Data Management Plan: This proposal will not generate oceanographic observations. The numerical model data used is from the Mercator Ocean project, and must be retrieved from them: https://www.mercator-ocean.fr/en/. The Mercator current data is guaranteed by agreement between the European Union and the Copernicus Marine Environmental Monitoring service to be available until March 2021 (http://marine.copernicus.eu/services-portfolio/service-commitments-and-licence/); they expect it to be available after that, and have a roadmap to run the model and provide the data until 2027 (personal communication with Mercator Ocean Communications). We will not serve this data both for licensing issues and because it is large (1.5 Tb/year for full data sets, 800Gb/year for just currents).

We will generate 4 products that will be available to the community:

1. From activity 1, a library of Lagrangian pathways released from all coastal points in the Mercator model, and all points adjacent to those points, for 60 days. A new path will be generated every quarter-day from 2007 to at least 2020. Particles will be released at every 10 meters from 0 to 200m and will be both fixed to that depth and allowed to drift vertically. This data will be archived at BCO-DMO.
2. From activity 2, life-history data and vertical swimming speeds will be gathered from the literature that supplement the data in Selkoe and Toonen (2011). This will be included as an appendix in the article published from this work.
3. For activity 3, we will create global connectivity matrices for each of the depths of particle release, both for particles confined to a depth and able to move freely vertically, globally, at the Mercator 1/12° resolution, for larval durations of 5 to 60 days, and with a competency period as described in the proposal. These sparse and compressed matrices are modest in size (300Mb per global matrix), and can easily be stored at BCO-DMO.
4. The maps generated in activity 4 are small in size (10s of Mb) and will be stored in a standard data format (netCDF) at BCO-DMO.

Data is not really shared unless tools are provided to use it. The code used in the analysis in activities 1 through 4 and described in the broader impacts of those sections will written in the open-sourced language Python, using standard, open-source libraries. We will also provide wrappers that will allow them to be called from R using the open-source "reticulate" project (https://github.com/rstudio/reticulate). We have used reticulate to allow R knowledgeable scientists to use our python codes, and have found it easy to use, and to show others how to use, since it is a part of the Rstudio project. We will test these interfaces well, since one of the PI's labs is Python based, and the other is more comfortable in R.

The code and the documentation to use them (as described in the proposal text) will be hosted by GitHub, a standard code and documentation repository for open-source projects, and links to the code and documentation will be provided by BCO-DMO. We will release our code under the permissive MIT open source license (https://en.wikipedia.org/wiki/MIT_License).

Instructional videos on the use of the software packages will be served by YouTube, and will be released under the Creative Commons Attribution license. The videos will be linked to by BCO-DMO.

The data analysis and particle tracking codes will require extensive temporary data sets. We have budgeted for archiving and backup infrastructure for these data sets.