PROCESSING TIGER FILES AND STF FILES

- **THE CREATION OF AN ARCHIVE OF CENSUS RELATED PRODUCTS:**

    - **TIGER BOUNDARY FILES**
    - **STF3A STANDARD EXTRACT FILES**
    - **ENHANCED MIGRATION FILES**
    - **ACCESSIBLE VIA ANONYMOUS FTP**
    - **ACCESSIBLE VIA WWW/Mosaic**

- **Hendrik K. Meij, CIESIN**
- **John Blodgett, UIC**

- **credits: CIESIN operations**
    - **Dennis Schorn & Jeff Blaine**

- **a new, and improved, ...**

SHEREKHAN
II

Processing TIGER92 Files

- **programs**
- **products**
- **examples**

Processing STF3A Files

- **programs**
- **products**
- **examples**

Archive Structure

- **directory structure**
- **documentation**

Retrieval Mechanisms

- **anonymous ftp**
- **Mosaic**

Usage

- **building of a "prototype"**

Processing TIGER92 Files

- **programs**

  input: state level TIGER 92 files, ZIPPED

  access: while looping through zip archive file by county,

  data files are uncompressed, sorted, and transferred to “pipes”

  loops: SAS(r) program files are built on-the-fly and submitted, containing statements to “chain” (connect-the-dots) at specified geographic levels, build the ***POLID*** unique geocode, load region names if appropriate, and build boundary datasets

  post-process: Transform boundary datasets into boundary ascii files (BNA format; fairly simple layout used by ATLAS*GIS(r), generate report file, ZIP all files)

- **products**

  BNA files by state for

  - county, mcd (fips), place (fips), and 1980 tract and mcd
  - pumas, 1% and 5% (?)

  BNA files by MSA code for

  - blockgroup, tract
  - MSA code “**9999**” contain all non-metro parts of the state

  BNA files by county for

  - block

  SAS(r) Data Library Transport Files by county

  Comma Separated Value Intersection Datasets by county

Processing TIGER92 Files

- **examples**

BNA Files

```
"211739802.00106A","106A","1211739802.00106A","blk",4
-83.940732, 38.067433 -83.942302, 38.066794
-83.940509, 38.067160 -83.940732, 38.067433

"211739802.001","1","2211739802.001","blkgrp",43
-83.945120, 38.079182 -83.946037, 38.078917
-83.948815, 38.078116 -83.949625, 38.077933 ...

"211739802.00","9802.00","3211739802.00","tract",128
-83.884252, 38.090194 -83.894784, 38.088679
-83.897823, 38.088451 -83.898575, 38.088452 ...

"2117391872","Jeffersonville","42117391872","fmcd",618
-83.903578, 38.013851 -83.907691, 38.012052
-83.905928, 38.009353 -83.904918, 38.008056 ...

"2117312052","Camargo","62117312052","fplace",64
-83.881803, 37.981387 -83.882416, 37.984456
-83.882857, 37.987048 -83.882876, 37.987501 ...
```

CSV Files

```
"NAME1","NAME2","COUNTY","NODE","XLONG","YLAT"
"AARONS RUN","GRASSY LICK","21173",280,
    -84.002622,38.127251
"AARONS RUN","US HWY 460","21173",100,
    -84.036651,38.113545
```

Processing STF3A 1990 Files

- **programs**

  input: state level STF3A 1990 Dbase(r) files from LBL

  access: Hockey's "dbftosas.sas" enhanced SAS(r) macro reads in cdrom files into SAS(r) datasets, separated by ***SLVL*** (h=070,080,090, i=others)

  loops: SAS(r) program files contain statements to build the ***POLID*** unique geocode, load region names if appropriate, post-processes table cells (into %, medians, ratios), and build the state level standard extract files (ix)

  post-process: Transform ix dataset into comma-separated-value files by ***SLVL***, and generate a contents file.

- **products**

  CSV files by state for

  - county, mcd, metro, place (within state and within county)

  CSV files by MSA code for

  - blockgroup, tract, zipcode (5-digit)
  - MSA codes "**9999**" contain all non-metro parts of the state

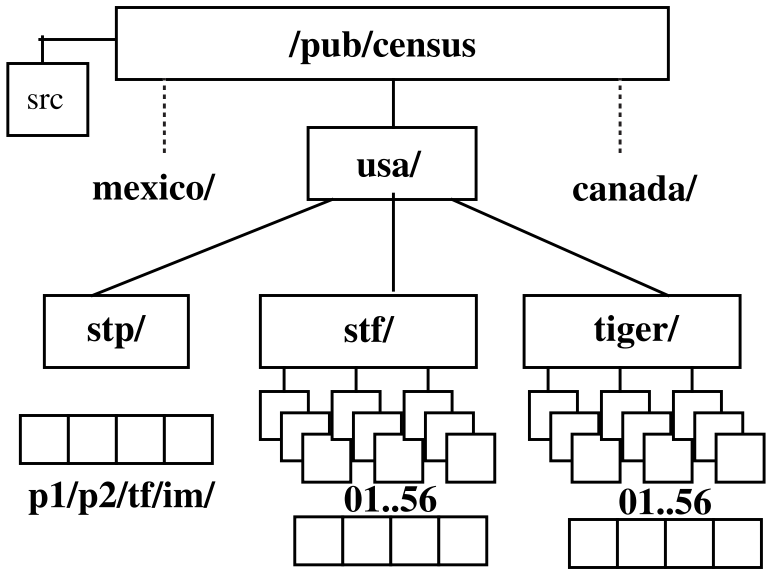
- **examples**

```
"POLID","SLVL","GEOCODE","AREAAME","COUNTY","MCD",
"PLACE","TRACT","BG","XCENTRM","YCENTRM","TOTPOP",...

"211739801.001","150","21173-9801.00-1","BG1","21173",
" "," ","9801.00","1",664.235,-129.331,600,...
```

Archive Structure

- Directory



- /pub/census/src zip & unzip programs
- /pub/census/usa
- /pub/census/usa/stp enhanced migration files

  p1, p2, tf, im directories
- /pub/census/usa/stf standard extracts files
- /pub/census/usa/tiger boundary files

  bna_st, bnablk, csvisd, xptsdl directories

- documentation and contributed programs

Retrieval Mechanisms

- **ftp ftp.ciesin.org**
- **Name (ftp.ciesin.org:hmeij): ftp**
- **Password: <sherekhan@jungle.book.edu>**
- **ftp> cd /pub/census**

250-Most files in this tree need ftp mode set to BINARY
250-(issue the command “binary” at the ftp prompt).
250-
250-Within this directory you will find ...
250-
250-   usa/   files related to U.S. Census Data
250-   src/   files related to Zip/Unzip (version 201)
250-
250-Please email problems to: **ciesin.info@ciesin.org**

- **URL http://www.ciesin.org/**
  follow links to **“Data Access”**
  “Dataset Guides”
  “Us Demography”

Usage

*How to build a local GIS database (applies to ATLAS*GIS/DOS)*

- **GIS database**
  - retrieve TIGER boundary files for desired geographic area

    this may contain entire STATEs, entire MSAs ... or sections of multiple states (ie, the Greater Chicago Area)
  - UNZIP all files (for example in C:\TEMP)
  - transform TIGER BNA files in AGIS databases using

    ATLAS Import/Export. The “names=4” option instructs *ATLAS*GIS* to assign the 4th field as layer name.

    `\agis\ie bg219999.bna /names=4`
  - enter ATLAS*GIS and combine all layers if desired

    `FILE-GEOGRAPHIC-NEW (create destination)`

    `FILE-GEOGRAPHIC-MERGE (select all)`

    for an entire state,including blockgroups, this can take a night. (You may also combine same type of geography BNA layers together [blockgroup/tract] with the DOS copy).

    Following these procedures the prototype database will contain one or more layers of geographic boundaries all identified by the layer name.

  - Within each layer, **_NAME** (created by *ATLAS*GIS*) contains the unique value of ***POLID*** (the first field in the names=4 list.

• Attribute database

- retrieve STF files for desired geographic area matching the TIGER files of previous section
  - this may contain entire STATEs, entire MSAs ... or sections of multiple states (ie, the Greater Chicago Area)
  - the standard extract files are split in 2 files per geographic ***SLVL***, an [**a**] and [**b**] file. The “**9999**” contains the non-metro part of the state.
- UNZIP all files (for example in C:\TEMP)
- enter *ATLAS*GIS*, and make a geographic file active

  `FILE-GEOGRAPHIC-USE`

  this database should contain the layers of the selected standard extract files retrieved (ie, blockgroup, tract)
- Import the unzippped attribute files (*.csv), select

  `FILE-ATTRIBUTE-TOOLS-IMPORT`

  select the format **Comma-Delimited-With-Field-Names**

  enter the destination file to create

  accept the database structure presented (press **F10**)
- Field to match between Geographic and Attribute files

  you will be prompted next to identify the matching fields

Field to Match From........POLID

Expression to Match to...._NAME

Duplicate Matches..........Reject

Non-Matches................Reject

Press **F10** to accept

• Datapoint database

- the Intersections Datasets can be imported into *ATLAS*GIS/* DOS in a similar manner
- retrieve ISD files for desired geographic area matching the TIGER files of previous section (county level files)
- UNZIP all files (for example in C:\TEMP)
- enter *ATLAS*GIS*, and make a geographic file active

  `FILE−GEOGRAPHIC−USE`
- Import the unzippped intersection files (*.csv), select

  `FILE−DATAPOINT−TOOLS−IMPORT`

  select the format **Comma-Delimited-With-Field-Names**

  enter the destination file to create
- the database structure needs to be edited to assign**LAT-LON**, from xlong-ylat. Simply edit the variable names and set the filed type to *numeric* and set: LAT width 10, 6 decimal places; LON width 11, 6 decimal places.

  accept the database structure presented (press **F10**)
- You may copy the datapoint locations (street intersections) to the Geographic file (for example intop a new layer "intersections", with the command

  `EDIT−GEOGRAPHIC−COPY−DATAPOINTS`

`<end>`