Crossing Boundaries
Building Tools to Combine Public-Use (and Misaligned) Data

Matthew Graham
LEHD Program
Center for Economic Studies
U.S. Census Bureau

2018 FCSM Research and Policy Conference

Disclaimer

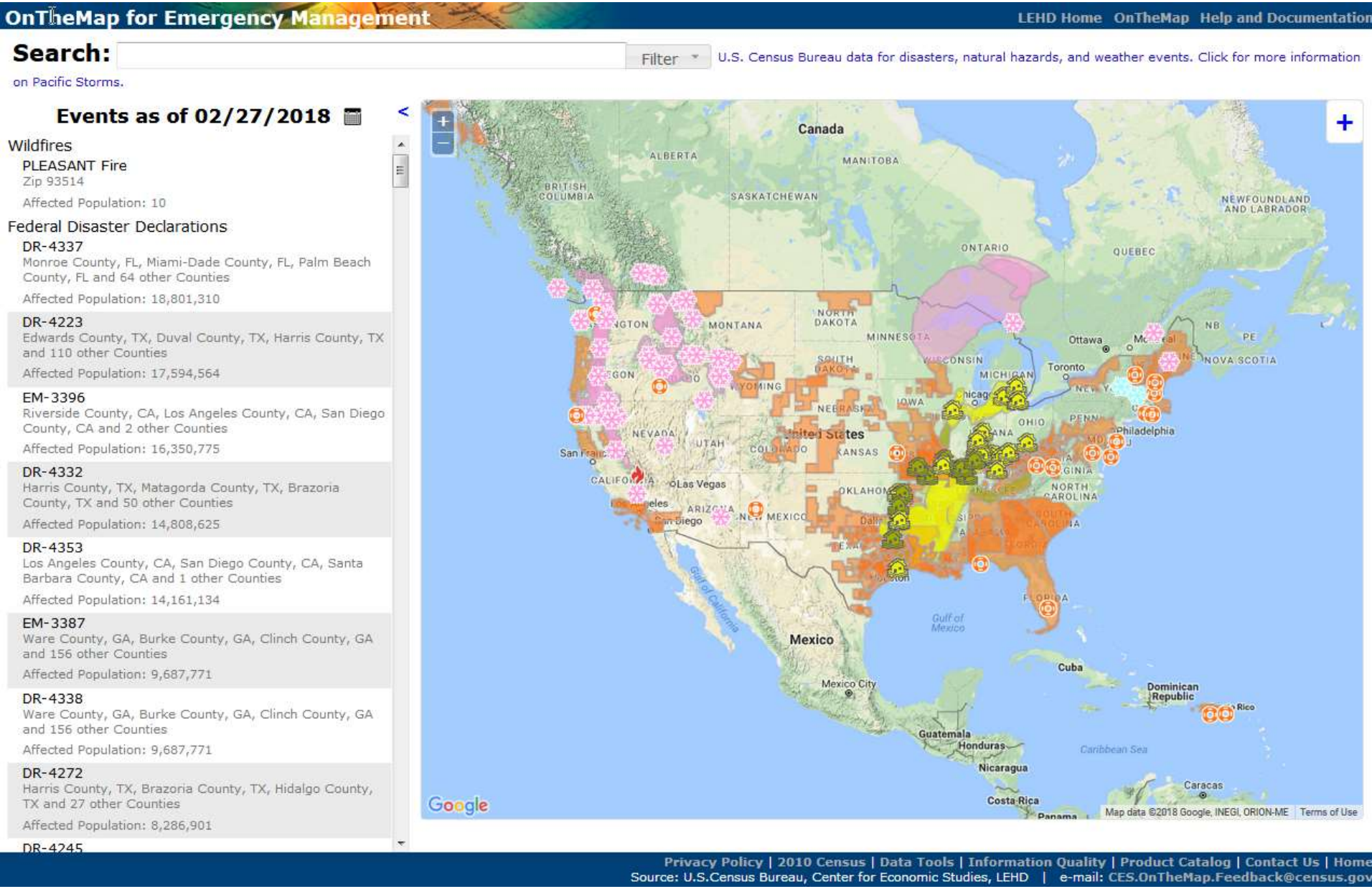
- Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.
- Additionally, these opinions and conclusions are not representative of other data products or programs within the Census Bureau.

Background

- LEHD Program
  - Longitudinal data on firms, workers, and jobs
  - >95% coverage of jobs
- Public-use data products (QWI, LODES, J2J) and dissemination tools (e.g. OnTheMap)
- Record of developing user-focused analysis tools with data partners in mind
  - More

OnTheMap for Emergency Management

- Free public web application that gives emergency preparedness/disaster recovery analysts easy access to economic & demographic summaries for emergency events in near real time.
- Disseminates some statistics from 2010 Decennial, American Community Survey (ACS), and LEHD Origin Destination Employment Statistics (LODES).
- Automatically incorporates real-time weather and hazard event data from various authoritative federal sources into a single searchable archive and map display, for rapid access, visualization, and reporting.
- Building early LEHD tools (OnTheMap) resulted in lots of requests for analysis, particularly during emergency events or disasters.
- But we're not emergency management analysts…
  - So we built OTMEM to automate the mechanical work.



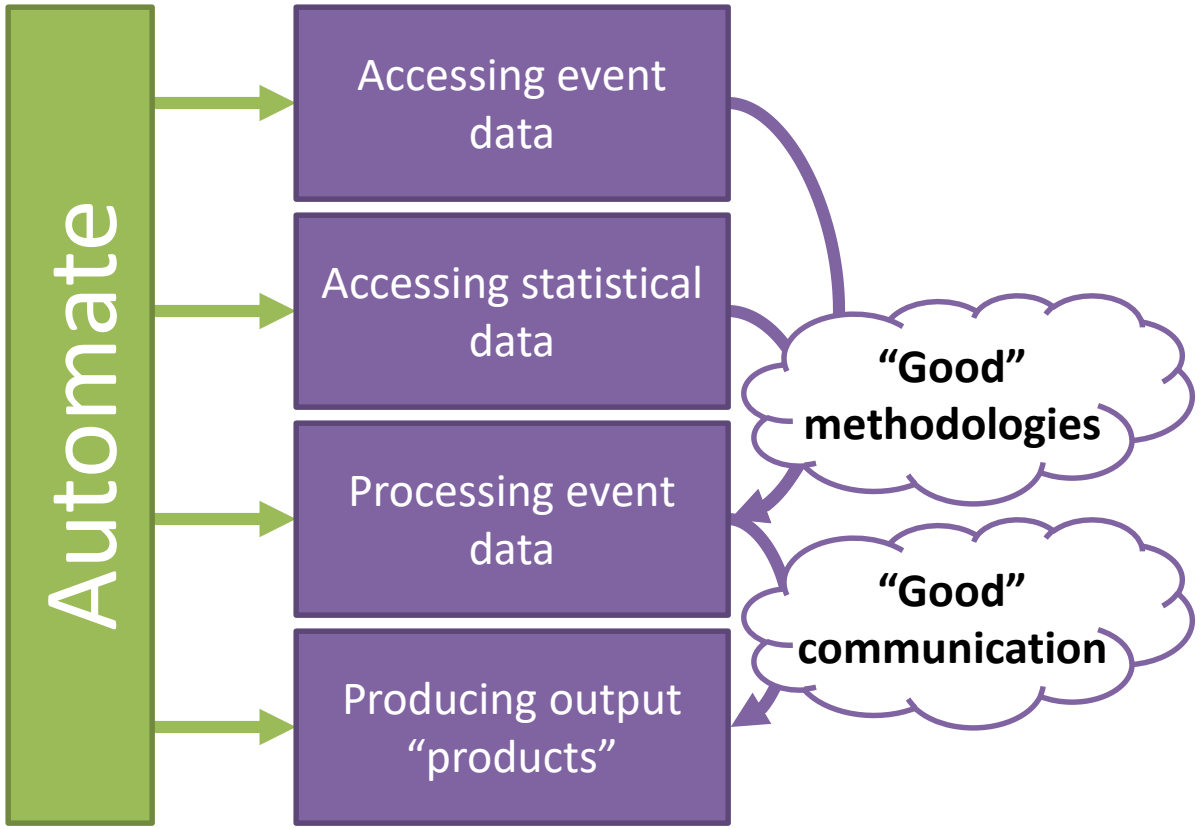
United States Census Bureau | U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

https://onthemap.ces.census.gov/em

Partial Timeline

- Late 1990s – LEHD Program started
- 2005 – 1st release of OnTheMap (data product and application)
- 2005 – Hurricane Katrina
- 2007 – I-35W Mississippi River bridge collapse
- 2007 – Southern California wildfires
- 2008 – Mississippi River flooding in Iowa
- 2010 – Deepwater Horizon oil spill
- 2010 – 1st release of OnTheMap for Emergency Management
- 2014 – Addition of ACS data to OTMEM

Building OTMEM



Data Sources

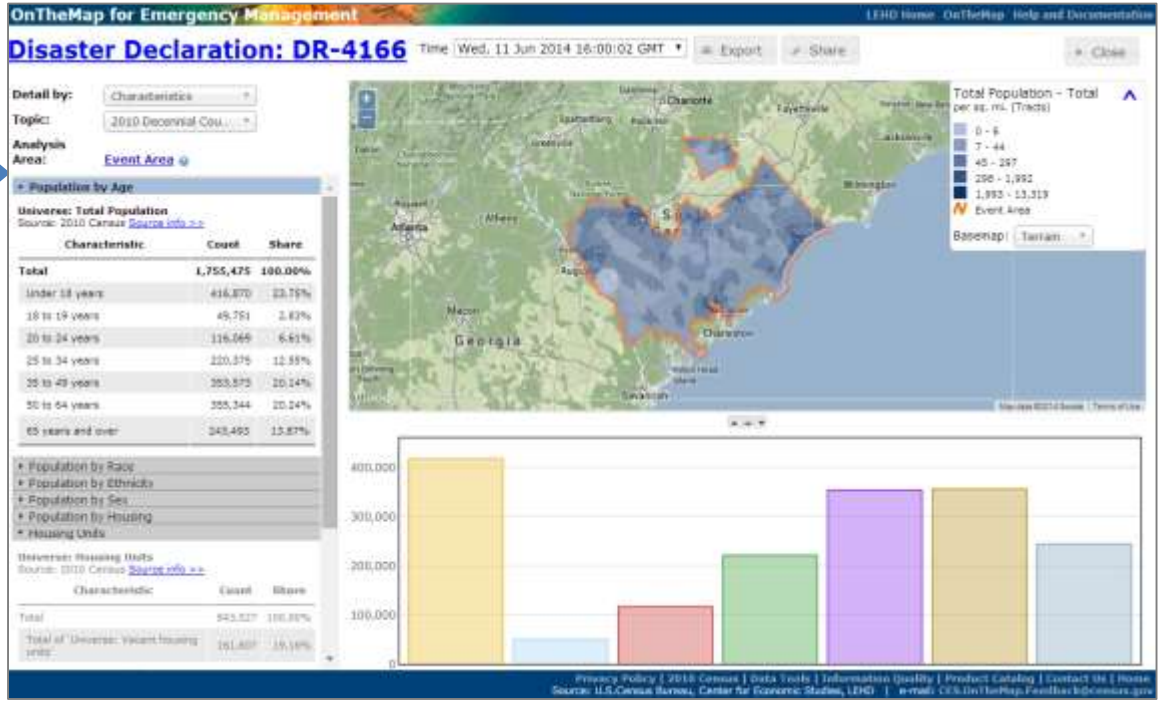
Hurricanes, Floods, Winter Storms

Disaster Areas

Wildfires

Demographic & Economic Data

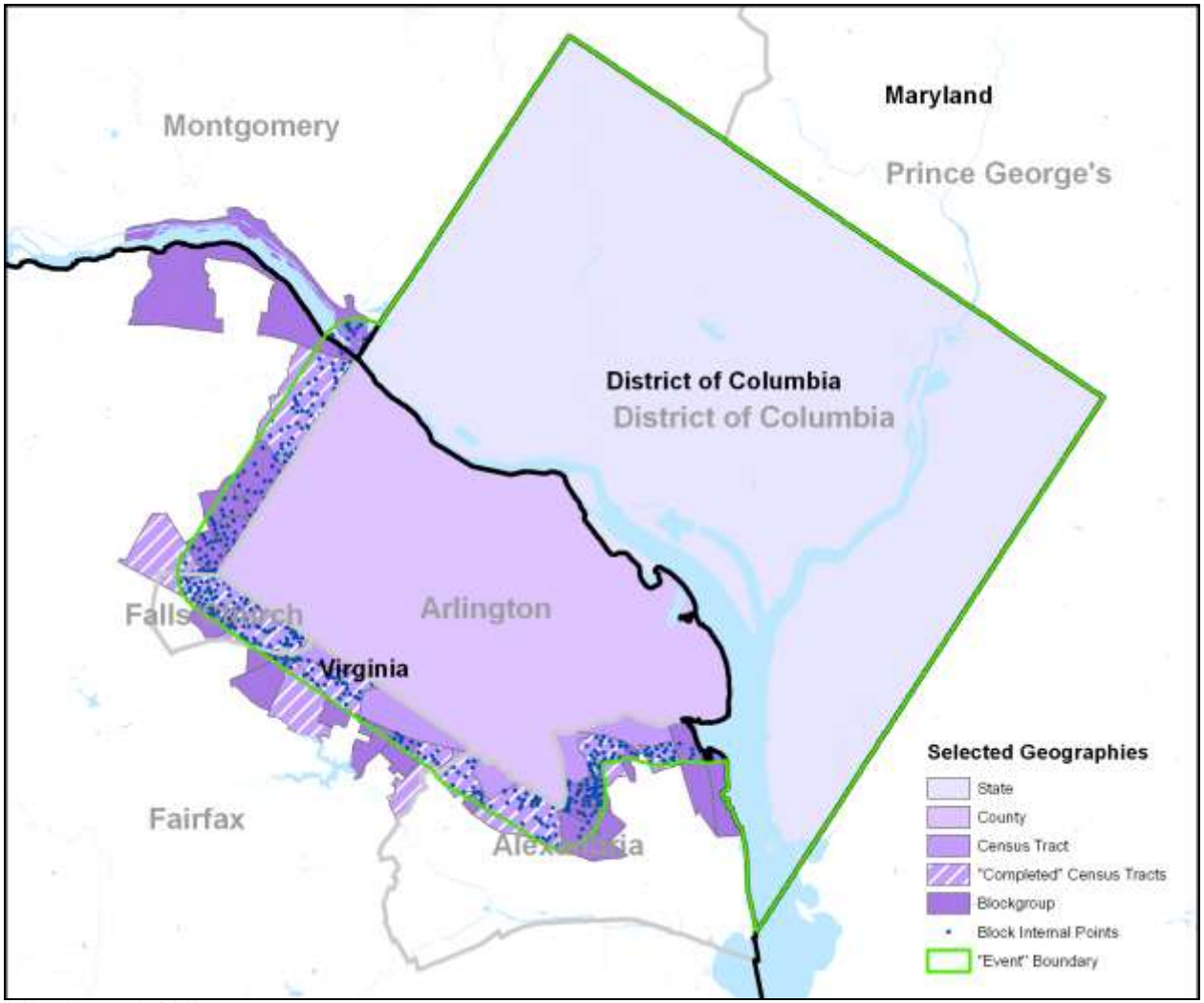
OnTheMap for Emergency Management



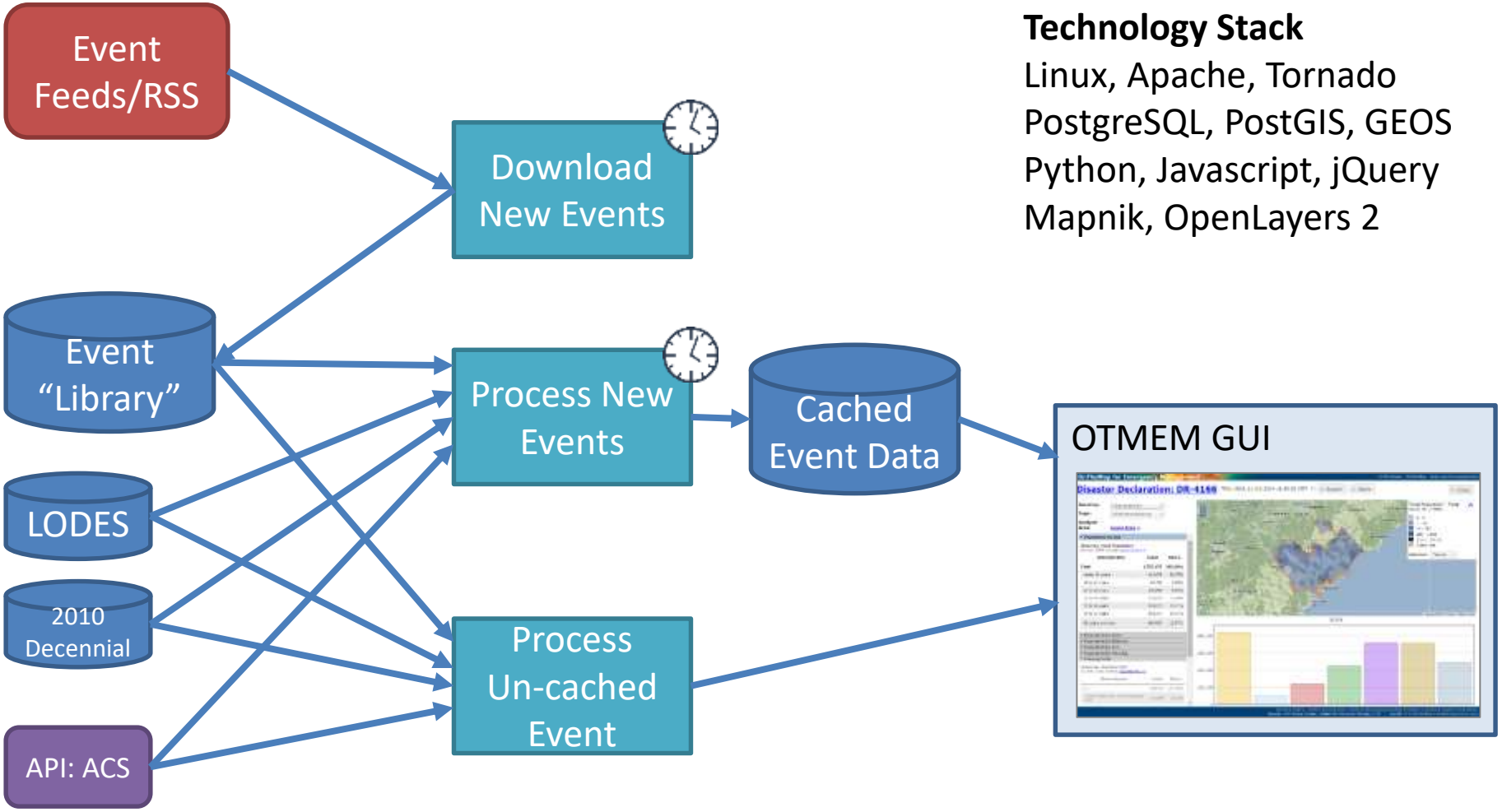
Geographic Approximation

- Event boundaries do not align with statistical boundaries…
- Many possibilities:
  - Inclusive?/Exclusive?
  - Minimize areal difference? MOE?
  - Data-driven?
  - Deterministic/probabilistic?
- Methodology should be consistent with statistical constraints and be explained to users.

Visual Example



Technical Details



Technology Stack
Linux, Apache, Tornado
PostgreSQL, PostGIS, GEOS
Python, Javascript, jQuery
Mapnik, OpenLayers 2

These happen every 4 hours.

Generalizing this Service (FUTURE!?)

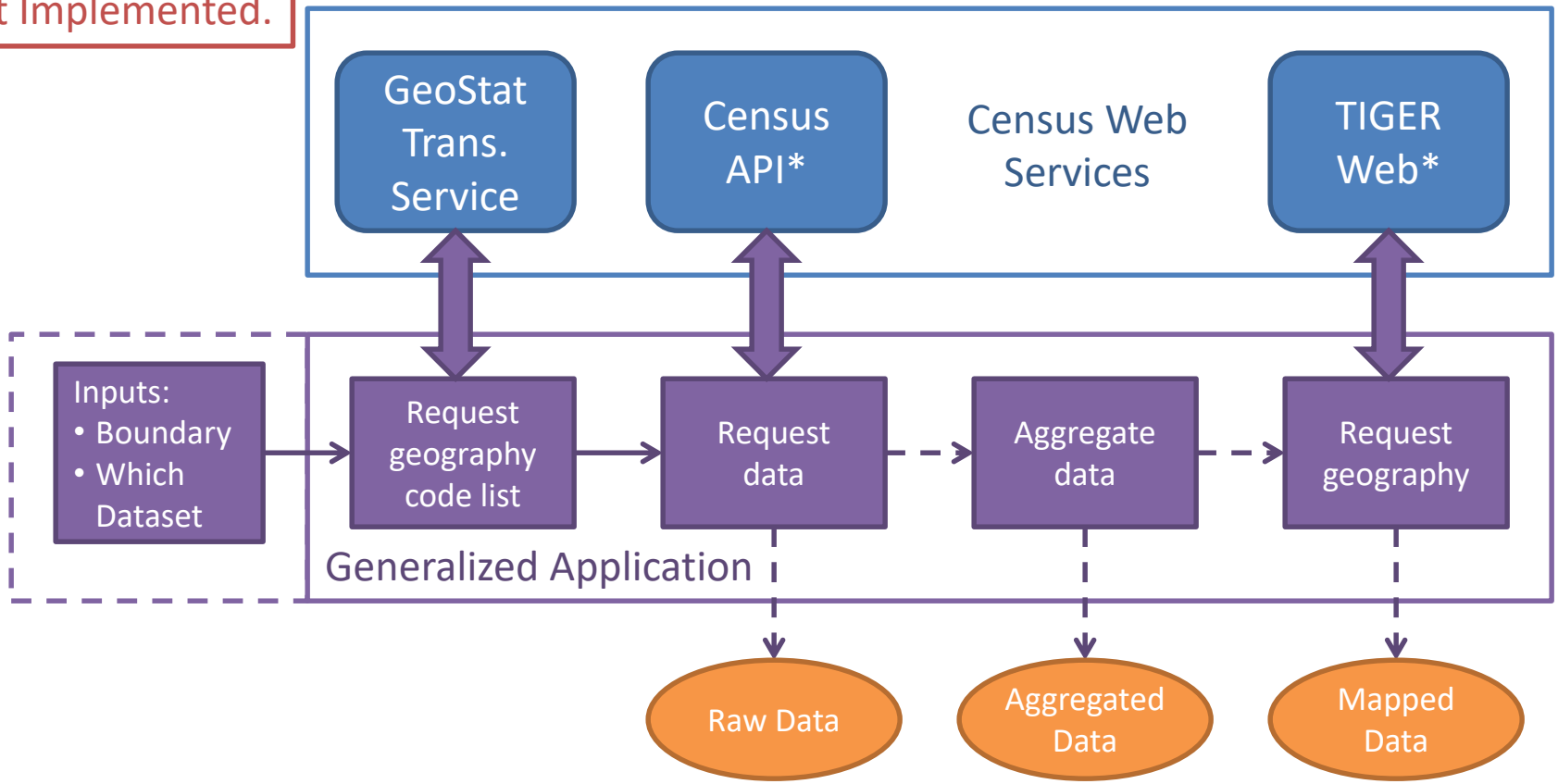
The ultimate goal is to **aggregate data to arbitrary boundaries**.

1. Start with an area of interest (boundary).
2. What data do we want to aggregate?
3. What geography matches that data?
4. What aggregation algorithm should we use?
   - Minimize MOE?
   - Inclusive/exclusive?
   - Areal/population?
   - Other requirements?

Now make it into a web service! What does that look like?

"GeoStatistical Transformation Service"

Conceptual.
Not Implemented.



* These services currently exist.

Potential Challenges

- Lots of datasets → Lots of geography → Large database
- Some algorithms may be slow or scale badly with query size
- Generally non-cacheable (mostly unique queries)
- Need to communicate clearly to the user community:
  - Simple/easy to explain algorithms or defaults for basic users
  - Expanded set of fully documented choices for advanced users
- "Bad" algorithms
  - Though there may be scope for letting users implement their own algorithms

Thank You

Questions?

Links/Contacts

- LEHD
  - https://lehd.ces.census.gov/
- OnTheMap for Emergency Management
  - https://onthemap.ces.census.gov/em
  - CES.OnTheMap.Feedback@census.gov
- Questions
  - matthew.graham@census.gov