U.S. Census Bureau

Spatial Data Storage and Topology in the Redesigned MAF/TIGER System

Jennifer DeVore and Danielle Lessard

March 10, 2005

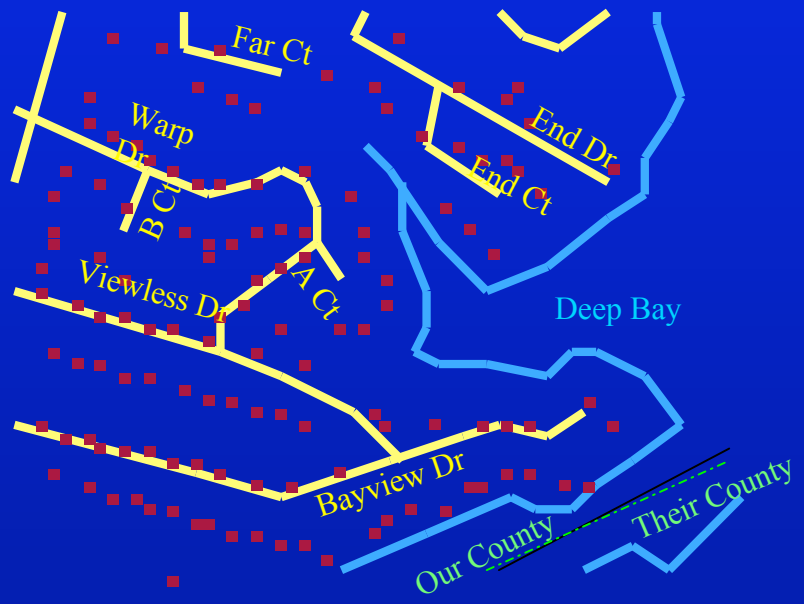
What is TIGER?

Topologically Integrated Geocoding and Referencing System

A street center-line "digital map" (geographic data base) of the entire United States, Puerto Rico, and the associated Island Areas

TIGER Content

- **Street center-lines and their names**
- **Lakes, streams, and their names**
- **Railroads**
- **Geographic entity boundaries, names, and codes (for governmental units, census tracts, census blocks, etc.)**
- **Housing unit locations (selected areas)**
- **Key geographic locations (for airports, schools, etc.)**
- **ZIP Codes and address ranges (for streets with city-style addresses)**



What is the MAF?
Master Address File

An accurate and up to date inventory of all known living quarters in the United States and Puerto Rico.

The content of the MAF is:

- Mailing address, if one exists
- Descriptive address, when no city-style address exists
- Census geographic location
- Source and history data
- Non-spatial data
- Supports data collection efforts, address canvassing operations and questionnaire deliveries to each residence.

Title 13 requires that all addresses/locations be kept confidential

MAF/TIGER:
Mission Critical Corporate Resource

- System provides storage, processing, products and services that support agency's statistical programs.
  - Geocoding
  - Maps
  - Residential Address Lists
  - Geographic Reference Files
- Continually updated with new address and geographic information.
- Wide public use of geographic information.

Issues with Legacy System

- Homegrown database system doesn't integrate well with COTS and Web Technology
- Cumbersome to change
- Difficult for new developers to learn
- Does not allow multi-user access
- Not accessible via a Standard Query Language

Storage of Geometry: Layers or Integrated Topology?

- Roads, rivers, and other linear geography in the TIGER database often also serve as boundaries for geographic areas, such as places or counties.

- The boundaries of over 75 different types of tabulation and collection geographic areas are managed and maintained by the Census Bureau.

- Multiple vintages must be maintained simultaneously. These areas often share portions of their boundaries with each other and/or with linear features.

Topology: Calculate on-the-fly or Store Persistently?

The majority of processing for the MAF/TIGER system is spent on large batch processes that run on the whole nation and utilize topology to improve performance. On-the-fly topology calculation could be a performance issue for these national batch runs.

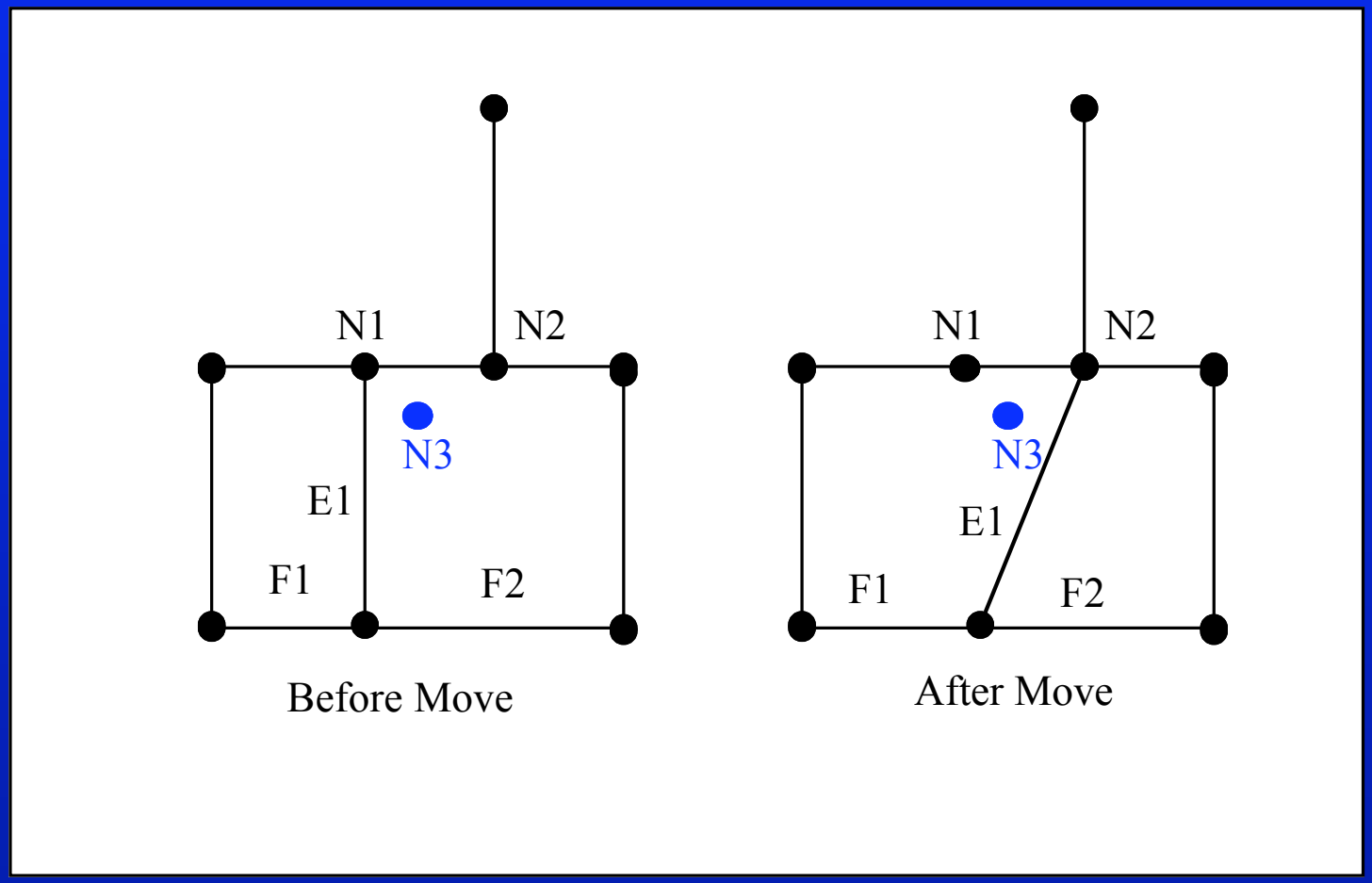
Managing Topology

The redesigned MAF/TIGER system will use the persistent topology data structure that is part of Oracle Spatial, starting with the release of Oracle 10g. This system, called the Oracle Spatial Topology Data Model, provides persistent topology to support batch or interactive applications.

Isolated Nodes: Separate Layer or Integrated with Topology?

In the Legacy TIGER system, isolated nodes can be linked to a face but they are not actually part of the topology. The Oracle Spatial Topology Data Model allows isolated nodes to be integrated into the topology layer.

Example: reshaping a geometric feature



Features, Attributes, and Topology

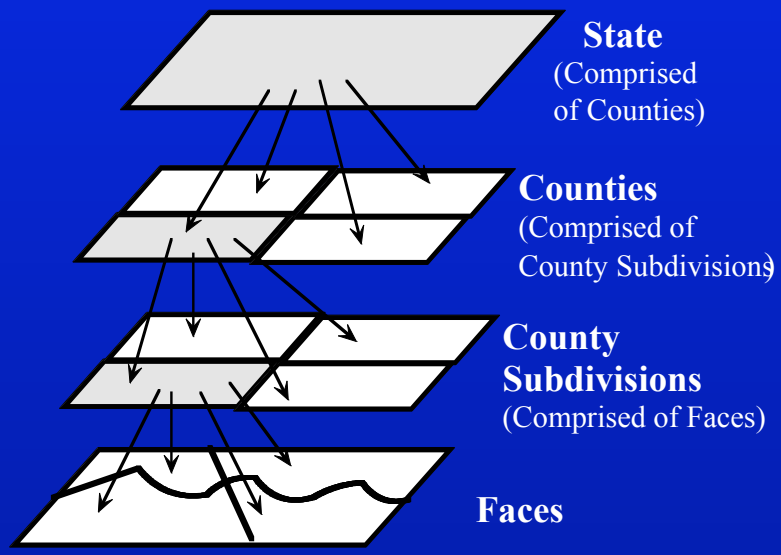
- The Oracle Spatial Topology Data Model manages the creation of edges and faces, therefore these topological primitives are not readily modified by the user, in the event that it is necessary to store attributes at the primitive level.
- The redesigned MAF/TIGER makes greater use of higher level features and assigns attributes on the features where possible.
- Certain attributes vary over the extent of a feature and are more appropriately stored at the topological primitive level.

Hierarchical Features

•Areal features are often hierarchical in nature. For example, states are comprised of counties, and counties are comprised of county subdivisions.

•The Census Bureau will be using the Oracle Spatial Topology Data Model to define geographic areas up through the County level in terms of other geographic areas, rather than defining them directly in terms of faces.

Hierarchical Diagram



Advantages of the New System

- Seamless national database improves ability to respond to customer needs for product delivery mechanisms or partitioning.
- Integrates spatial and non-spatial data.
- Facilitates data exchange with local partners.
- Improves data accessibility and currency, using standard GIS tools.
- Greater flexibility and efficiency in responding to new or changed requirements.
- Use of commercial database and current technology facilitates recruitment/training of users.
- More efficient data exchange .

Questions?