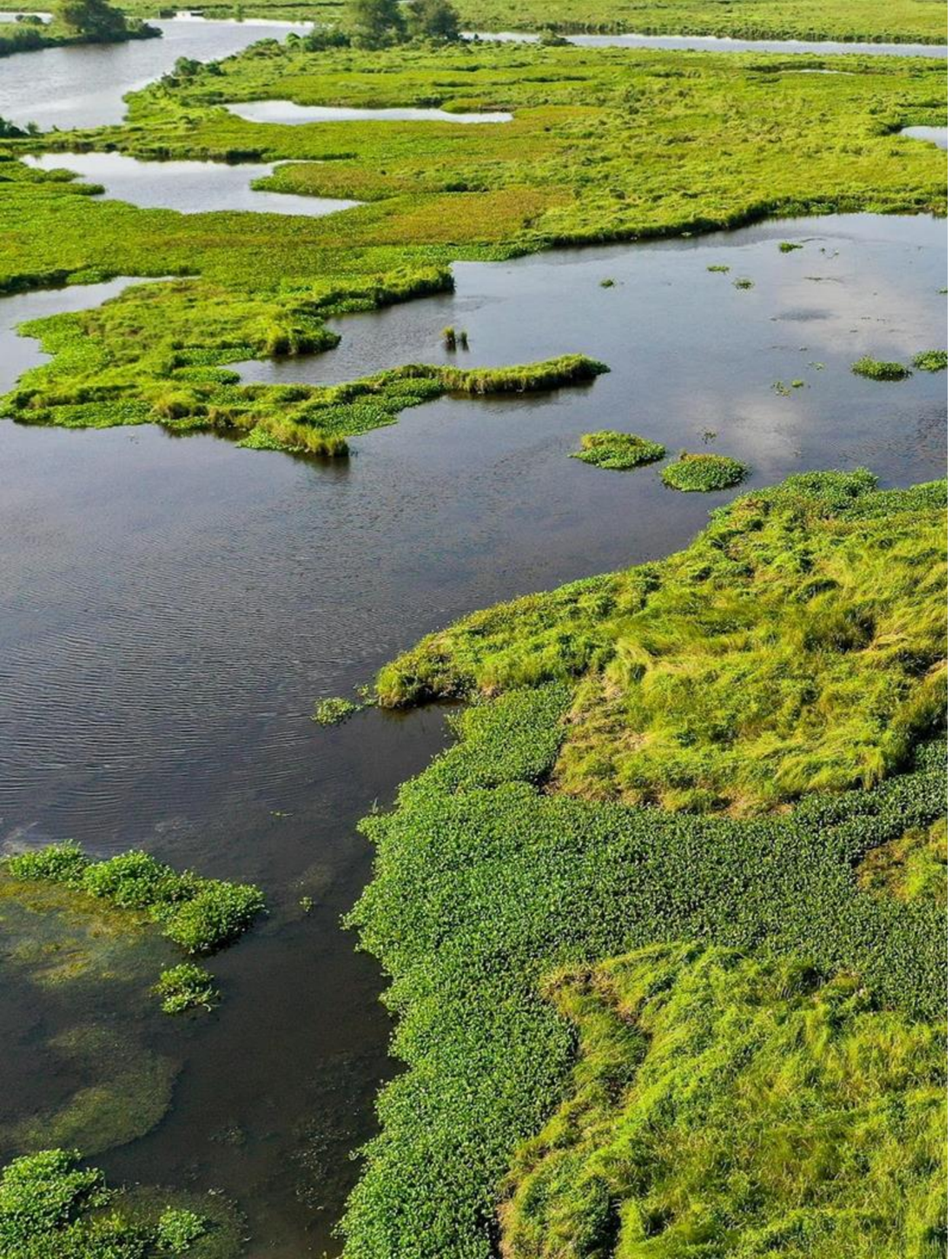


2029 COASTAL MASTER PLAN
COMMITTED TO OUR COAST

SMART AND LAZY

INTELLIGENT DATA STORAGE FOR ON-THE-FLY ANALYSIS AND VISUALIZATION

MATT YODER, PITTSBURGH SUPERCOMPUTING CENTER

MAY 21, 2025

PSC AND ME

Pittsburgh Supercomputing Center

- Research computing center founded in 1986
- Provides research computing capability, education, and expertise to the national research community

Matt Yoder

- Software engineer at PSC since 2021
- Background in urban planning and GIS
- Interests: open source software and decision support systems

THESIS

How we store data matters!

THESIS (EXTENDED)

Choosing **smart data formats**
allows for **lazy visualization**,
saving time and effort.

WHAT IS LAZY VISUALIZATION?

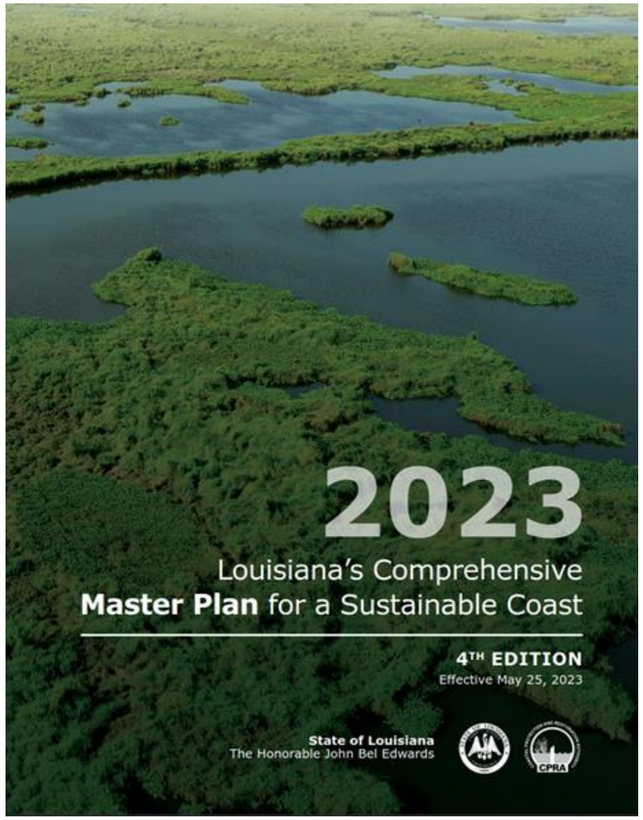
- **Visualizations are rendered in real-time as needed.**
  - Using the original data source
  - No need for a separate display format
- **Simple analysis can be performed on the fly.**
  - Calculating statistics
  - Spatial aggregation
  - Dynamic color ramps

WHAT ARE SMART DATA FORMATS?

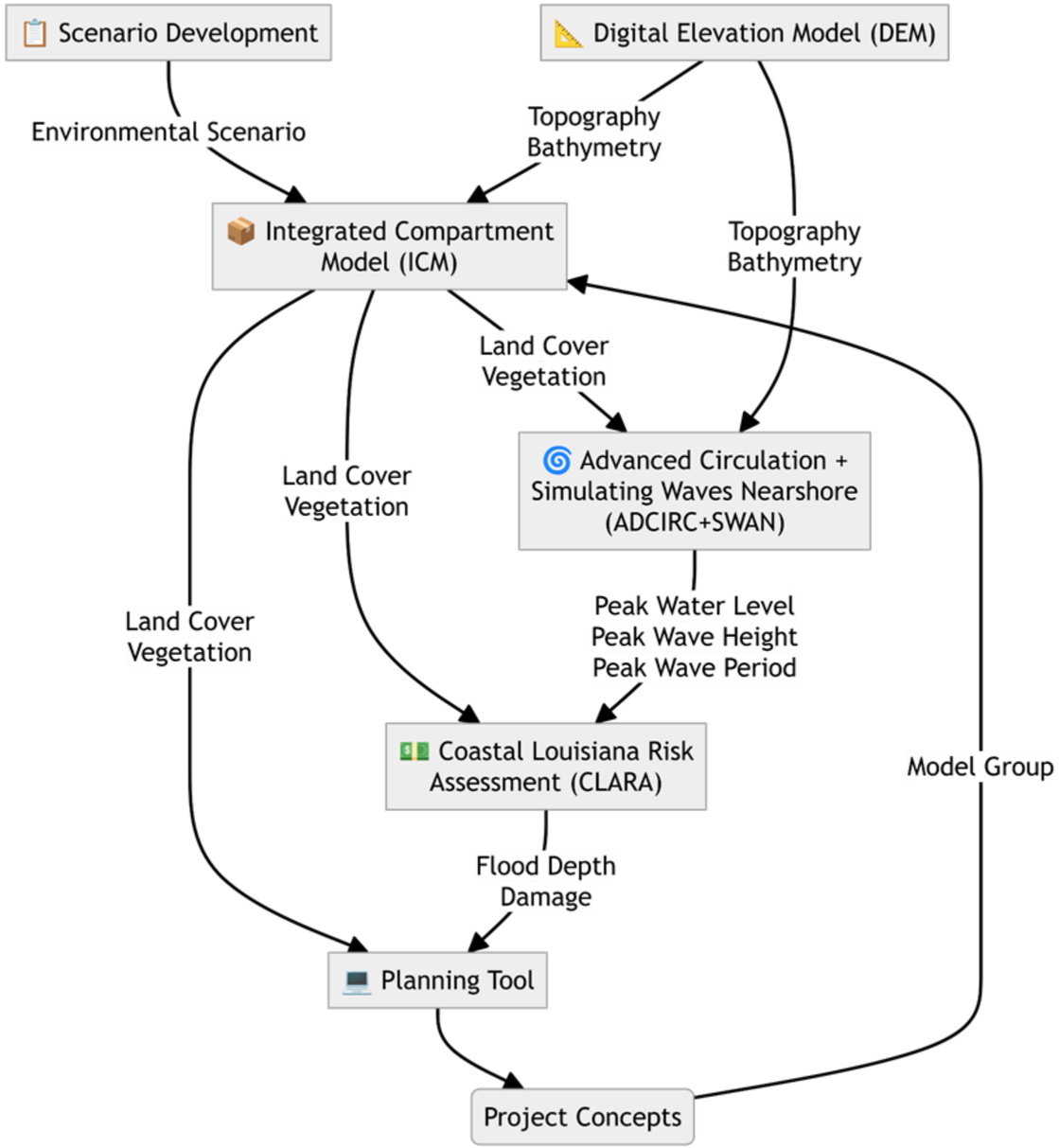
- **Store data efficiently**
  - Indexing
  - Random access
  - Compression
- **Data structure matches access patterns**
  - Column-oriented storage (tabular data)
  - Built-in overviews (image data)

CASE STUDY: LOUISIANA'S COASTAL MASTER PLAN

- **Researchers are using PSC's Bridges-2 supercomputer to model the future of Louisiana's coastline.**
  - Predict land loss, vegetation change, and storm-surge based flooding
  - Evaluate the impacts of mitigation projects
- **The 2023 plan produced over 600 TB of data.**
  - Tabular attribute data
  - Raster data



MASTER PLAN MODELS

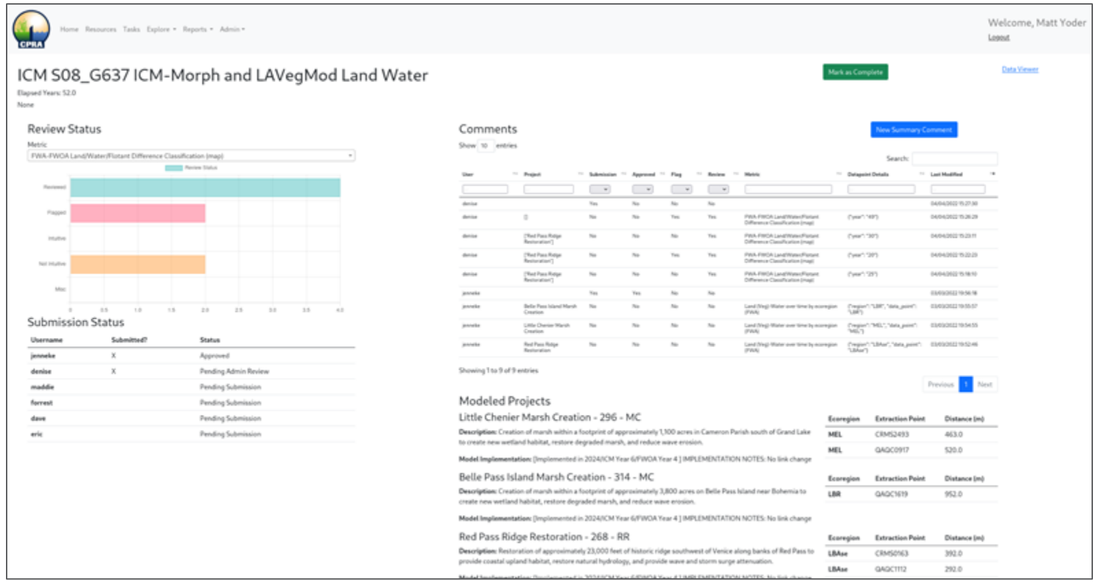


QUALITY CONTROL / QUALITY ASSURANCE

2012 and 2017 Master Plans

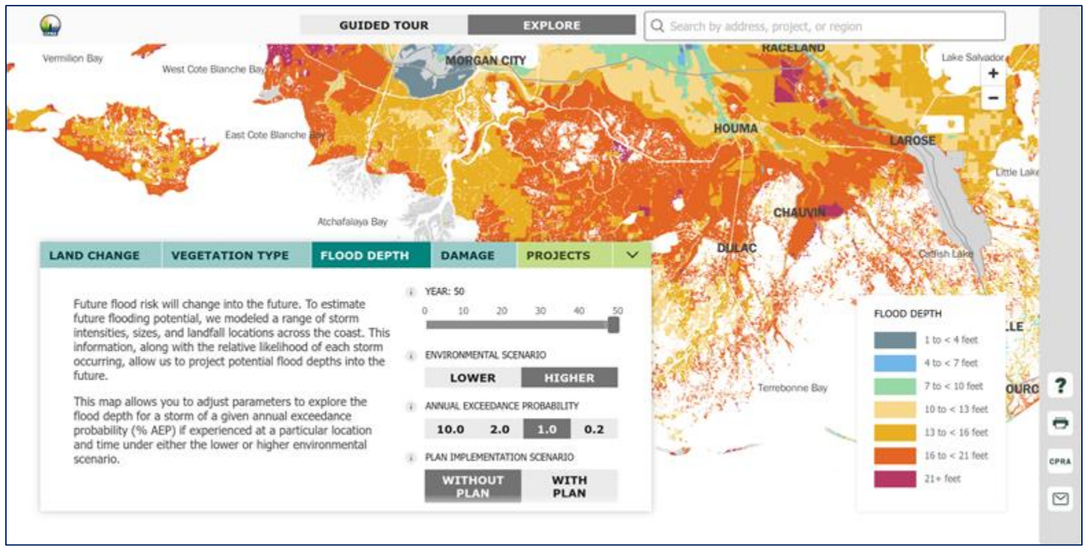
ArcGIS

2023 Master Plan

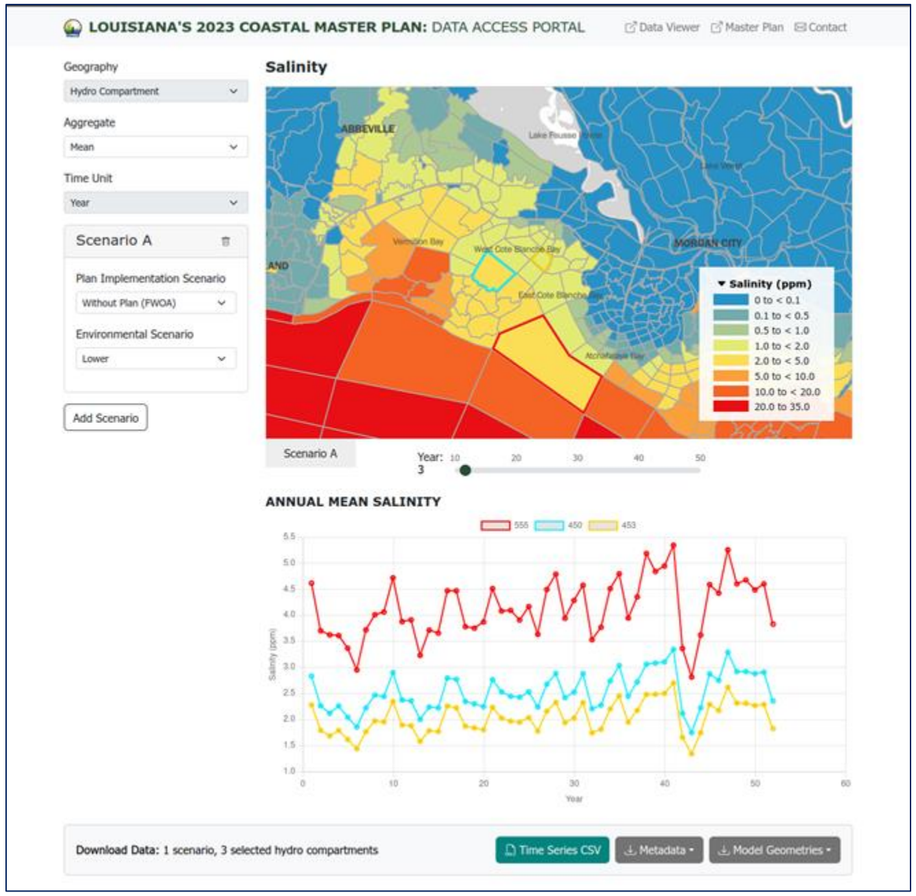


DATA VISUALIZATION

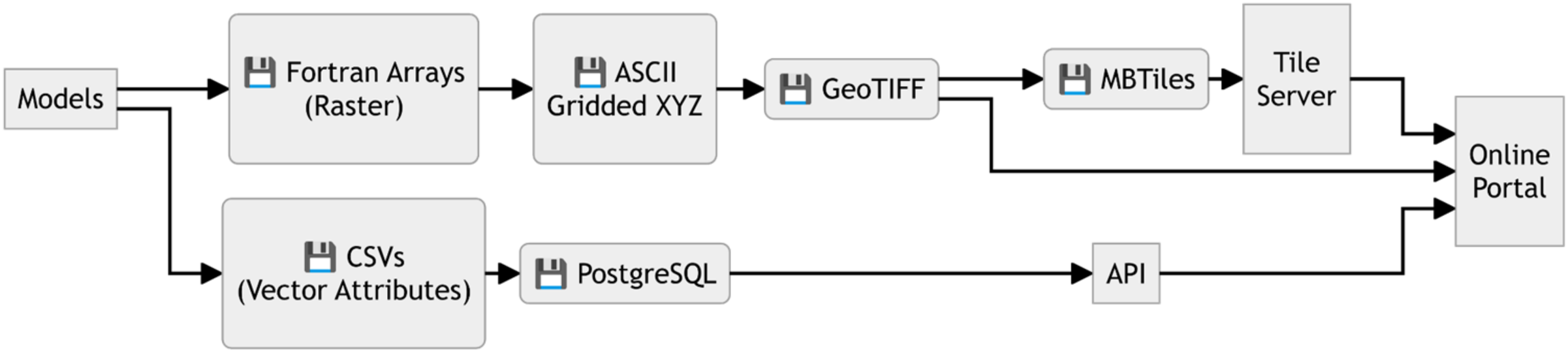
Master Plan Data Viewer



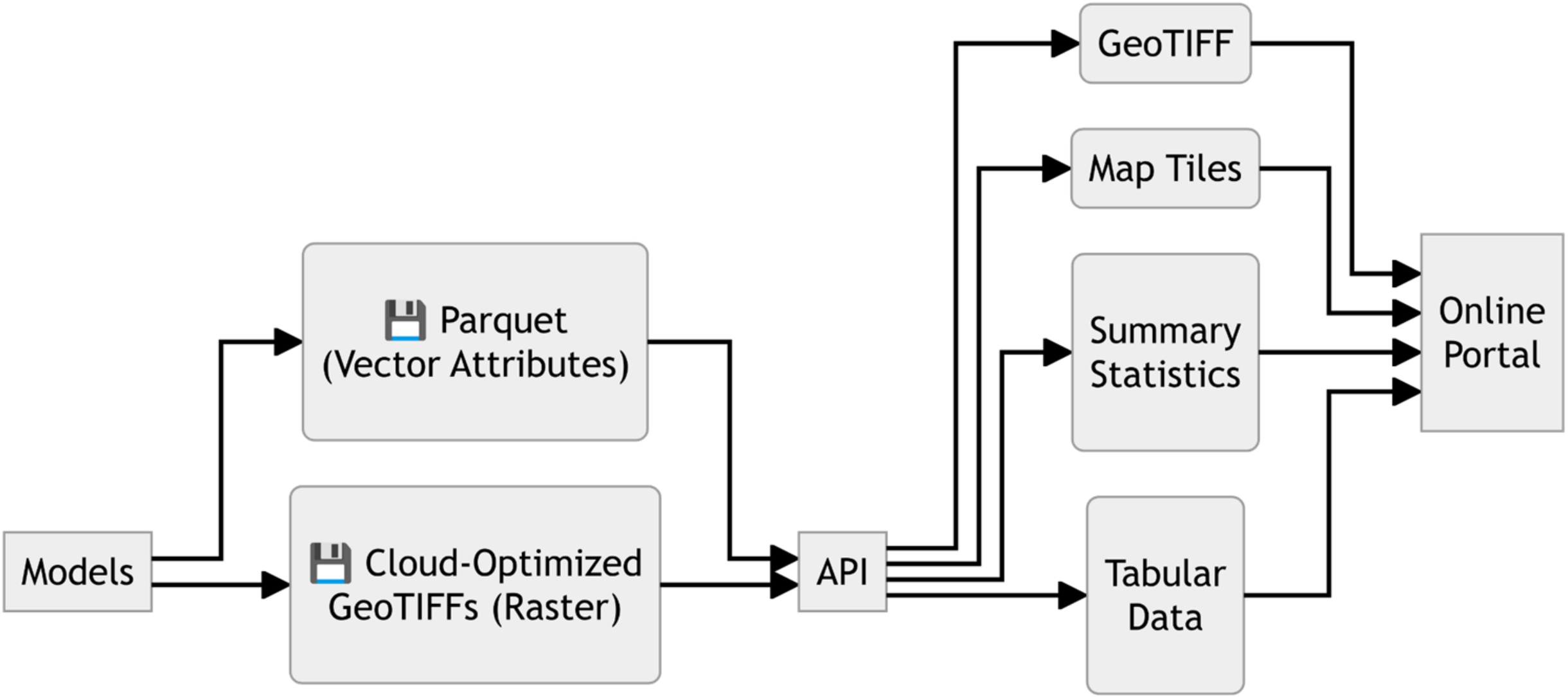
Master Plan Data Access Portal



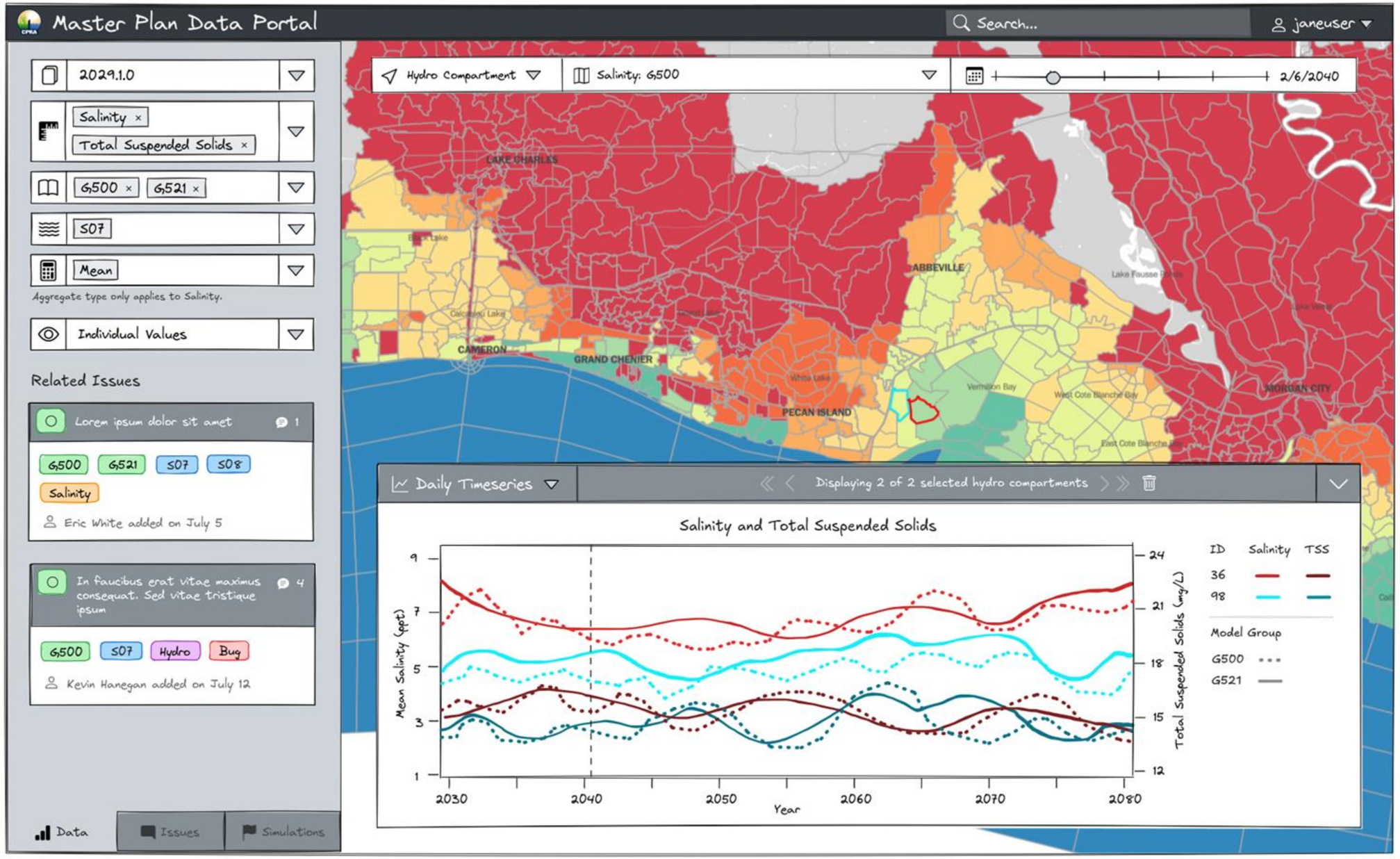
2023 PLAN DATA WORKFLOW



2029 PLAN DATA WORKFLOW



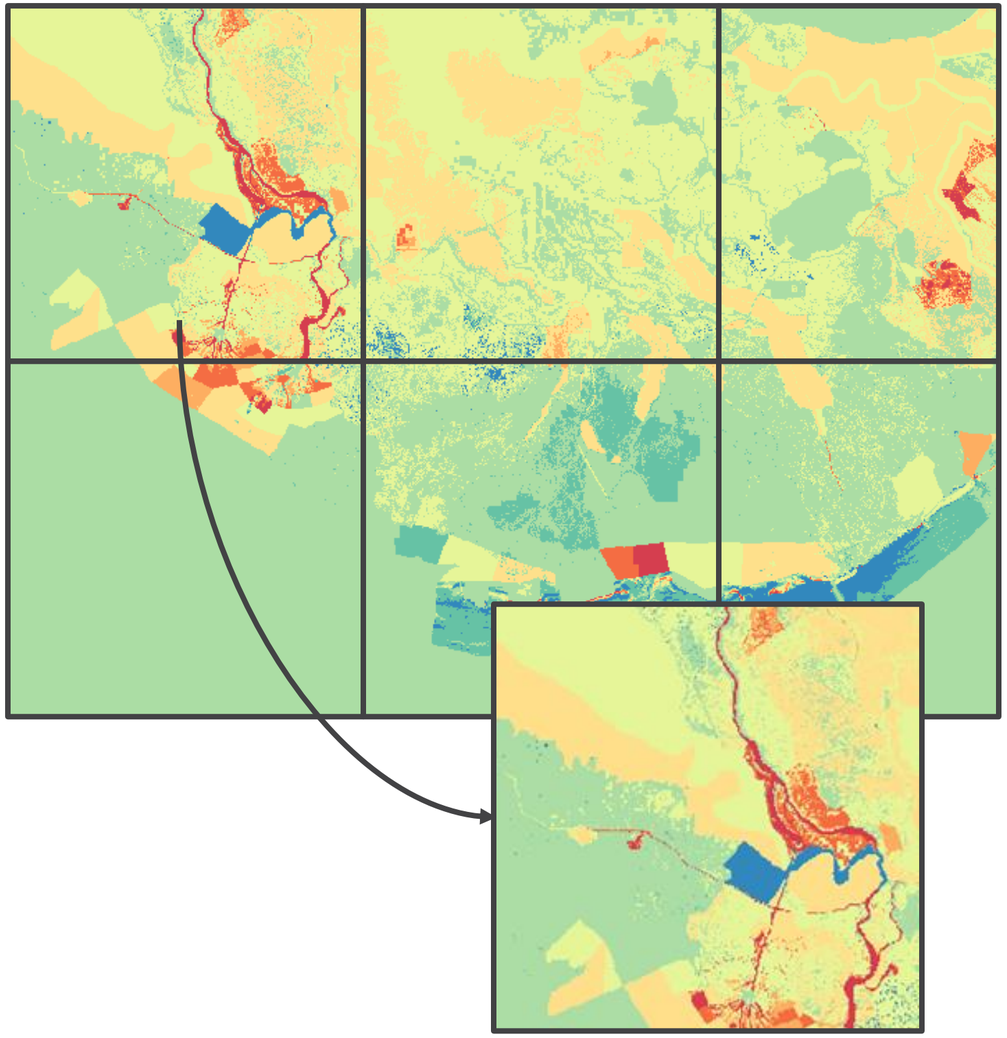
2029 PORTAL


Master Plan Data Portal
Search...
janeuser
2029.1.0
Salinity ×
Total Suspended Solids ×
G500 × G521 ×
S07
Mean
Aggregate type only applies to Salinity.
Individual Values
Related Issues
Lorem ipsum dolor sit amet 1
G500 G521 S07 S08
Salinity
Eric White added on July 5
In faucibus erat vitae maximus consequat. Sed vitae tristique ipsum 4
G500 S07 Hydro Bug
Kevin Hanegan added on July 12
Data
Issues
Simulations
Hydro Compartment
Salinity: G500
2/6/2040
CAMERON
GRAND CHENIER
PECAN ISLAND
ABBEVILLE
MORGAN CITY
Daily Timeseries
Displaying 2 of 2 selected hydro compartments
Salinity and Total Suspended Solids
Mean Salinity (ppt)
Total Suspended Solids (mg/L)
Year
2030
2040
2050
2060
2070
2080
ID Salinity TSS
36
98
Model Group
G500
G521

APACHE PARQUET

	flavor (str)	**scoops** (int32)	**sale_date** (date)
Row Group 1	Chocolate	2	2024-05-28
	Vanilla	1	2024-06-01
	Strawberry	3	2024-06-10
Row Group 2	Chocolate	1	2024-06-15
	Chocolate	3	2024-06-15
	Mint	2	2024-07-02
Row Group 3	Strawberry	2	2024-07-08
	Mint	2	2024-07-14

CLOUD OPTIMIZED GEOTIFF (COG)



GeoTIFF with Superpowers

- Tiled
- Internal overviews
- Compression
- Can act like a map service when served over HTTP

DATA STRUCTURE

- **Current structure**
  - 47 variables
  - 28 dimensions
  - 8 grids
  - Dimension values and legends
- **Developed in a shared spreadsheet**
- **Stored in Python library**
  - Serves as a single source of truth (e.g., for data validation, documentation, etc.)

Data Structure

Variables

Label	Name	Model	Subroutine	Data Type	Geographic Units	Time Units	Unit
Inundation Depth	inun	icm	morph	Float32	morph_pixel	annual	m
Topobathymetric Elevation	elev	icm	morph	Float32	morph_pixel	annual	NAVD88 m
Land Type	lnd_type	icm	morph	Int32	morph_pixel	annual	
Vegetation Type: FFIBS	ffibs	icm	lavegmod	Float32	veg_pixel	annual	
Land Change	lnd_chng	icm	morph	Int32	morph_pixel	annual	
Salinity	sal	icm	hydro	Float32	hydro_compartment	daily, monthly, annual	ppt
Daily Tidal Range	trg	icm	hydro	Float32	hydro_compartment	daily, annual	m
Stage (Water Level)	stg	icm	hydro	Float32	hydro_compartment	daily, monthly, annual	NAVD88 m
Sediment Accumulation in	sedacc	icm	hydro	Float32	hydro_compartment	daily,	g/m2/yr

PERFORMANCE TESTING

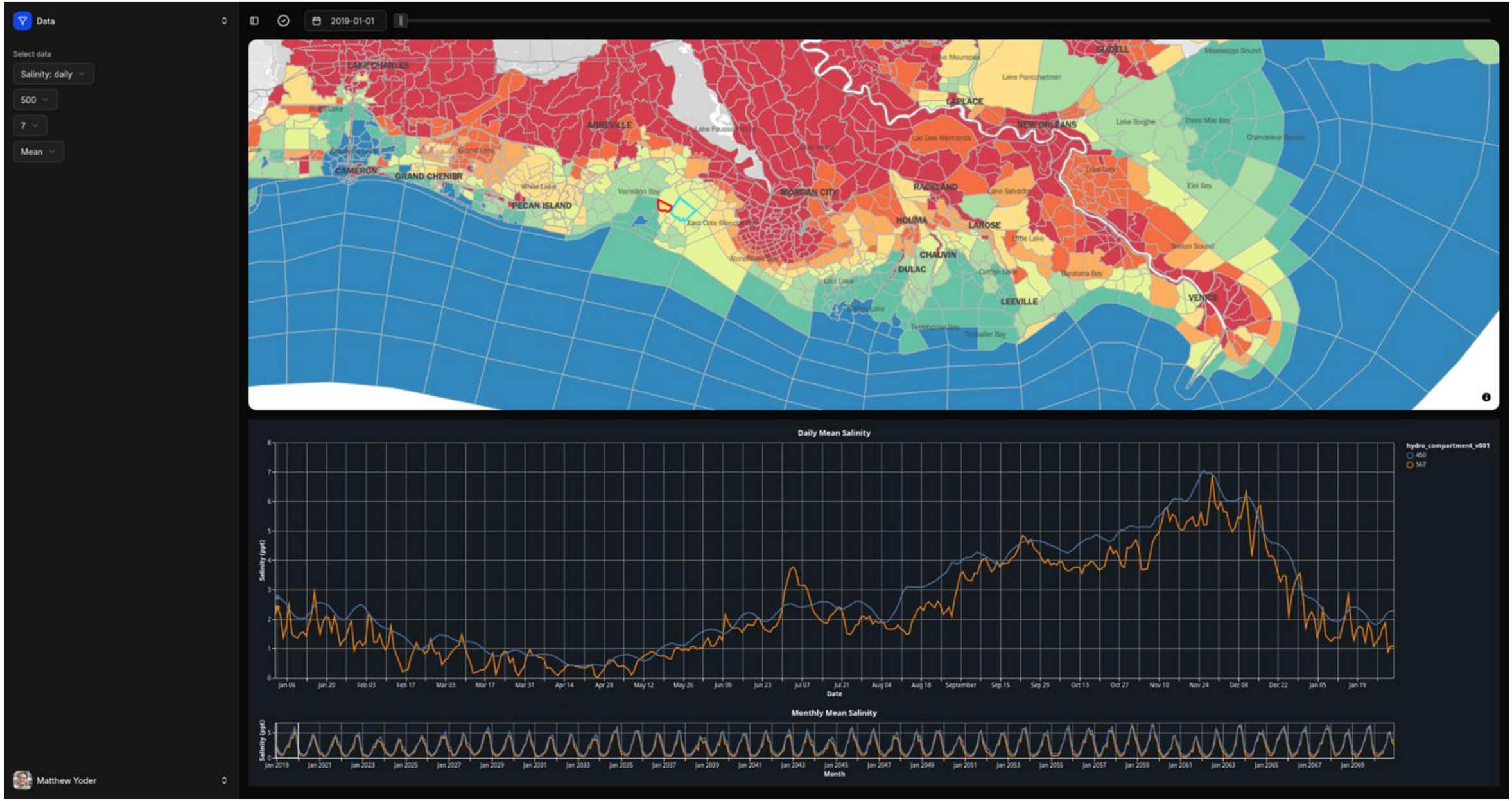
- **Variables tested**
  - Data shape
  - Sorting
  - Compression algorithm
  - Compression level
  - Row group size
- **Metrics**
  - File size on disk
  - Write speed
  - Read speed
    - Entire dataset
    - Sample queries

Row Group Size	Algorithm	Compression Level	File Size (MB)	Read 1 Date All Geos (s)	Read All Dates 1 Geo (s)	Read All Dates 10 Geos (s)	Read Whole File (s)	Write (s)
100,000	brotli	2	179.30	0.86	0.03	0.15	1.27	4.97
100,000	brotli	4	174.06	0.82	0.02	0.15	1.19	5.72
100,000	brotli	6	160.22	0.74	0.03	0.14	1.18	8.84
100,000	gzip	4	165.58	1.20	0.03	0.17	1.80	29.48
100,000	gzip	6	162.80	1.04	0.03	0.17	1.43	16.93
100,000	gzip	8	165.12	0.89	0.03	0.15	1.31	12.35
100,000	lz4	4	189.24	0.54	0.05	0.18	0.56	7.95
100,000	lz4	6	188.70	0.50	0.03	0.18	0.62	8.35
100,000	lz4	8	188.69	0.42	0.02	0.14	0.68	7.96
100,000	lz4	10	188.68	0.41	0.02	0.13	0.64	11.24
100,000	lz4	12	188.67	0.40	0.03	0.16	0.64	12.15
100,000	zstd	3	166.17	0.48	0.03	0.15	0.92	4.13
100,000	zstd	6	177.04	0.46	0.02	0.14	0.88	4.75
100,000	zstd	9	162.88	0.47	0.03	0.14	0.83	6.68
100,000	zstd	12	162.45	0.53	0.03	0.11	0.65	9.15
100,000	zstd	15	157.00	0.60	0.02	0.11	0.72	17.52
100,000	zstd	18	157.04	0.59	0.02	0.11	0.79	31.31
500,000	brotli	2	101.32	0.68	0.02	0.15	0.98	3.96
500,000	brotli	4	92.27	0.69	0.02	0.16	1.06	5.25
500,000	brotli	6	85.72	0.65	0.02	0.24	0.96	7.74
500,000	gzip	4	130.43	0.95	0.02	0.20	1.25	23.52
500,000	gzip	6	128.40	0.81	0.02	0.17	1.14	12.99
500,000	gzip	8	130.29	0.72	0.02	0.15	1.08	9.73
500,000	lz4	4	142.77	0.37	0.02	0.12	0.68	6.54
500,000	lz4	6	142.19	0.35	0.02	0.11	0.63	6.89
500,000	lz4	8	142.17	0.35	0.01	0.11	0.67	7.05
500,000	lz4	10	142.15	0.37	0.01	0.12	0.68	9.65
500,000	lz4	12	142.14	0.35	0.01	0.11	0.67	10.61
500,000	zstd	3	93.41	0.38	0.01	0.09	0.73	3.54
500,000	zstd	6	92.33	0.37	0.01	0.09	0.70	4.26
500,000	zstd	9	89.20	0.37	0.01	0.09	0.66	5.62
500,000	zstd	12	88.80	0.37	0.01	0.10	0.59	7.32
500,000	zstd	15	88.82	0.38	0.01	0.09	0.57	10.06
500,000	zstd	18	85.35	0.43	0.01	0.10	0.65	23.30
1,000,000	brotli	2	94.44	0.66	0.02	0.23	0.96	3.71
1,000,000	brotli	4	86.97	0.68	0.02	0.24	0.97	5.17
1,000,000	brotli	6	80.02	0.65	0.02	0.22	0.98	7.88
1,000,000	gzip	4	124.46	0.94	0.03	0.32	1.32	22.60
1,000,000	gzip	6	122.63	0.80	0.02	0.26	1.10	12.53
1,000,000	gzip	8	124.46	0.73	0.03	0.25	1.04	9.45
1,000,000	lz4	4	136.25	0.37	0.02	0.14	0.65	6.39
1,000,000	lz4	6	135.60	0.34	0.01	0.13	0.64	6.93
1,000,000	lz4	8	135.57	0.35	0.01	0.13	0.62	6.98
1,000,000	lz4	10	135.56	0.34	0.01	0.13	0.63	9.94
1,000,000	lz4	12	135.53	0.34	0.01	0.13	0.61	11.28
1,000,000	zstd	3	87.84	0.39	0.02	0.15	0.71	4.36
1,000,000	zstd	6	85.16	0.38	0.02	0.14	0.72	4.30
1,000,000	zstd	9	83.24	0.40	0.01	0.14	0.70	5.42
1,000,000	zstd	12	82.64	0.41	0.02	0.15	0.58	8.71
1,000,000	zstd	15	82.67	0.37	0.02	0.16	0.58	10.72
1,000,000	zstd	18	79.11	0.40	0.02	0.16	0.64	24.80

INITIAL RESULTS

- **Master Plan 2023 data for 20 variables converted so far**
- **Compression ratios vary widely by dataset**
  - Usually at least 2x compression compared to the original
  - Higher ratios compared to PostgreSQL (e.g., 47GB -> 174MB)
- **API response times are good**
  - Under 500 milliseconds for small grids
  - 1-2 seconds for large grids
  - Supports real-time visualization

2029 PORTAL PROGRESS



NEXT STEPS

- **Issue tracking and commenting** (in progress)
- **Simulation dashboard**
- **Analysis capabilities**
  - Difference maps
  - Custom charts
- **Data export for other visualization tools**

OPEN DATA ECOSYSTEM

- Vision: Create a data ecosystem based on open formats and tools
- Store data once; use it anywhere
  - Master plan data portal
  - Maps and charts embedded in the plan website
  - Jupyter notebooks for data analysis and visualization
  - Custom dashboards
  - Web-based storytelling tools
  - And more!

QUESTIONS