Things that go together

Matt Peters | Greg Bunce

NSGIC Annual Conf | September 2020

UTAH AGRC
LOCATION MATTERS
gis.utah.gov

Goals from the Snowbird Conference Presentation….

- Better public access
- Internal-workflow conflicts resolved
- Popularity of data as a service was driving us
- We wanted to focus efforts on data that is being used
- Take care of a bloated state GIS database

What we had...

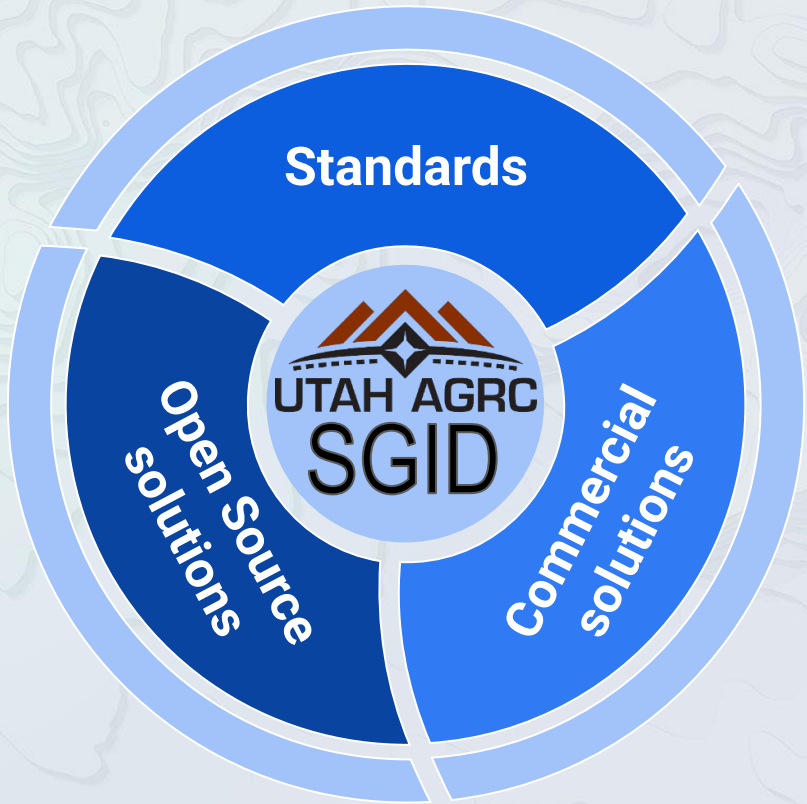
- SQL ESRI Enterprise Database
  - State network access only
- Combined editing and production DB
- Limited AGOL implementation
- Data download links on website
  - Using Google Drive
- FTP site remnants
- Imagery/basemaps on Google Cloud

AGRC Getting Schooled

Unlock Geospatial Data

This is what we learned from our journey…

- **Standards are important**
  - OGC/ISO, FGDC, W3C
- **Data Models play a key role**
  - NENA; NSDI; DOT
- **Accessibility and Interoperability**
  - Open Source and Commercial Solutions



This is what we learned from our journey...

- **Interoperability (cont'd.)**
  - QA/QC
    - Empty geometry
    - Invalid geometry
    - Duplicate data
  - Domains
    - code = description
  - Metadata
    - What standard to use?
    - www.gis.utah.gov/data (data pages)

Resources

General Policies
https://gis.utah.gov/about/policy/sgid/
Metadata Guidelines
https://gis.utah.gov/about/policy/metadata/
Data QA/QC
https://github.com/agrc/sweeper
Domain Policy
https://docs.google.com/document/d/12bdtmtv-ZVF9g-lFshbobx8DbKX9LfbL44HawLjFTVw/edit

Coded Values:

Code	Description
A	Class A - State
B	Class B - County
C	Class C - Municipality
D	Class D - County, Other
F	Federal Agency Maintenance

Coded Values:

Code	Description
Bike	Bike
Pedestrian	Pedestrian
Multiuse	Multiuse

This is what we learned from our journey...

- **Coordination and Data Aggregation**
  - When ingesting from other agencies consider
    - Naming convention
    - Metadata
    - Data Quality (QA/QC)
    - AGOL / Open Data

AGRC Getting Schooled

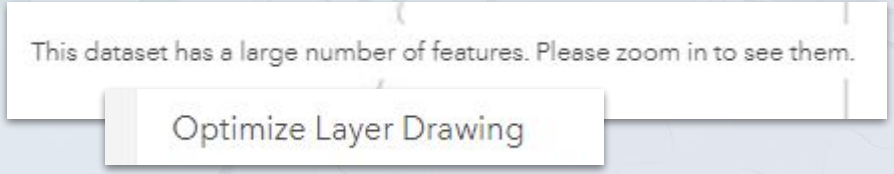
- **Moving to the Cloud**
  - Azure, AWS, GCP, etc. (plenty of options and configurations)
  - VPN tunnel
  - The comfort of moving out of the state network
- **Performance Testing**
  - Internal testing
  - Public beta period (issue tracking)

Google Cloud

PostGIS

esri ArcGIS Online

Hub



The dawning of a new age

Now we are too cool for school

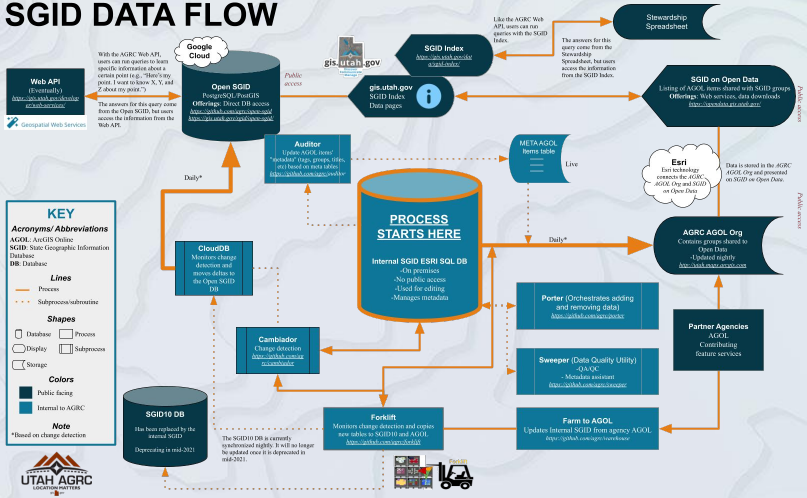
This is where we ended up - where we are today

- We made the data process more robust
  - Internal only db
- Two public offerings:
  - PostgreSQL/PostGIS
  - Open Data (via HUB)
- The option of adding much more data from agencies

PostGIS

ESRI Open Data Hub

SGID DATA FLOW



UTAH AGRC
Automated Geographic Reference Center

Meet the players...

The importance of organizing tools and processes and procedures around the data lifecycle to stay organized

Porter
github.com/agrc/porter

Auditor
github.com/agrc/auditor

Farm to AGOL
github.com/agrc/warehouse

Conductor

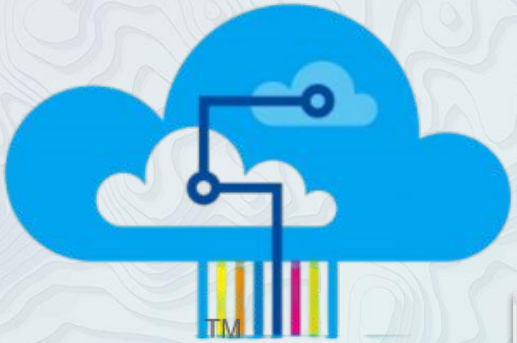
Sweeper
github.com/agrc/sweeper

Cambiador
github.com/agrc/cambiador

Forklift
github.com/agrc/forklift

How we formed the system

- Project based
- Automation
- Python
- Cloud resources (storage, compute, etc.)



Meet the players...

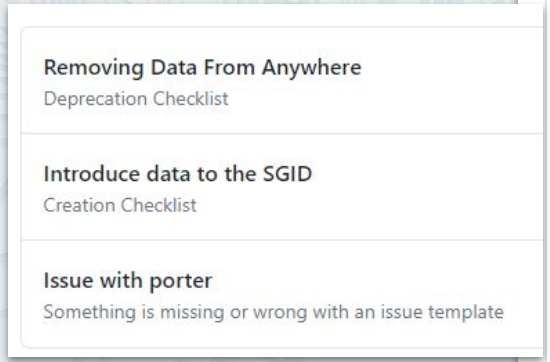
Porter and Conductor

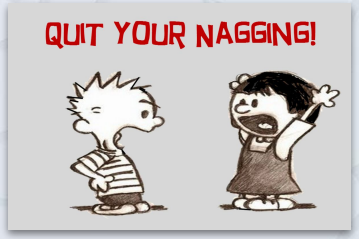
- **Porter -** handles the details
- **Conductor -** does the nagging

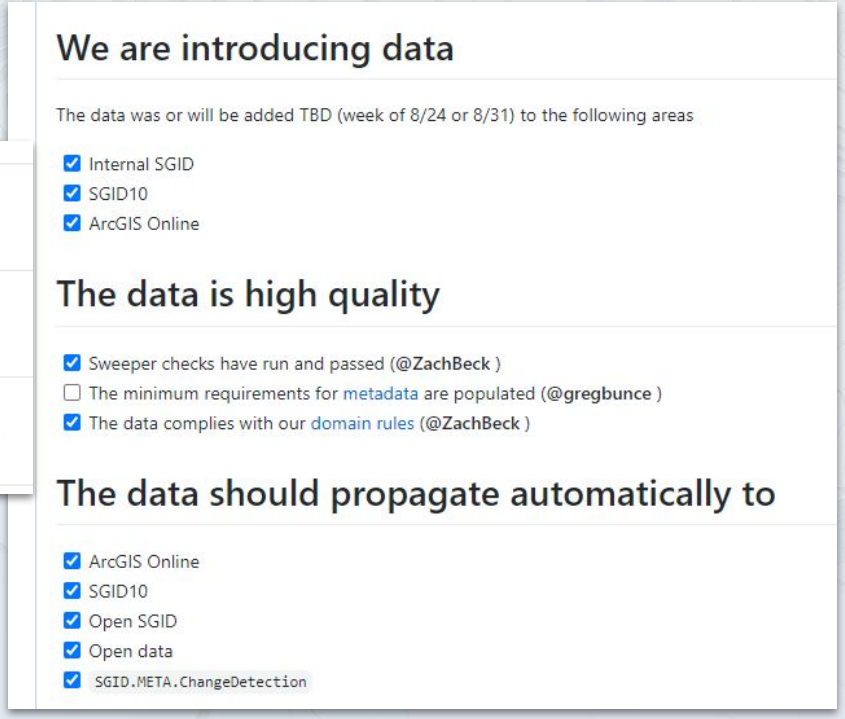
Porter

github.com/agrc/porter

Conductor







Meet the players...

Sweeper and Cambiador

- **Sweeper -** cleans the data (QA/QC)
- **Cambiador -** inspects the data for changes (change detection)

Sweeper

github.com/agrc/sweeper

- Remove duplicates
- Remove Empty Geometries
- Add standard metadata
- Check for minimum address components

Cambiador

github.com/agrc/cambiador

- Checks for data changes
- Hashing (using XXHash)
- Determines what data needs to be updated in public sources

Meet the players...

Auditor and 'Farm to AGOL'

- **Auditor -** ensures that all the required elements exist and are accurate before moving the data is moved to AGOL/Open Data
- **Farm to AGOL -** imports agency data into our in-house enterprise database

Farm to AGOL

github.com/agrc/warehouse

Auditor

github.com/agrc/auditor

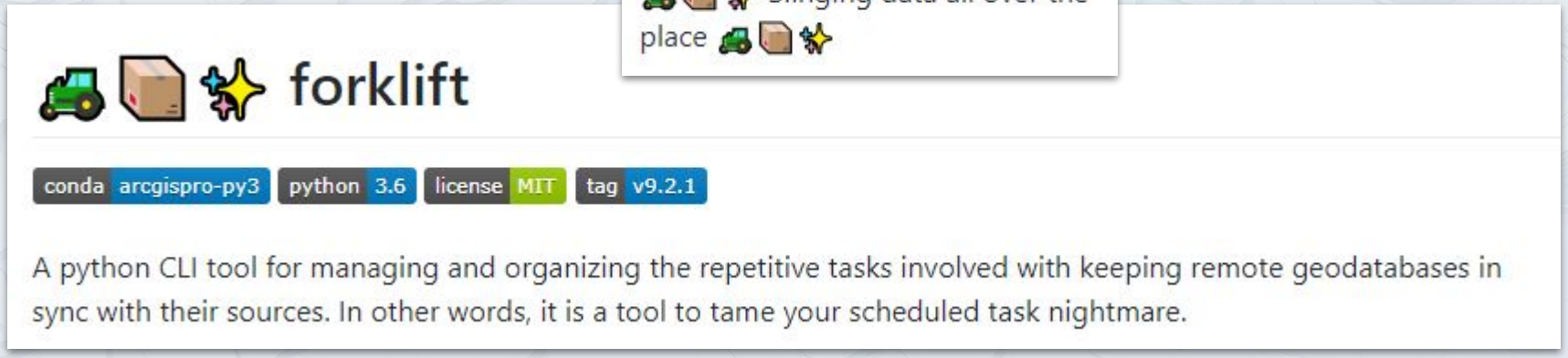
- *Title*
- *Group*
- *Folder*
- *Metadata*
- *Tags*
- *Description note for shelved/static items*
- *Path to appropriate thumbnail image*
- *Sets the flag for delete protection*
- *Sets the flag to 'Allow others to export to different formats', which opens up GDB downloads in Open Data*
- *Marks the item as Authoritative*

Meet the players...

Forklift

- **Forklift -** moves the pallets (data) around the warehouse (platforms/geodatabases)
  - Enterprise database to AGOL
  - Enterprise database to PostgreSQL/PostGIS

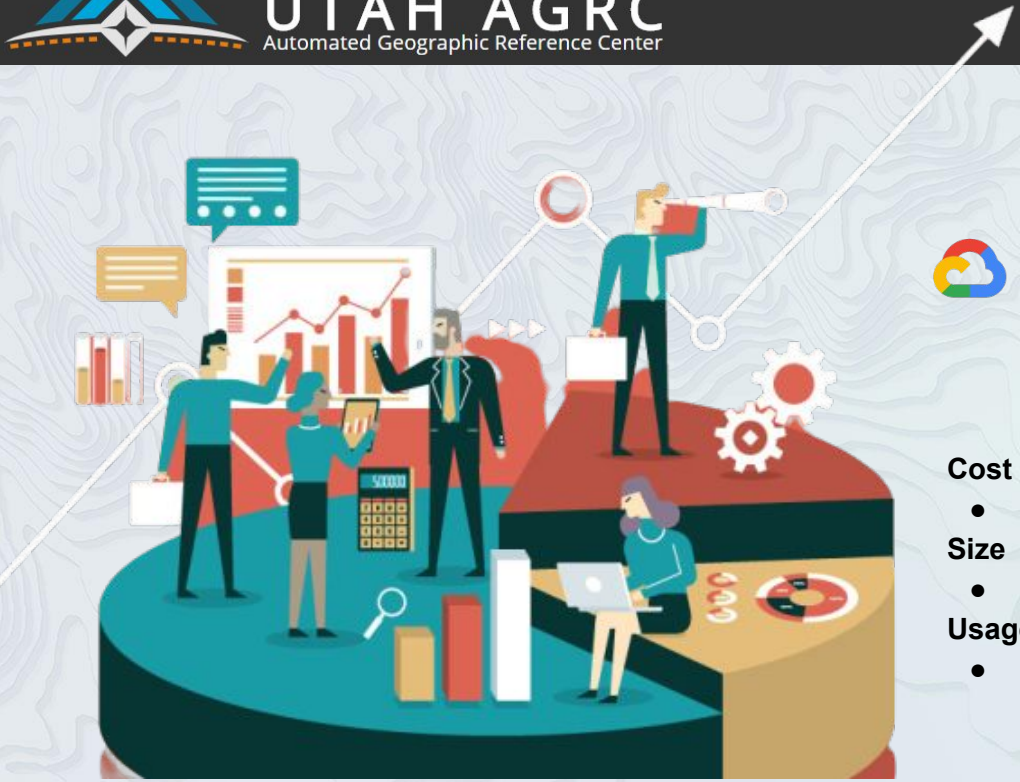
Forklift
github.com/agrc/forklift



UTAH AGRC
Automated Geographic Reference Center

So, what's the GISt of all this?

show me the numbers...



390 Statewide Datasets

Google Cloud

PostGIS

Cost
- **$140/mo** (state db storage)

Size
- **10GB** (db tables)

Usage
- 25 to 35 database connections a day

esri ArcGIS Online

Hub

Cost
- **$400/mo** (state db storage)

Size
- **16GB** (hosted feature layers)

Usage

30 day average requests
- Address Points
  - 6,000/day
- County Boundaries
  - 1,200/day

Framework is Key

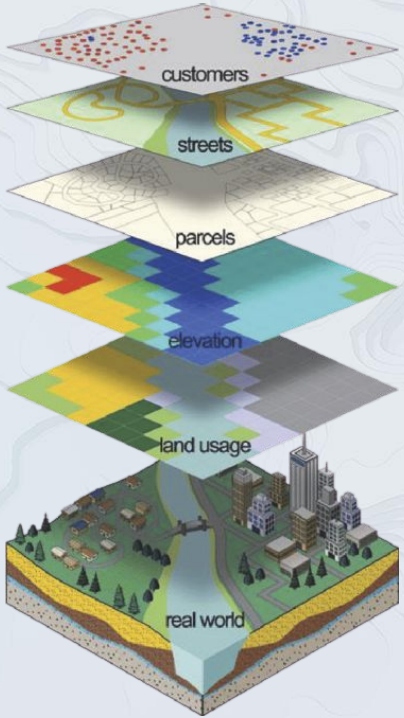
Framework layers are still king

- **“Tier 1” layers**
  - **Roads, Address Points, Parcels, Political Boundaries, Water**

- These layers are “running the business of government”
  - money and revenue is often tied to the inventory and availability of these layers
  - making them desirable to maintain

- Create processes to keep these layers current, complete, clean, available, and easily accessible

- The data we thought was important years ago is the same data that is important today
  - By taking care of these layers, we hit the majority of our users’ and needs



Things that go together

Matt Peters | Greg Bunce

NSGIC Annual Conf | September 2020