Environmental Data Science Standards For the Offshore Wind Development Industry

And why this is critical

Dave Steckler
Founder / CTO - Mysticetus, LLC

Who am I?

Mathematician, Data Scientist, and Software Engineer

Goals: Invent technologies to solve targeted problems for industry and society

Previously:

Co-Founder - **Nobeltec** - ***How do we safely navigate ships around the world's oceans?***

Microsoft Research - ***How do we organize, analyze, and make use of insanely huge (and growing) quantities of geo-spatial data?***

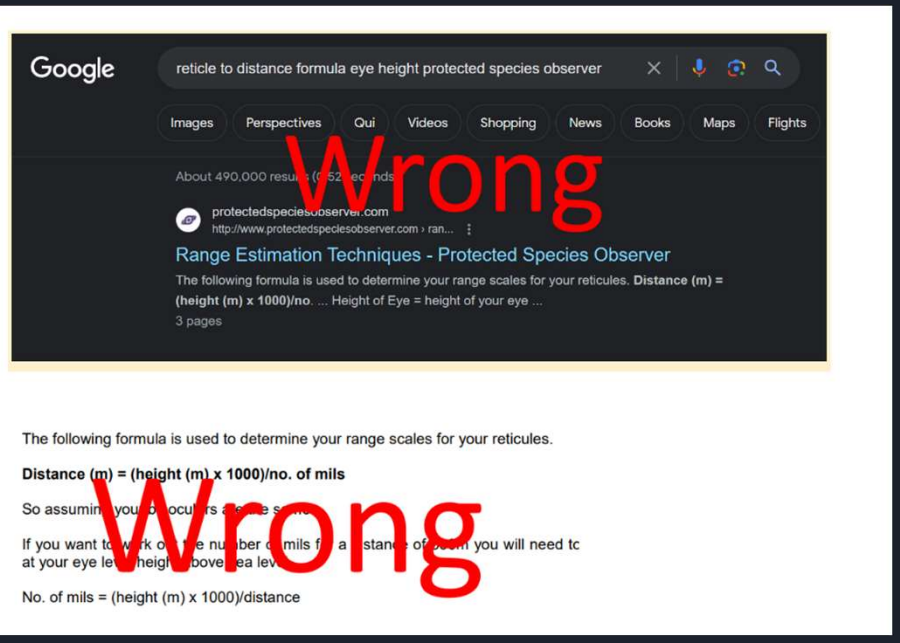
Currently: Founder and CTO - **Mysticetus**

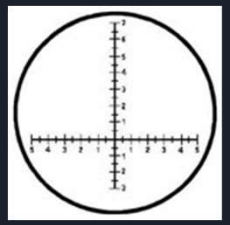
How do we - with data science rigor - collect, audit, and analyze data to SUPPORT and PROVE compliance with environmental law in the face of legal or regulatory challenges?

Early 2010's Protected Species Mitigation Data Science Problems to Solve

Wrong Animal Localization	Wrong Localization Timing	Wrong Effort Times	Wrong Mitigation "Story"	Wrong. Just Wrong.
Wrong formula - distance 30%+ too far	Times recorded to the minute (min error bar ~300m)	Times recorded to the minute	Times recorded to the minute	Witness problem
Made up data "kindof" matches	Times from memory (max error bar unlimited)	Times from memory	Story is from memory	Times and story is from memory
Time only recorded to the minute (min error bar ~300m)		Made up data in attempt to make it "kindof" match	Made up data in attempt to make it "kindof" match	Made up data in attempt to make it "kindof" match

In the early 2010's - Many PSOs were using the wrong formula to calculate distance to animal - too far by 30% or more





1. ***Lerczak, James & Hobbs, Roderick. (2006)**. Calculating sighting distances from angular readings during shipboard, aerial, and shore-based marine mammal surveys. Marine Mammal Science. 14. 590 - 598. 10.1111/j.1748-7692.1998.tb00745.x. Note that the original had an error in a key equation; it was fixed in an errata published in the next issue. I have corrected the equation in this copy but the errata should be cited with the paper.*

2. **Conversion Factors for Reticle Binoculars**, Kinzey and Gerrodette, Marine Mammal Science 2001 – empirically determined that binoculars typically used for marine mammal work (e.g. marine-specific Fujinon, Bushnell, Nikon, etc.) are not produced with optics of mils exactly at 1/5th a degree. Instead, the mean for most 7x50 binoculars has a 0.2779 reticle-to-degree factor.

   Official Citation:

   Kinzey, Douglas, and Tim Gerrodette. "Conversion factors for binocular reticles." Marine Mammal Science 17, no. 2 (2001): 353-361.

3. Additional work by Jeff Laake (NOAA) working with the National Marine Mammal Laboratories - (N)MML, part of Alaska National Marine Fisheries Service, and Calambokidis of the Cascadia Collective (https:/

   Laake and Calambokidis performed precise in-fi Gerrodette, and also selected 0.2779 as a mear binoculars. Of note, **this is the value NMFS us surveys**.

4. Much of these works were usefully summarized, (Texas A&M, Galveston) in 2000 (and beyond) v

Right

Pythagorean uses a modification of the approximation proposed by Lerczak and Hobbs (1998) to determine the distance along the surface of the ocean from the station to the object being fixed:

$$\beta = \frac{\pi}{2} - \alpha - \theta = 180 - \varpi$$

$$D_h = (R_E + h)\cdot\cos(\beta) - \sqrt{(R_E + h)^2\cdot\cos(\beta)^2 - (2hR_E + h^2)}$$

$$\delta = \arcsin\left(\sin(\beta)\frac{D_h}{R_E}\right)$$

$$D = \delta \cdot R_E$$

Once the distance to an object along the surface of the ocean (D) is known, the great circumference equation is used to determine geographic position of the fixed object given the location of the station and the azimuth and distance to the subject:

$$\tau = \eta - \rho$$

$$Lat_F = \sin^{-1}\left(\cos(\tau)\cdot\sin(D/60/1852)\cdot\cos(Lat_S) + \left[\sin(Lat_S)\cdot\cos(D/60/1852)\right]\right)$$

$$Lon_F = \cos^{-1}\left(\frac{\cos(D/60/1852) - \left[\sin(Lat_S)\cdot\sin(Lat_F)\right]}{\cos(Lat_S)\cdot\cos(Lat_F)}\right) + Lon_S$$

where

D = distance (m) between the two points along the surface of the Earth.

τ = bearing from station to subject.

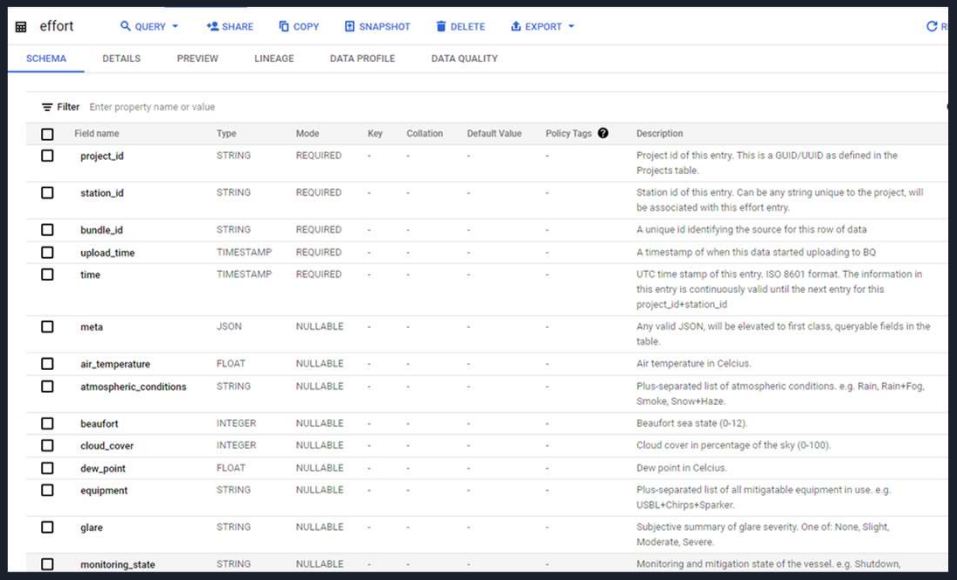
η = azimuth or horizontal angle estimated with the theodolite

ρ = reference azimuth (bearing from station to reference point)

We need your help

Data Collection, Analysis, Auditing, and Sharing Only Works When We ALL Agree On

Common Data Format and Schema



effort

QUERY | SHARE | COPY | SNAPSHOT | DELETE | EXPORT

SCHEMA | DETAILS | PREVIEW | LINEAGE | DATA PROFILE | DATA QUALITY

Filter: Enter property name or value

Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
project_id	STRING	REQUIRED	-	-	-	-	Project id of this entry. This is a GUID/UUID as defined in the Projects table.
station_id	STRING	REQUIRED	-	-	-	-	Station id of this entry. Can be any string unique to the project, will be associated with this effort entry.
bundle_id	STRING	REQUIRED	-	-	-	-	A unique id identifying the source for this row of data
upload_time	TIMESTAMP	REQUIRED	-	-	-	-	A timestamp of when this data started uploading to BQ
time	TIMESTAMP	REQUIRED	-	-	-	-	UTC time stamp of this entry. ISO 8601 format. The information in this entry is continuously valid until the next entry for this project_id+station_id
meta	JSON	NULLABLE	-	-	-	-	Any valid JSON, will be elevated to first class, queryable fields in the table.
air_temperature	FLOAT	NULLABLE	-	-	-	-	Air temperature in Celcius.
atmospheric_conditions	STRING	NULLABLE	-	-	-	-	Plus-separated list of atmospheric conditions. e.g. Rain, Rain+Fog, Smoke, Snow+Haze.
beaufort	INTEGER	NULLABLE	-	-	-	-	Beaufort sea state (0-12).
cloud_cover	INTEGER	NULLABLE	-	-	-	-	Cloud cover in percentage of the sky (0-100).
dew_point	FLOAT	NULLABLE	-	-	-	-	Dew point in Celcius.
equipment	STRING	NULLABLE	-	-	-	-	Plus-separated list of all mitigatable equipment in use. e.g. USBL+Chirps+Sparker.
glare	STRING	NULLABLE	-	-	-	-	Subjective summary of glare severity. One of: None, Slight, Moderate, Severe.
monitoring_state	STRING	NULLABLE	-	-	-	-	Monitoring and mitigation state of the vessel. e.g. Shutdown,

Data Format - We need to pick one (Pro-tip: pick GeoJSON)

Microsoft Access

```
:10001B00204100002028000020000000000000000C
:10002B0000000000000000000000000000000000C5
:10003B0000000000000000000000026401016000A00E3
:10004B0020B00A0020700A0020600A0020B00A00CD
:10005B00208100002068000020000000000000004C
:10006B0000000000000600A00200000000000000FB
:10007B0000000000000000000020802021000A004D
:10008B0020600A0020100A0020100A0020100A002D
:10009B0020C1000020A8000020000000000000008C
:1000AB0000000000000000000000000000000045
:1000BB0000000000000000000040000204400000CD
:1000CB00208400002000000000000000000000061
:1000DB0000000000000000000FFFFFFF01000018
:1000EB0000000000000000000000000000000005
:1000FB0000000000000000000000000FFFFFFF8
:10010B00FFFFFFFFFF0000000000000000000E9
:10011B0000000000000000000000000000000D4
:10012B0000000000000000000FFFFFFFFFFF00CA
```

Shapefile
GeoDB
Raw Text
RTF
…

CSV

Time (UTC)	Night	PSO 1	PSO 2	Effort Type 1	Effort Type 2	PSO Locat	PSO Locat	DMA Stat	Vessel Spe	Vessel Activity	Monitorin	Source Fre	GT Operat	Borehole I	Co
2024-01-26T00:00:00.0	TRUE	Meghan	N/A	NVD	Off	Bridge	N/A	TRUE	0.1	Surveying	Monitorin	None	FALSE	FALSE	F
2024-01-26T00:30:00.0	TRUE	Meghan	N/A	NVD	Off	Bridge	N/A	TRUE	0.1	Surveying	Monitorin	None	FALSE	FALSE	F
2024-01-26T01:00:00.0	TRUE	Nicole H	N/A	NVD	Off	Bridge	N/A	TRUE	0.1	Surveying	Monitorin	None	FALSE	FALSE	F
2024-01-26T01:30:58.0	TRUE	Nicole H	N/A	NVD	Off	Bridge	N/A	FALSE	0.1	Surveying	Monitorin	None	FALSE	FALSE	F
2024-01-26T02:00:24.5	TRUE	Nicole H	N/A	NVD	Off	Bridge	N/A	FALSE	0.1	Surveying	Monitorin	None	FALSE	FALSE	F
2024-01-26T02:30:18.5	TRUE	Nicole H	N/A	NVD	Off	Bridge	N/A	FALSE	0.2	Surveying	Monitorin	None	FALSE	FALSE	F
2024-01-26T03:00:23.1	TRUE	Nicole H	N/A	NVD	Off	Bridge	N/A	FALSE	0.5	Surveying	Monitorin	None	FALSE	FALSE	F
2024-01-26T03:30:31.8	TRUE	Nicole H	N/A	NVD	Off	Bridge	N/A	FALSE	0.4	Surveying	Monitorin	None	FALSE	FALSE	F
2024-01-26T04:00:58.8	TRUE	Nicole H	N/A	NVD	Off	Bridge	N/A	FALSE	0.2	Surveying	Monitorin	None	FALSE	FALSE	F
2024-01-26T04:30:00.0	TRUE	Nicole H	N/A	NVD	Off	Bridge	N/A	FALSE	0	Surveying	Monitorin	None	FALSE	FALSE	F

GeoJSON

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "geometry": {
        "type": "Point",
        "coordinates": [102.0, 0.5]
      },
      "properties": {
        "prop0": "value0"
      }
    },
    {
      "type": "Feature",
      "geometry": {
        "type": "LineString",
        "coordinates": [
          [102.0, 0.0],
          [103.0, 1.0],
          [104.0, 0.0],
          [105.0, 1.0]
        ]
      },
      "properties": {
        "prop0": "value0",
        "prop1": 0.0
      }
    },
    {
      "type": "Feature",
      "geometry": {
        "type": "Polygon",
```

Google Earth KML (XML)

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
  <Document>
    <Placemark>
      <name>CDATA example</name>
      <description>
        <![CDATA[
          <h1>CDATA Tags are useful!</h1>
          <p><font color="red">Text is <i>more readable</i> and
          <b>easier to write</b> when you can avoid using entity
          references.</font></p>
        ]]>
      </description>
      <Point>
        <coordinates>102.595626,14.996729</coordinates>
      </Point>
    </Placemark>
  </Document>
</kml>
```

What's a Data Schema?

A formal database ***schema*** is far more than just field names

1. *Field Name*

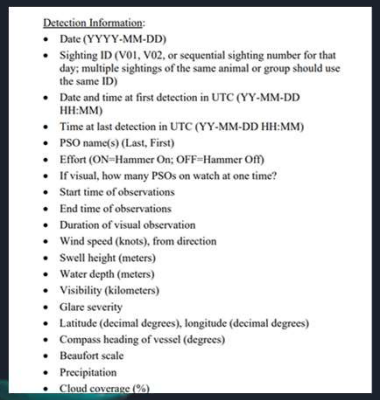
1. *Formal Representation Format*

1. *Validation Rules and Warnings*

2. *Query Semantics*

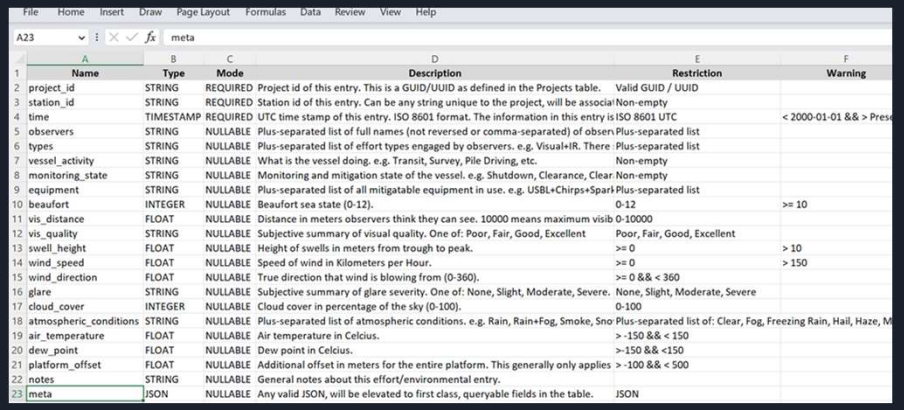
3. *Adheres to Data Science Best Practices*

Lists of required data fields is a good start, but not good enough



Detection Information:

- Date (YYYY-MM-DD)
- Sighting ID (V01, V02, or sequential sighting number for that day; multiple sightings of the same animal or group should use the same ID)
- Date and time at first detection in UTC (YY-MM-DD HH:MM)
- Time at last detection in UTC (YY-MM-DD HH:MM)
- PSO name(s) (Last, First)
- Effort (ON=Hammer On; OFF=Hammer Off)
- If visual, how many PSOs on watch at one time?
- Start time of observations
- End time of observations
- Duration of visual observation
- Wind speed (knots), from direction
- Swell height (meters)
- Water depth (meters)
- Visibility (kilometers)
- Glare severity
- Latitude (decimal degrees), longitude (decimal degrees)
- Compass heading of vessel (degrees)
- Beaufort scale
- Precipitation
- Cloud coverage (%)



File Home Insert Draw Page Layout Formulas Data Review View Help

A23 meta

	A	B	C	D	E	F
1	Name	Type	Mode	Description	Restriction	Warning
2	project_id	STRING	REQUIRED	Project id of this entry. This is a GUID/UUID as defined in the Projects table.	Valid GUID / UUID	
3	station_id	STRING	REQUIRED	Station id of this entry. Can be any string unique to the project, will be associa	Non-empty	
4	time	TIMESTAMP	REQUIRED	UTC time stamp of this entry. ISO 8601 format. The information in this entry is	ISO 8601 UTC	< 2000-01-01 && > Pres
5	observers	STRING	NULLABLE	Plus-separated list of full names (not reversed or comma-separated) of obsen	Plus-separated list	
6	types	STRING	NULLABLE	Plus-separated list of effort types engaged by observers. e.g. Visual+IR. There	Plus-separated list	
7	vessel_activity	STRING	NULLABLE	What is the vessel doing. e.g. Transit, Survey, Pile Driving, etc.	Non-empty	
8	monitoring_state	STRING	NULLABLE	Monitoring and mitigation state of the vessel. e.g. Shutdown, Clearance, Clear	Non-empty	
9	equipment	STRING	NULLABLE	Plus-separated list of all mitigatable equipment in use. e.g. USBL+Chirps+Spar	Plus-separated list	
10	beaufort	INTEGER	NULLABLE	Beaufort sea state (0-12).	0-12	>= 10
11	vis_distance	FLOAT	NULLABLE	Distance in meters observers think they can see. 10000 means maximum visib	0-10000	
12	vis_quality	STRING	NULLABLE	Subjective summary of visual quality. One of: Poor, Fair, Good, Excellent	Poor, Fair, Good, Excellent	
13	swell_height	FLOAT	NULLABLE	Height of swells in meters from trough to peak.	>= 0	> 10
14	wind_speed	FLOAT	NULLABLE	Speed of wind in Kilometers per Hour.	>= 0	> 150
15	wind_direction	FLOAT	NULLABLE	True direction that wind is blowing from (0-360).	>= 0 && < 360	
16	glare	STRING	NULLABLE	Subjective summary of glare severity. One of: None, Slight, Moderate, Severe.	None, Slight, Moderate, Severe	
17	cloud_cover	INTEGER	NULLABLE	Cloud cover in percentage of the sky (0-100).	0-100	
18	atmospheric_conditions	STRING	NULLABLE	Plus-separated list of atmospheric conditions. e.g. Rain, Rain+Fog, Smoke, Sno	Plus-separated list of: Clear, Fog, Freezing Rain, Hail, Haze, M	
19	air_temperature	FLOAT	NULLABLE	Air temperature in Celcius.	> -150 && < 150	
20	dew_point	FLOAT	NULLABLE	Dew point in Celcius.	>-150 && <150	
21	platform_offset	FLOAT	NULLABLE	Additional offset in meters for the entire platform. This generally only applies	> -100 && < 500	
22	notes	STRING	NULLABLE	General notes about this effort/environmental entry.		
23	meta	JSON	NULLABLE	Any valid JSON, will be elevated to first class, queryable fields in the table.	JSON	

Set Aside Time, Focus, (and Money) to Collaborate

Need industry-wide agreement on Mitigation (PSO/PAM/Video) Data:

1. (Data) Science Best Practices
2. Schemas and Formats
3. Sharing and Transport Mechanisms
4. Query & Analysis Methods and Modalities
5. Incorporate Indigenous/Traditional Knowledge Systems

Thanks!

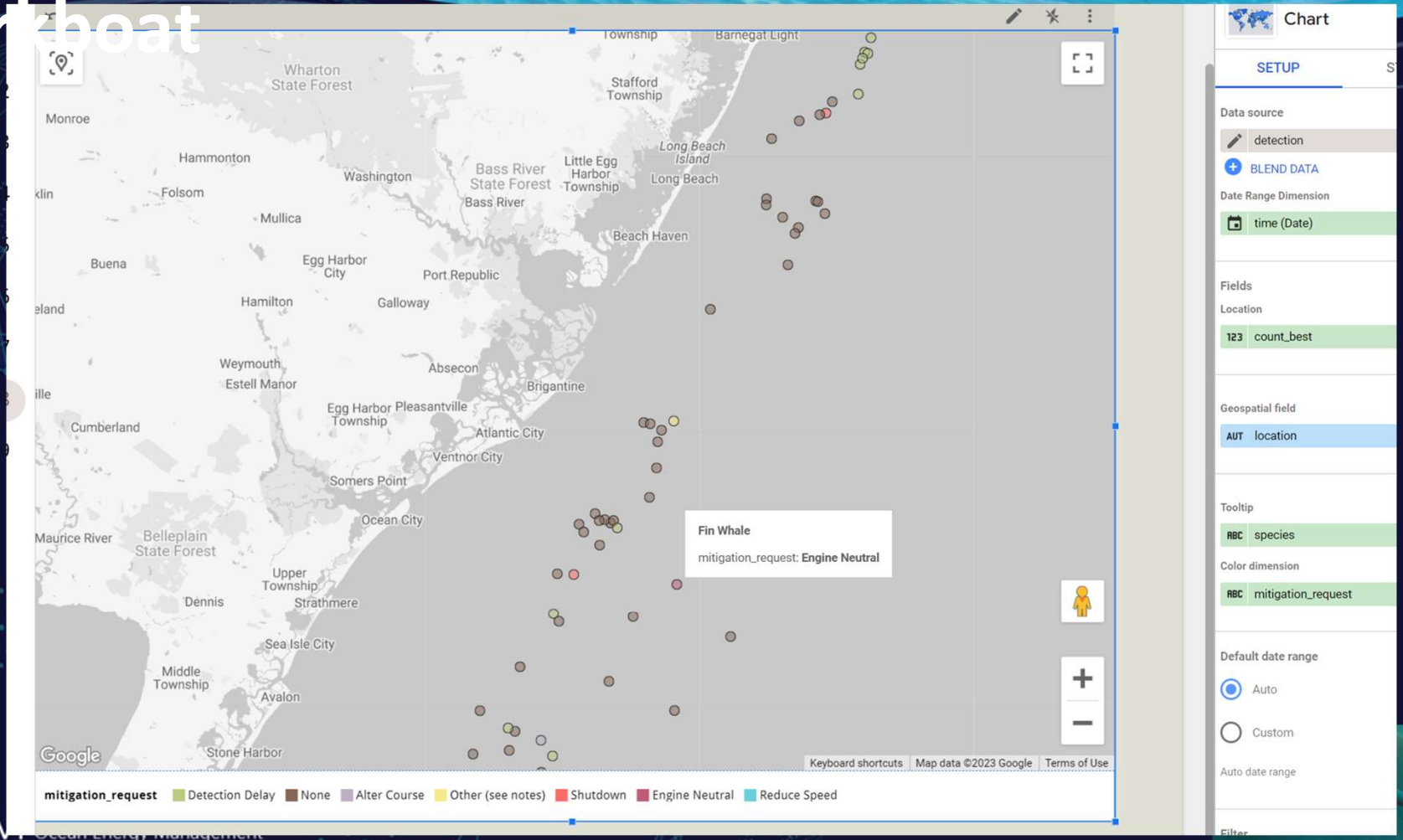
Extra Slides

Geospatial Data Science Best Practices

For Protected Species Scientists, Government Regulators, and Industry Managers

David Steckler - 2023-02-05

2 fin whales popped up in front of moving workboat



Watch a PSO prevent a possible vessel strike

```
SELECT * FROM `mysticetus-boem.main.detection`
WHERE project_id = "9FCEF5D2-7350-42A7-BAFA-CD56DD4B6B88"
  AND species = "Fin Whale"
  AND mitigation_request = "Engine Neutral"
  ORDER BY time
  LIMIT 1000
```

This query finds the appropriate raw sighting record in the detection table

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

row	project_id	station_id	id	time	species	distance_estimation_technique
1	9FCEF5D2-7350-42A7-BAFA-C...	OceanObserverVis	V77	2021-11-09 18:35:00 UTC	Fin Whale	UE

speed_over_ground
7.512334

A similar query is used to find the corresponding effort record based on timestamp

effort_timestamp	effort_station_id	ef	monitoring_state	vessel_activity
2021-11-09 18:15:26.600000 U...	OceanObserverVis		Monitoring - Transit	Transit

Query results

SAVE RESULTS | EXP

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

mitigation_request	mitigation_response	observer_name	dead	injured	notes	species_notes	distance	location	meta
ngine Neutral	Engine Neutral	Mauricio Guimaraes	fal...	fal...	Two animals were sighted at 18:35 UTC 20m from the vessel`s bow (50m from PSO) blowing at the surface. Officer on watch, put the engines in neutral and altered course to port away from whales	Dark brown body, tall bushy blow, very small dorsal fin visible relative to large body size. Long dark back, with dorsal fin positioned	50.0	POINT(-74.2090179280977 39....	null

BOEM Bureau of Ocean Energy Management

Mysticetus
mysticetus.com

Fix Fin Whale ~600m Location Error (1/2)

id	time	species
V77	2021-11-09 18:35:00 UTC	Fin Whale

1. 1 in 60 events are probabilistically at the top of a minute - ":00" seconds merits a deeper look

1. Why: In critical situations (*2 fin whales pop up 20m in front of moving boat*) PSOs are trained to FIRST protect animal(s), THEN go back and retroactively enter data

2. Human recollection of exactly when something happened can be coarse ("a few minutes ago")

3. At 7.5 kts, a boat travels ~233 meters / minute (which is why **recording to the second is critical**)

1. Sighting locations are tied to a boat's location by exact time stamp (to the second)

2. **If time of sighting is off by ~2.5 minutes, sighting location is wrong by ~600 meters**

This hinders sighting sharing (mitigation on other vessels) as well as future scientific study

Even deeper - Fix Fin Whale ~600m Location Error (2/2)

PSO Collection of high-fidelity GPS tracks pays huge quality dividends

```
SELECT * FROM mysticetus-boem.main.tracks
WHERE project_id = "9FCEF5D2-7350-42A7-BAFA-CD56DD4B6B88"
AND time BETWEEN TIMESTAMP("2021-11-09T18:30:00") AND TIMESTAMP("2021-11-09T18:40:00")
```

Query for high fidelity GPS track marks near reported time of sighting of Fin Whales

Press Alt+F1 for Accessibility Options.

ery results — SAVE RESULTS — EXPLORE DATA

INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Speed is kilometers per hour - multiply by 0.54 to get knots

station_id	time	fix_status	altitude	speed_over_gro	course_over_gro	alt_offset	location
OceanObserverVis	2021-11-09 18:32:20.673900 U...	Fixed	10.7	14.79748	126.64	NaN	POINT(-74.21514 39.21774166...
OceanObserverVis	2021-11-09 18:32:25.675190 U...	Fixed	10.8	14.79748	134.69	NaN	POINT(-74.21496 39.21762)
OceanObserverVis	2021-11-09 18:32:30.690750 U...	Fixed Engine Neutral - Rapid Slowing	9.6	14.55672	126.72	NaN	POINT(-74.214785 39.2175033...
OceanObserverVis	2021-11-09 18:32:35.697320 U...	Fixed Sharp turn to STARBOARD	8.3	13.89	140.48	NaN	POINT(-74.214605 39.2173733...
OceanObserverVis	2021-11-09 18:32:40.710300 U...	Fixed	8.2	15.20492	125.94	NaN	POINT(-74.2144083333334 39....
OceanObserverVis	2021-11-09 18:32:45.573200 U...	Fixed	7.8	14.46412	131.99	NaN	POINT(-74.2142283333333 39....

This is the only significant speed/course deviation in high fidelity GPS track +-5 minutes of reported sighting time (*also note: ship turned sharply to starboard, not port)*

1. **Reported time of sighting: 18:35:00**
2. **Actual time of sighting ~18:32:30 (~02m 35s earlier)**
3. Original Plotted Position: 39.2135168 N 74.2090179 W
4. Corrected Position: 39.2172304 N 74.2143645 W
5. Error Corrected: **618.61m**

BOEM Bureau of Ocean Energy Management