

Background

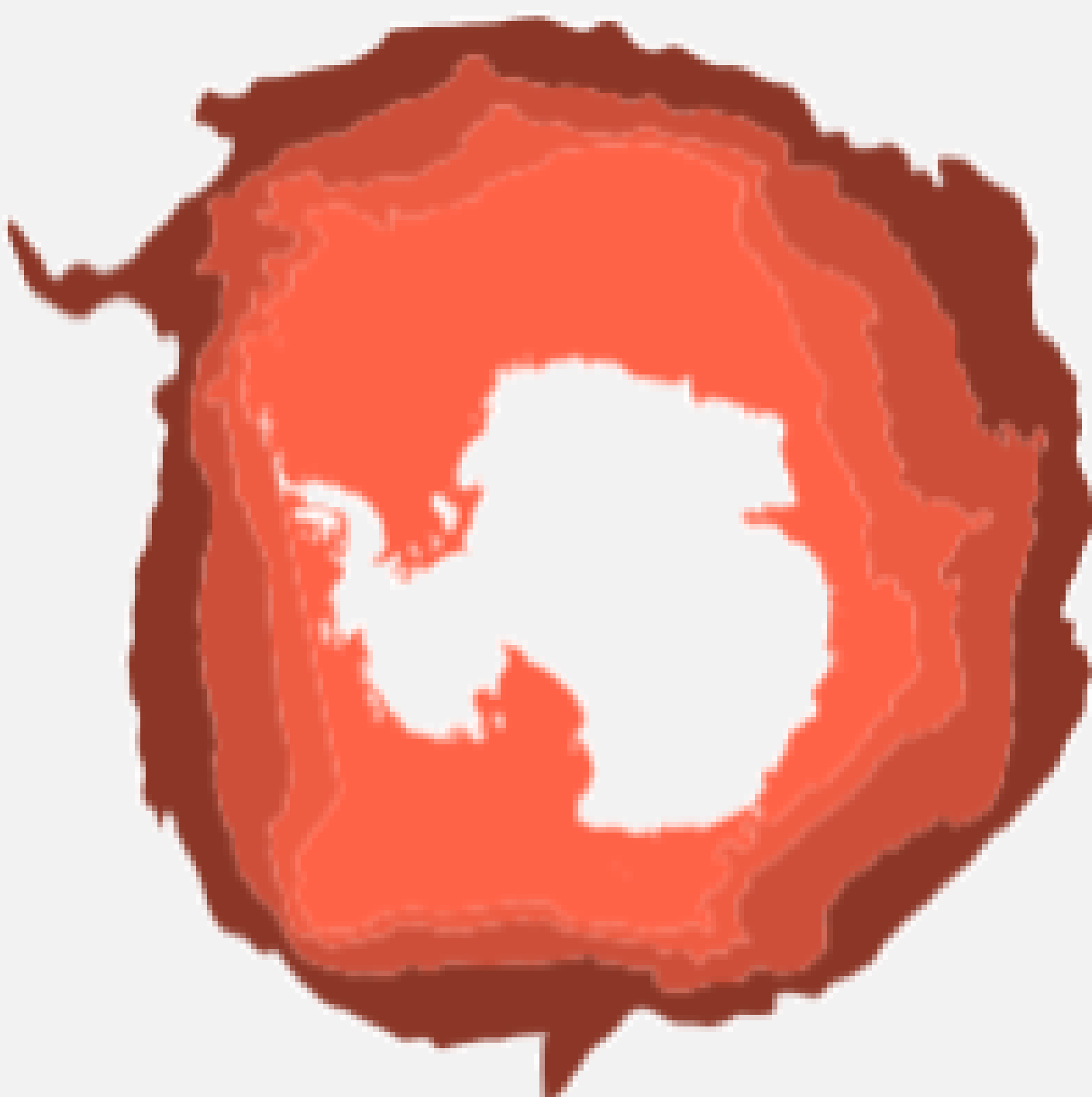
The Southern Ocean is considered a large part of the global carbon cycle and phytoplankton are a key driver by converting CO₂ to carbon in the deep ocean. Current databases examine phytoplankton presence and CO₂ flux but don't take community species composition into account.

We aim to determine the relationship between phytoplankton species and biological production

Methods

I added a frontal zone label to all the data by reading in a fixed front definition (Kim and Orsi, 2014), converting to polygon shapefiles in python, and performing a geospatial join.

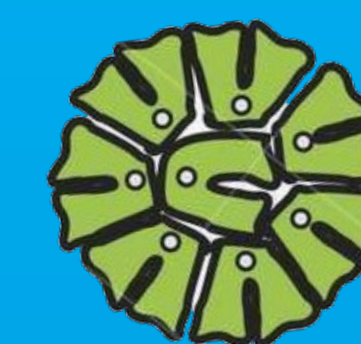
Researchers commonly filter data by the frontal zone and sector, so the join is performed at insert time (rather than runtime) which greatly increases query speed for the user



SOPhy

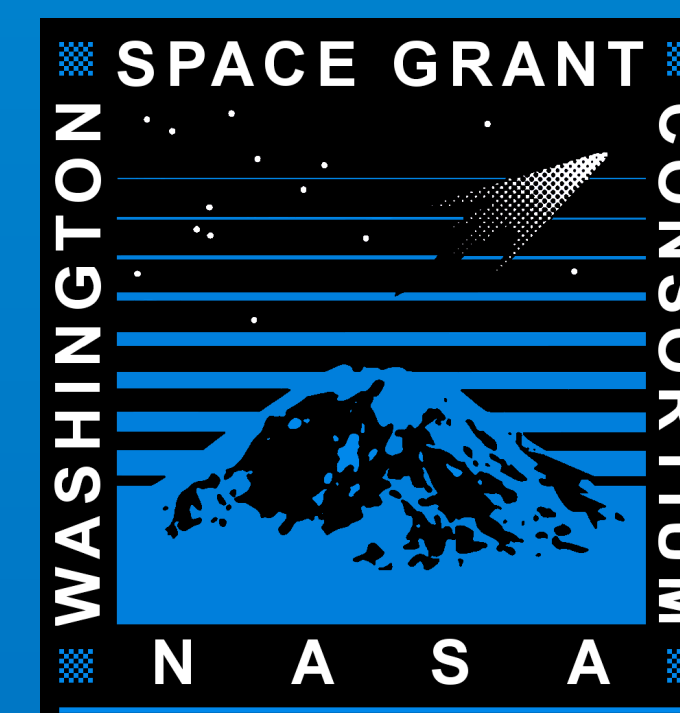
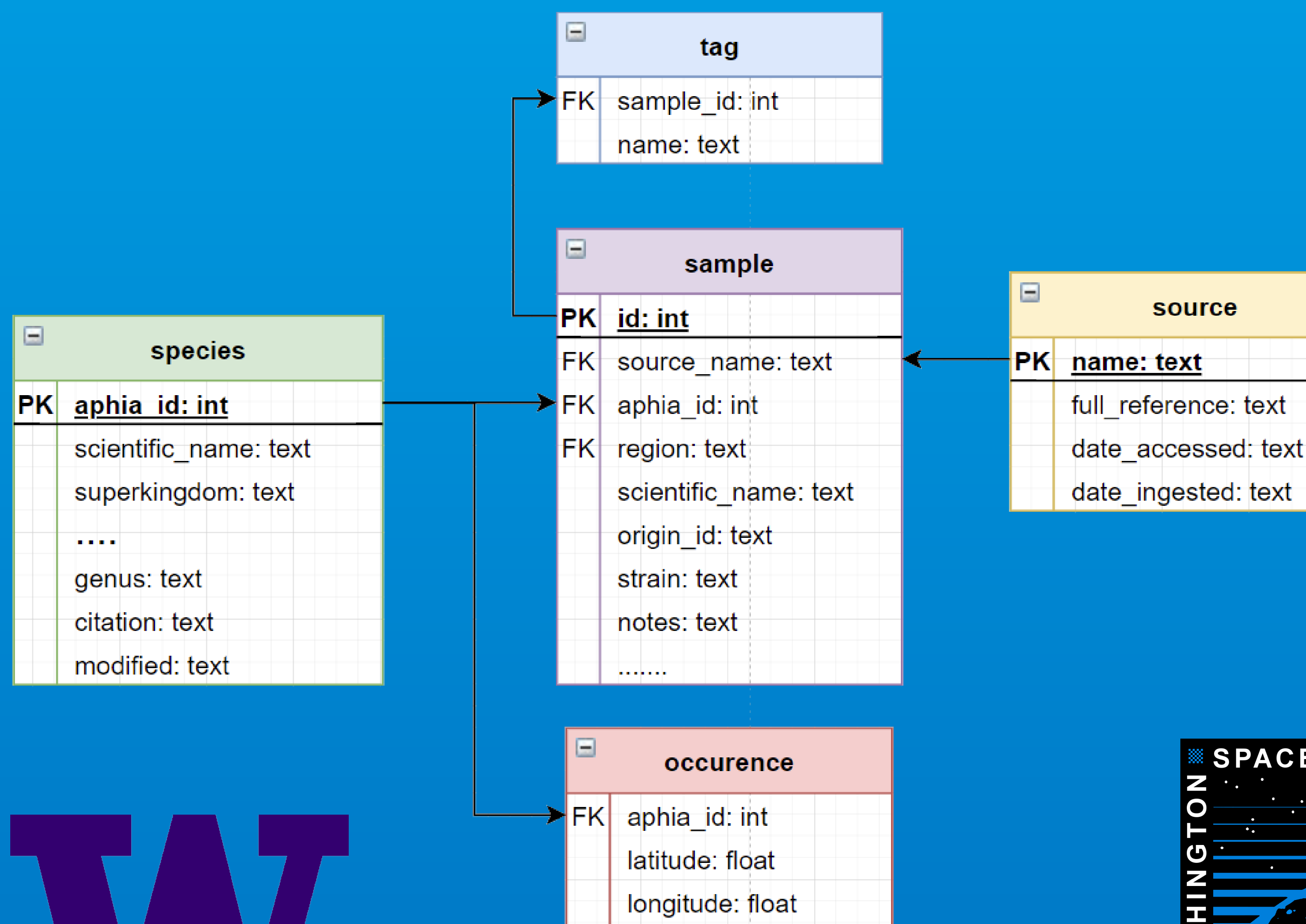
Southern Ocean Phytoplankton Database

Ayush Nag¹, Hannah Joy-Warren², David A.C. Beck¹



Combining in situ datasets from cruises and adding features such as full taxonomy and labeled Southern Ocean fronts and sectors.

Database schema and backend in SQLite. Used Python to transform the data and add new features such as visualizations, full taxonomy records, and ease of use functions.



Schema

The first step of this project was to design the schema and normalize the dataset. This is the heart of the database and an evolving document on version 8.

Below is a sample query that counts the total amount of genera across the data. Genus is not present in the main table however the data can be joined using the aphia_id key which represents the exact species of phytoplankton

```
select count(*) as entries, genus
from sample, microscopy
where sample.aphia_id = microscopy.aphia_id
group by genus
order by entries desc;
```

entries	genus
4491	Chaetoceros
2009	Fragilariopsis
1937	Tripos
1893	Corethron

Future Work

Next, we will use this data to train a machine learning model to help us study the relationship between phytoplankton species and biological production. We will get data from Argo and BGC-Argo floats to create clusters. We will study the correlation between the clusters and community composition

1. Paul G. Allen School of Computer Science and Engineering, University of Washington
2. Cooperative Institute of Climate, Ocean, and Ecosystem Studies, University of Washington