

Implementing a Rating System to Enhance Data Quality



Greg Bunce | Nathan Kota

Utah's SGID (State Geographic Information Database)



The early years...



Discussions about statewide geographic data



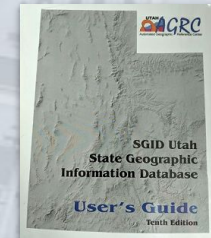
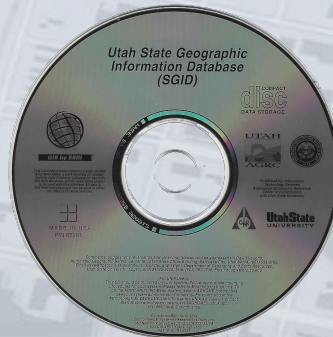
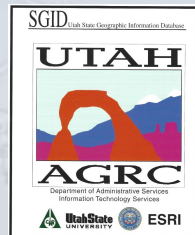
State signed contract with Esri and established GIS office



SGID established via legislative action



SGID meets the world wide web. Interactive website (ArcIMS)



Utah's SGID (State Geographic Information Database)

The early years...

1970s

Discussions about statewide geographic data



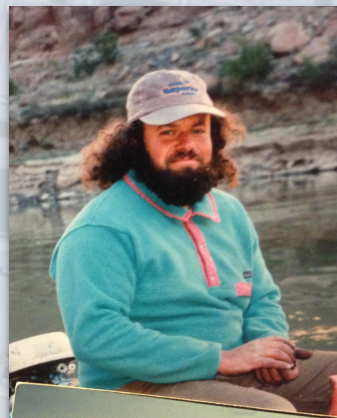
1980s

State signed contract with Esri and established GIS office



1990s

SGID established via legislative action



2000s

SGID meets the world wide web. Interactive website (ArcIMS)



Database

PostgreSQL
+PostGIS

UGRC website

data pages
& index

Web API

http service for
geocoding and
search

Discover

basemaps,
imagery, web app

AGOL

hosted feature
services

Open Data

opendata.gis.utah.gov



SGID

STATE GEOGRAPHIC
INFORMATION DATASOURCE

The SGID aims to be the primary source for statewide GIS data in Utah by creating a single portal to datasets created by many different agencies and organizations. With such a broad scope, SGID datasets live in many different places depending on their type, availability, and popularity.

"A collection of data resources distributed among a variety of platforms."

Database

PostgreSQL
+PostGIS

UGRC website

data pages
& index

Web API

http service for
geocoding and
search

Discover

basemaps,
imagery, web app

AGOL

hosted feature
services

Open Data

opendata.gis.utah.gov



SGID governance provided by:



gis.utah.gov/about/policy/sgid



Policies and procedures

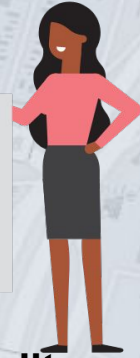


The screenshot shows the UGRC website's 'SGID Policies' page. The page header includes the UGRC logo and navigation links for News, Data, Developer, About, and Status. The main content area is titled 'SGID Policies' and contains a paragraph explaining that the policies govern data interaction in the SGID. Below this is a bulleted list of policy categories: SGID Dataset Qualifications (General, Index, Open Data, Open SGID), Dataset Deprecation and Removal (General Removal, Reasons for Removal, Static, Shelved), and Miscellaneous Policies (Dates in Dataset Names, Internal Database Standards, Editing Policies). The page also includes a section for 'A Brief Word on Language' and a list of terms used intentionally, such as 'Must' and 'Should'.

SGID governance provided by:



Policies and procedures



Auditor

checks AGOL feature
service items
github.com/agrc/auditor



Automation
& process tools



Porter

tracks additions,
replacements, deletions
github.com/agrc/porter



Farm to AGOL

steward data>internal gdb>SGID
github.com/agrc/warehouse



Cambiador

gdb change detection
github.com/agrc/cambiador



Sweeper

data cleaner
github.com/agrc/sweeper



Forklift

slinging data to and fro since 2015
github.com/agrc/forklift



✓ Policies and procedures

✓ Automation & process tools

✓ Defined data flow

opensgid.agrc.utah.gov



Database

PostgreSQL
+PostGIS

UGRC website

data pages
& index

AGOL

hosted feature
services

utah.maps.arcgis.com

gis.utah.gov

Open Data

opendata.gis.utah.gov

opendata.gis.utah.gov

Web API

http service for
geocoding and
search

api.mapserv.utah.gov

Discover

basemaps, imagery,
web app

gis.utah.gov/discover

A Strategic Plan for the SGID

3 year plan

- **Branding**
 - logo
- **Renaming**
 - From database to datasource
- **Collaboration v2**
 - ESRI HUB / Portal
 - OpenStreetMap
- **Data rating systems**
 - (SDI dashboard)



SGID

STATE GEOGRAPHIC
INFORMATION DATASOURCE

Data Rating Systems

Where?

- Clean up the database
- Framework data
 - Data Tiers



Collective energy



Data Rating Systems

Why?

- Where are we missing the mark?
- What can we improve or remove?
- User confidence score
- Internal health check
- Feedback



Data Rating Systems

How?

The Challenge

- Translate qualitative to quantitative

Metrics

- Numeric measures that describe the data

Weights

- A weight given to each metric/measure based on its importance



Data Rating Systems

Nine Metrics for Measuring Data



Authoritative

- “Statewide authoritative” - *Would the steward promote this data?*

Business Systems

- *Is this layer used in state business systems? Critical business systems?*

Completeness

- *Are there any gaps or holes in the data? Is it statewide?*

Data quality

- *How accurate is the attribute and geometry quality?*

Foundational layer

- *Is this layer a building block to build/create/derive other layers?*

Interoperability

- *Does the data look the same on all systems and file formats?*

Metadata completeness

- *Data has [minimum required metadata](#) with [license and disclaimer](#)?*

System of Record

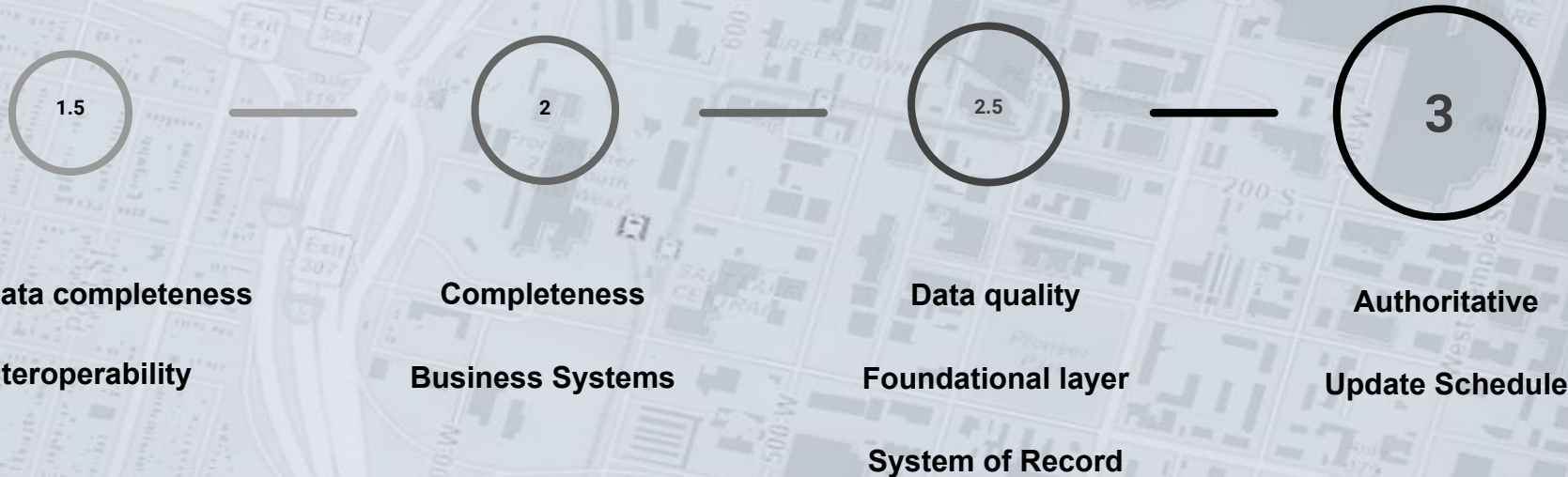
- *Is this data directly mandated by statute or does it support other data or programs mandated by statute?*

Update Schedule/Currency

- *Does it use schedule - is it adhering to the schedule?*

Data Rating Systems

Weighting the Metrics Based on Importance



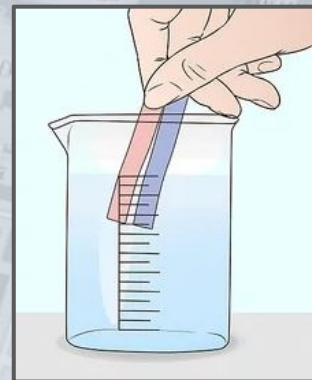
Data Rating Systems

Let's do the numbers...



Layer Name	Weighted Score
County Boundaries	100.00%
Land Ownership	100.00%
Census Block Groups	100.00%
Sales Tax Areas	100.00%
Political Districts	100.00%
Address Points	100.00%
Road Centerlines	96.34%
Water Related Land Use	93.90%
Big 5 County Parcel Layers	87.80%
Utah State Boundary	85.37%
Utah Statewide Parcels	80.49%
Rural County Parcel Layers	75.61%
TURN GPS Stations	70.73%
Small Statistical Health Areas	68.29%
State Fuel Sites	54.88%
Zipcodes	48.78%
Utah Wilderness Areas	46.34%
City Town Locations	39.02%
Grazing Allotments	36.59%
US States	29.27%
Railroads	23.17%
Rural Telecom Boundaries	7.32%

Gold = 80 +
Silver = 50-79
Bronze < 50



Data Rating Systems



Let's do the numbers...

Layer Name	Weighted Score	Completeness	Update Schedule/Currency	Authoritative	Metadata	Interoperability	Data Quality	Foundational	Business System	System of Record	Total
County Boundaries	100.00%	2	2	2	2	2	2	2	2	2	41
Land Ownership	100.00%	2	2	2	2	2	2	2	2	2	41
Census Block Groups	100.00%	2	2	2	2	2	2	2	2	2	41
Sales Tax Areas	100.00%	2	2	2	2	2	2	2	2	2	41
Political Districts	100.00%	2	2	2	2	2	2	2	2	2	41
Address Points	100.00%	2	2	2	2	2	2	2	2	2	41
Road Centerlines	96.34%	2	2	2	2	1	2	2	2	2	39.5
Water Related Land Use	93.90%	2	2	2	2	2	2	1	2	2	38.5
Big 5 County Parcel Layers	87.80%	2	2	2	2	2	1	2	2	1	36
Utah State Boundary	85.37%	2	2	2	1	2	2	2	1	1	35
Utah Statewide Parcels	80.49%	2	1	2	2	2	1	2	2	1	33
Rural County Parcel Layers	75.61%	1	1	2	2	2	1	2	2	1	31
TURN GPS Stations	70.73%	2	2	2	2	2	2	0	1	0	29
Small Statistical Health Areas	68.29%	2	1	2	2	2	2	0	2	0	28
State Fuel Sites	54.88%	1	1	2	1	2	2	0	1	0	22.5
Zipcodes	48.78%	2	1	0	2	2	1	1	1	0	20
Utah Wilderness Areas	46.34%	2	1	0	1	2	2	0		1	19
City Town Locations	39.02%	2	1	0	1	2	1	0	1	0	16
Grazing Allotments	36.59%	2	0	1	0	2	2	0	0	0	15
US States	29.27%	1	1	0	1	2	1	0	0	0	12
Railroads	23.17%	1	0	0	0	2	1	0	1	0	9.5
Rural Telecom Boundaries	7.32%	0	0	0	0	2	0	0	0	0	3

Data Rating Systems

Automate the Ratings...

Highly unlikely

Highly Likely



System of Record

Foundational layer

Authoritative

Metadata completeness

Business Systems

Update Schedule

Interoperability

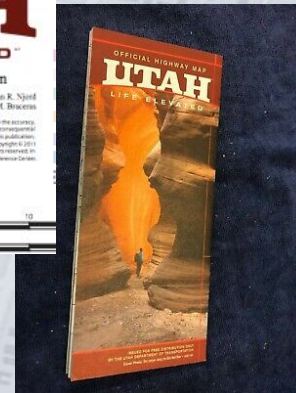
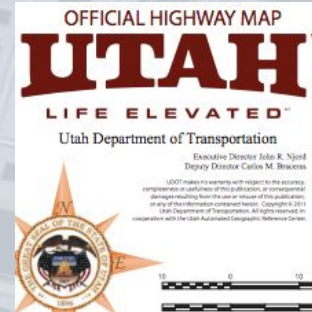
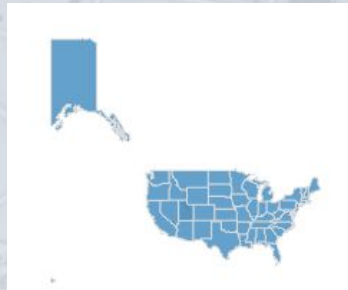
Completeness

Data Quality

Data Rating Systems

Insights and Surprises

- Scoring
- Cities and Towns layer
- Roads layer
 - Coded domains
- US States



UGRC



**Utah
Geospatial
Resource
Center**

Location matters

Nathan Kota

email: nathankota@utah.gov

twitter: <NULL>

Greg Bunce

email: gbunce@utah.gov

twitter: [@bunce_greg](https://twitter.com/bunce_greg)