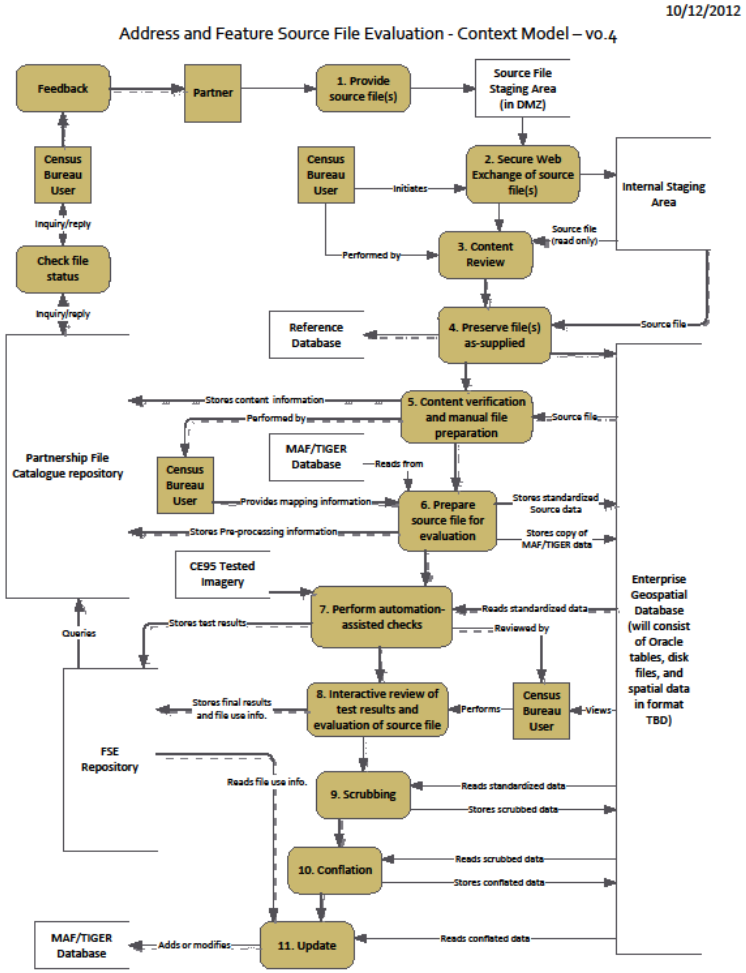
## An overview

- Implemented on Oracle Spatial Topology Data Model
  - Custom topological primitive features
  - Over 50 feature “layers” implemented in dozens of feature tables
  - Hierarchical features (features built on other features)
- Many software systems in MAF/TIGER system – all using Oracle
  - GWCS (GSSI Workflow Control System)
  - School District Review Program (SDRP) Crowd Source application
  - PCPTR (Problem Capture, Prioritization, Tracking, and Reporting)
  - QIs (Quality Indicators)

United States™ Census Bureau


# Geographic Partnership Program

# Context Model




# Geographic Partnership Program

## Acquisition of data

- Data are provided at the state, county, and local level.
  - A partner provides a set of source files
  - The source files are moved inside the Census firewall via a secure web-exchange module
  - The content inventory of the files undergoes initial verification
    - The data and metadata provided are verified against published standards
  - The files are preserved, as supplied, for later reference

United States™ Census Bureau


# Geographic Partnership Program

## Evaluation of data

- Data are evaluated
  - More detailed content assessment is done, including verification the files meet minimum guidelines for content and metadata
  - The files are prepared for automated processing, including re-projection and mapping to a standardized schema
  - A series of (mostly) automated checks is run, which provides metrics about the data in the files

United States™ Census Bureau


# Geographic Partnership Program

## Evaluation and integration of data

- Update method is chosen
  - An interactive review is conducted, in which the files and their associated metrics are reviewed and a decision is made how to capture any new data
  - Some feature files are processed by an interactive editor, using the partner file, imagery, and other sources
  - Some feature files are processed through a semi-automated conflation process
  - Address files are processed through an automated update system

United States™ Census Bureau


# Geographic Partnership Program

## Integration of data - Conflation

- Data are integrated into the MAF/TIGER database
  - Scrubbing: Any data that are not useful for updating the MAF/TIGER database get removed from the files
  - Matching: Features or addresses are matched to the MAF/TIGER database, using an automated process. An interactive review identifies any issues and allows adjustments to be made
  - Update: All approved changes are applied to the MAF/TIGER database, using an automated process.
  - Metadata is captured and stored in the database, with quality metrics about each feature

United States™ Census Bureau


# Geographic Partnership Program

## Integration of technology architecture

- COTS software integration
  - ArcGIS
    - Partner data is spatially enabled using ArcSDE
    - ArcMap is used to view and analyze data
  - FME (Feature Manipulation Engine)
    - Extraction of reference layers, ETL (to load partner data into an SDE geodatabase)
  - 1Spatial Radius Studio
    - Used for automation of conflation processes

United States™ Census Bureau


# Geographic Partnership Program

## Integration of technology architecture – cont'd

- COTS software integration
  - ADE (Acquis Data Editor)
    - Built on MapViewer
    - Utilized by GATRES (Geographic Acquis-Based Topological Real-time Editing System) for interactive editing
  - Redwood Cronacle
    - Used for process control, scheduling, and monitoring

United States™ Census Bureau


# Oracle Exadata

## Delayed Calculations – TABBLOCK table

| Field | Type | Description |
|---|---|---|
| ACT | VARCHAR2 (2) | |
| AREALAND | NUMBER (14) | Adds up areas of land topological faces (which are calculated using SDO_GEOM.SDO_AREA) |
| AREALANDHIST | NUMBER (14) | Saves AREALAND value, if AREALANDHIST is null |
| AREAWATER | NUMBER (14) | Adds up areas of water topological faces |
| AREAWATERHIST | NUMBER (14) | Saves AREAWATER value, if AREAWATERHIST is null |
| BLOCKCE | VARCHAR2 (4) | |
| BOUNDARYEDGES | SDO_TOPO_GEOMETRY | Updates topology for use cases that Topology Manager does not handle automatically |
| CENTROID | SDO_GEOMETRY | Uses SDO_GEOM.SDO_CENTROID |
| COUNTYFP | VARCHAR2 (3) | |
| CYCLES | NUMBER (3) | Counts external polygon loops |
| CYCLESMAX | NUMBER (3) | |
| FUNCSTAT | VARCHAR2 (1) | |
| GENSDOGEOMETRY | SDO_GEOMETRY | |
| HOUSING | NUMBER (9) | |

United States™ Census Bureau


# Oracle Exadata

## Delayed Calculations – TABBLOCK table – cont'd

| Field | Type | Description |
|---|---|---|
| INTPOINT | SDO_GEOMETRY | Uses centroid, if inside polygon; if not, chooses first quadrant centroid that falls inside polygon (recursively) |
| LSAD | VARCHAR2 (2) | |
| LWBLKTYP | VARCHAR2 (1) | Set aggregate value, based on values of composite topological faces |
| MTFCC | VARCHAR2 (5) | |
| NAME | VARCHAR2 (100) | |
| OID | NUMBER | |
| PERIM | NUMBER (9) | Adds lengths of topological edges in BOUNDARYEDGES |
| POPDEC | NUMBER (9) | |
| SDOGEOMETRY | SDO_GEOMETRY | Uses GET_GEOMETRY; calls FME to aggregate some topologically hierarchical features |
| STATEFP | VARCHAR2 (2) | |
| SUFFIX1CE | VARCHAR2 (1) | |
| SUFFIX2CE | VARCHAR2 (1) | |
| TOPOGEOM | SDO_TOPO_GEOMETRY | |
| TRACTCE | VARCHAR2 (6) | |

United States™ Census Bureau


# Oracle Exadata

## Performance improvement example

- Delayed Calculations – Day to day
  - Before Exadata migration, nightly calculations (for daily changes) rarely completed – and were automatically terminated at 6 a.m., causing a consistent backlog
  - After Exadata migration, nightly calculations almost always complete – sometimes in less than an hour

United States™ Census Bureau


# Oracle Exadata

## Performance improvement example

- Delayed Calculations – Benchmark Update
  - Includes calculation of SDO_GEOMETRY values from SDO_TOPO_GEOM types, as well as other calculations (like areas and internal points)
  - All delayed calculations are calculated for each "benchmark" (snapshot) of the MAF/TIGER database
  - Processes for benchmarks in 2010 and 2011 took between one and two months
  - Process after migration to Exadata took five days

United States™ Census Bureau

# Q&A