Updated Data Management Plan

Overall statement of data accessibility

Our team and global data collaborators strongly support policies of data openness and transparency along with the FAIR principles (Findability, Accessibility, Interoperability, and Reusability). Openness and sharing provides an efficient space where researchers can benefit from earlier efforts to continue the iterative process of learning and gaining knowledge. To that end, all raw and derived data products from this study will be openly shared with the marine science community through standard data access portals.

Data sources and types

This will be a complicated and empirical data-driven research project, with a wide range of data types to be used and integrated, and an even wider variety of derivative data products that will be generated. Data will be obtained and integrated from existing datasets, such as the European Unions' Earth Observation Program Copernicus (www.copernicus.eu) global ocean model reanalysis product (GLORYS) and chlorophyll (chl-a) data from NASA's Ocean Biology Processing Group (OBPG). These datasets will be stored as either NetCDF4 or Zarr formats which can be accessed efficiently using parallel processing tools such as the Dask framework. Temperature at depth within 300 km radii will be differenced to produce *estimates of thermal stratification for each seabird site under study* (estimated to be ~50 sites) at a daily to monthly temporal resolution, which will then be averaged to produce estimates for seabird pre-breeding and breeding seasons. A similar process will be used to produce *monthly to seasonal values of chl-a* (using OBPG data). Daily, monthly, and seasonal (pre-breeding and breeding seasons) values of stratification and chl-a for each colony will be shared.

Seabird productivity and diet data, as well as forage fish abundance data sets, are of much simpler formats since they are not spatial, and can be contained in comma separated value (.csv) files. Seabird breeding productivity data have been provided by collaborators around the world for their specific study sites and species (see Supplemental Materials of Sydeman et al. 2021; currently available on Github (https://github.com/DavidSchoeman/sydeman_et_al_seabirds). For this project, we will update this database with data from other sites not used in our previous analyses. The breeding productivity data is expressed per species in the form of *number of young fledged per female per year*, so these data will be shared in a time series format. Prey use data will be in the form of *proportion of prey species used per year*, and we will use only the top two prey species, as measured on decadal scales, in this study. The prey use data for species and colonies under study is not currently publically available, so this will be a new global dataset on seabird prey use. Other prey availability data are in the form of *biomass (metric tons)* and are available from fisheries stock assessments, many of which will be accessed through the RAM Legacy Stock Assessment Database (ramlegacy.org). Additional forage fish relative abundance data in the form of *catch per unit effort* (CPUE, available from trawl net samples) or *nautical area scattering coefficient* (NASC, available from hydro-acoustic surveys), will be available for certain sites, and we will contact data providers directly for these data sets. These data will also be provided as time series on the spatial scale (typically regional) available from data contributors.

Data storage and availability

Earlier, the large physical and lower trophic level datasets to be used in this study would have required substantial downloading and storage resources, but several of them are already available in the cloud on either AWS S3 or Google Cloud servers. These datasets and their metadata are accessible using the dataset search functionality of either AWS (https://registry.opendata.aws/) or Google (https://datasetsearch.research.google.com/). The cloud availability of these datasets makes them easily accessible to all project members, as well as the marine science community and the public, and in conjunction with cloud computing, now offers a tremendous opportunity for collaborative research on large datasets. For project data that are not currently on the cloud, we have budgeted *1 TB of S3 cloud storage* for making these data available. Moving our analysis to the cloud will allow us to process large amounts of data with little hardware investment, and this approach will also enhance the potential reproducibility of our research to advance our open data/open science research policies.

Global seabird productivity and diet data are provided as annual means by individual data providers or groups; raw data will not be made available to us for this project. Data used in our analyses, however, will be shared when the project concludes. For other data that are already publicly available, we will share the specific data set used in our analyses and the location/source where the original data were obtained. Raw data will be provided when possible.
We will contribute our project metadata and data to the Biological & Chemical Oceanography Data Management Office (BCO-DMO; www.bco-dmo.org/). Derived data from this project, including seabird breeding productivity and prey data, will be made available on Github and BCO-DMO within two years of project completion.

Sharing analytical code
Integration of datasets through GLMM and SEM modeling will be conducted using open-source cloud computing frameworks such as Python, Julia, Jupyterhub, and R. Our default policy for all code and data management software developed as part of this project is that it will be available publicly under an explicit open-source license within two years of project completion. All collaborators participating in the project will be encouraged to use permissive licenses (e.g., MIT) as well to allow for the reuse of code and other data that may be generated from project activities. All software will be stored on a publicly available repository as part of Farallon Institute's Github account (https://github.com/farallon-institute).

Roles and responsibilities
The PI (Sydeman) and co-PIs (García-Reyes and Kristiansen) of the project will be responsible for ensuring that this data management plan is implemented, and that any concerns regarding the plan be handled according to the above open data/open science policy framework. Project staff SA Thompson will be tasked with ensuring data is shared on appropriate time scales. All project data and code will be available within 2 years of project completion. Updates to data management will be made in the annual reports.