International Cartographic Association
Association Cartographique Internationale

Arctic seas data cube based on discrete global grid system

German Titov [a,*], Pavel Kargashin [a]

[a] *Lomonosov Moscow State University, German Titov - titovgs@my.msu.ru, Pavel Kargashin - p.e.kargashin@mail.ru*

* Corresponding author

Keywords: Arctic, data cube, discrete global grid system, H3, analysis ready data

Abstract:

The amount of spatial data is becoming so vast that traditional research methods are proving to be inadequate and heterogeneity turns out to be a more significant methodological problem than the amount of data (Li et al., 2016; Mechenich, Žliobaitė, 2023). Once little-explored the Arctic region moves towards a trend of abundance of data with the variety of open data sources (Olawoyin et al., 2021). Heterogeneous data can be integrated into a multi-dimensional construct known as a data cube (Purss et al., 2019; Shurygina, Titov, 2022).

Data integrated in the cube can be called analysis ready data as it can be utilized in, for example, machine learning, expert analysis, web applications with a minimum of additional user effort. The data cube as a methodology provides a framework to unify heterogeneous spatial data, helps to decouple analysis from preprocessing and minimizes the geographical complexity of the data to analyze. For the Arctic region geographical complexity lies in the context of distortions at high latitudes, various data formats and structures. Alternative approaches such as transforming data to a unified projection, eliminating differences in formats make spatial data easier to use but still require GIS training from the user.

The data cube encompasses spatial and temporal dimensions, which pinpoint the location, and attribute dimensions that encapsulate the properties of the location. The spatial-temporal dimensions comprise the date, the depth, and the index of the cell of the discrete global grid system H3. Using a raster grid is more performant but distortion of the raster cells damages geographical reliability of analysis. Therefore the process of designing a data cube includes determining the optimal cell size of the discrete global grid system, choice of date precision, and selection of area properties. The setting of these parameters is a bargain between the speed of further processing and the detail of the dataset.

Our data cube of the Arctic seas is designed to assess the habitats of marine animals. The data cube is built of hexagonal cells with a side of 4 kilometers and has a date precision of a month. This is due to focus on interactive use with the help of a web map, the detail of the source data and its coverage. The data cube integrates (1) biogeochemical and (2) physical variables from Copernicus Marine Data Store, (3) bathymetry from The General Bathymetric Chart of the Oceans (GEBCO), (4) known ranges of threatened species from The IUCN Red List, (5) animal occurrences from Global Biodiversity Information Facility (GBIF) and Ocean Biodiversity Information System (OBIS), ship load calculated from Automatic Identification System (figure 1).

Raster data (1, 2, 3) transformed using R library `terra`. Vector data (4, 5) transformed with `h3` and `h3ronpy` Python libraries. The data cube is implemented as a PostgreSQL database with PostGIS and h3-pg extensions

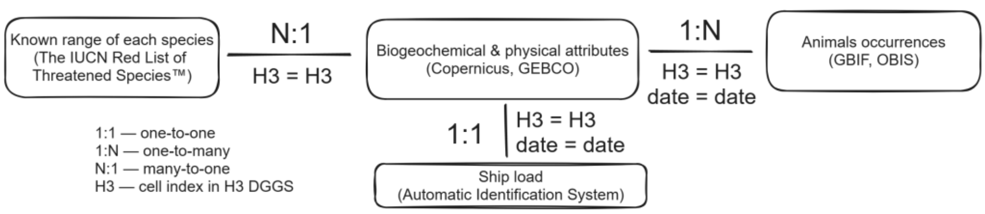

Figure 1. Relations of the data sources in the data cube.

A RESTful API for the data cube is developed with open-source Supabase and PostgREST software. DeckGL and React is utilized to visualise the data cube attributes as a set of web maps and add capacities of interactive filtering. The DeckGL mapping library and related VisGL tools (https://vis.gl) allow to organize interactive data filtering with immediate cartographic visualization, which is convenient for performing an expert assessment of the parameters of marine animal habitats or verifying the results of processing by machine learning tools (figure 2).

https://doi.org/10.5194/ica-abs-7-172-2024

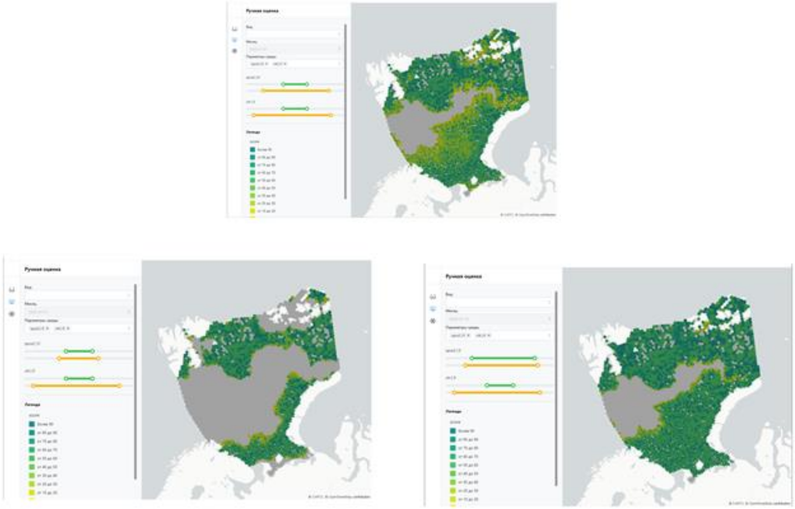

Figure 2. Interactive filtering with immediate cartographic visualization

The preliminary integration and harmonization makes data ready-to-analyze. The suggested data cube approach minimizes distortion problems at high latitudes, facilitates API development and spatial data visualization in the Web. The discrete global grid system as a spatial framework allows to perform effective calculations related to analysis by machine learning methods. Being more geographically reliable discrete global grid systems are less performant and less functional than raster grids for now. Integration of vector and raster data on the framework of a discrete global grid system is not lossless which requires careful attention to the choice of spatial and temporal resolution. Within these constraints the data cube approach can also be used for other cases where it is feasible to bring data to a unified regular spatial grid and beneficial to visualize it on a web map.

References

Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., Cheng, T., 2016. Geospatial big data handling theory and methods: A review and research challenges. ISPRS Journal of Photogrammetry and Remote Sensing 115, 119–133. https://doi.org/10.1016/j.isprsjprs.2015.10.012

Mechenich, M.F., Žliobaitė, I., 2023. Eco-ISEA3H, a machine learning ready spatial database for ecometric and species distribution modeling. Sci Data 10, 77. https://doi.org/10.1038/s41597-023-01966-x

Olawoyin, A.M., Leung, C.K., Cuzzocrea, A., 2021. Open Data Lake to Support Machine Learning on Arctic Big Data, in: 2021 IEEE International Conference on Big Data (Big Data). Presented at the 2021 IEEE International Conference on Big Data (Big Data), IEEE, Orlando, FL, USA, pp. 5215–5224. https://doi.org/10.1109/BigData52589.2021.9671453

Purss, M.B.J., Peterson, P.R., Strobl, P., Dow, C., Sabeur, Z.A., Gibb, R.G., Ben, J., 2019. Datacubes: A Discrete Global Grid Systems Perspective. Cartographica 54, 63–71. https://doi.org/10.3138/cart.54.1.2018-0017

Shurygina, A., Titov, G., 2022. Building Datacube for Maritime Applications on Discrete Global Grid System, in: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Kuala Lumpur, Malaysia, pp. 2446–2449. https://doi.org/10.1109/IGARSS46834.2022.9884498

https://doi.org/10.5194/ica-abs-7-172-2024 |