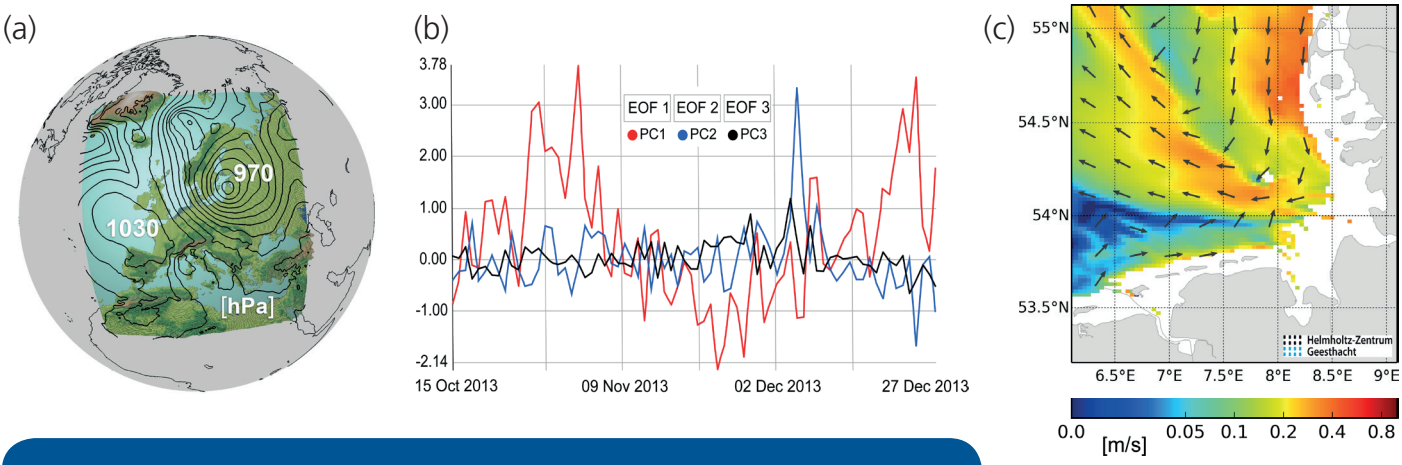
Data Management stories at DKRZ

Issue 3, November 2017



A selection of coastDat products: from regional atmospheric data (a) to local residual currents (c) with principal components (b) in the German Bight. More infos below in the table.

Data Management for coastDat at DKRZ

coastDat

Helmholtz-Zentrum Geesthacht
Centre for Materials and Coastal Research

CoastDat is a simulation-based data collection developed mainly for the assessment of long-term changes in data sparse regions such as the North Sea. A sequence of numerical models is employed to reconstruct various aspects of marine climate (e.g. storms, waves, surges, currents etc.) covering many decades, relying only on large-scale information such as atmospheric conditions or bathymetry. The idea of coastDat is to increase data homogeneity and consistency over long periods of time. Using quasi-realistic global and regional numerical models, the marine conditions are calculated in high resolution regarding time and space. The approach was developed more than ten years ago. coastDat has been applied successfully to various research questions concerning the North Sea, including, among others, assessments of the effectiveness of political measures to reduce chronic oil pollution or changes in wind and storm surge climate. The coastDat data set is used by more than 100 users (40% of them located in industry, 15% in authorities and 45% in research institutes). **http://www.coastdat.de**

German Climate Computing Center (DKRZ)

The mission of DKRZ is to provide high performance computing platforms, sophisticated and high capacity data management and services for premium climate science.

Besides providing HPC services, DKRZ supports projects in all aspects relevant to data management. This includes preparation, quality assessment, distribution, and long-term archiving of data.

www.dkrz.de

Overview & history of coastDat by the Institute of Coastal Research @ WDCC

2012	2013	2014	2015	2016	2017	Prospects
Atmosphere (SN-Remo)	Atmosphere (COSMO-CLM) North East Atlantic & Europe					Data from long-term regional (coupled) earth system model simulations and full 4-d reconstructions using advanced data science methods with high resolution in time and space.
Water level & currents North Sea	Water level & currents North Sea		Water level & currents Baltic Sea	Waves North Sea	Atmospheric chemistry Europe	
Waves Baltic Sea			Principal Component Analysis (PCA) of Residual Currents German Bight		Atmosphere (COSMO-CLM) – Bohai, Yellow and East China Sea – German Bight	
Waves North Sea						
Hydrodynamics North Sea			Global High Resolution Climate Reconstruction		Hydrodynamics & Biogeochemistry North & Baltic Sea	

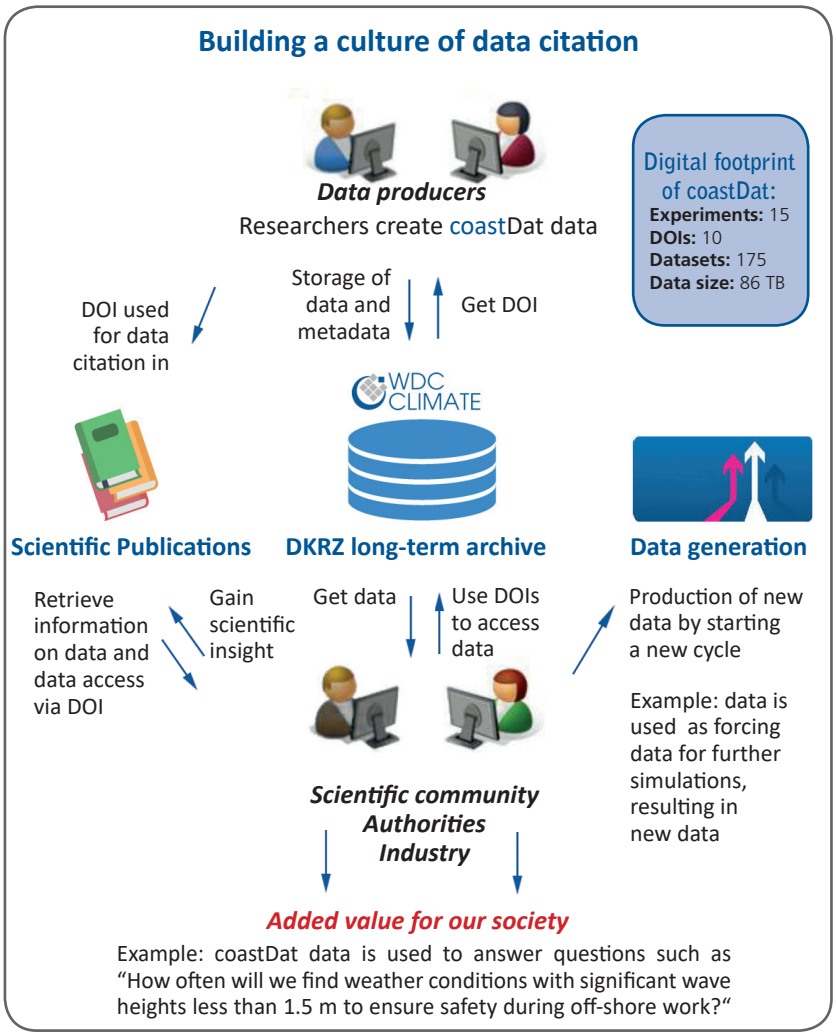
This table provides an overview of coastDat long-term data sets by the Institute of Coastal Research at the Helmholtz-Zentrum Geesthacht stored at the World Data Centre for Climate (WDCC). Blue arrows mark examples of the coastDat model chain. For example: global reanalysis data is used to force via spectral nudging a regional atmospheric model with higher resolution in time and space as the global model. The resulting regional atmospheric data , e.g. from COSMO-CLM, is used to force a hydrodynamic model to calculate water level and currents. Using this data, the residual currents of the German Bight are derived for the last 60 years. More information on ***www.coastdat.de, www.coastmap.org*** *and* ***www.coastdat.wdc-climate.de***

Data management for coastDat

The coastDat data sets are archived in the certified long-term archive at DKRZ and thus part of the World Data Centre for Climate (WDCC). Before data can be incorporated into the WDCC, thorough and systematic inspection of data and metadata is carried out by DKRZ in cooperation with the data providers. These tests include, for instance, verification of completeness, validity of formats and consistency of data and metadata. The results are meticulously documented and become part of the long-term archive. A special task of the coastDat project is the long production cycle, which requires a careful development of interfaces between data provider and archive. The final step of the archiving process is the assignment of DataCite DOIs, which provides citation information following best practices. Metadata collected by WDCC is published on **cera-www.dkrz.de** and distributed according to well-established standards. It is shared with other scientific portals, resulting in increased visibility of the data sets. WDCC provides long-term records of download statistics to the data producers, which indicate whether the data is of interest to users. The careful preparation of data and metadata by data providers and DKRZ enables the future use of the data by stakeholders.

Benefits for coastDat

- Long-term archiving of large model data volumes in WDCC, ensuring permanent and persistent access, preservation and discoverability.
- Provision of a high-quality data collection to facilitate long-lasting reuse.
- Technical quality verified and confirmed by an independent body.
- Handing over the responsibility for long-term archiving and data curation to WDCC.
- The assignment of DataCite DOIs as a recognized and well-adopted form of citation enables the close integration of scientific publications and research data and facilitates the creation of credit to the data producer for this dataset.
- Preservation in one common archive allows uniform search and access to all coastDat data sets.
- The visibility of the coastDat data collection is increased by sharing the metadata with other scientific portals.
- WDCC download statistics give the opportunity to draw conclusions regarding future extensions and improvements of the data set.



The publication of data in a long-term archive combined with a standardized form of citation supports increased transparency, traceability and reproducibility in science. Data with a DOI is easy to find and a necessary contribution to good scientific practice. Scientific data is increasingly used by third parties and cited in publications. Only cited data can be counted and tracked (in a similar manner to journal articles) to measure impact.

DOIs for dynamic datasets

Some coastDat data sets are dynamic in time. New records are created and added over time to extend the time series; existing archived data sets are not changed. The DOIs for these data sets are not affected by this as long as the same data creation methods are in place and the citation metadata is still valid. This procedure brings advantages for data producers as well as data users by establishing a transparent and clear assignment of the dynamic data and the DOI.
Example: doi:10.1594/WDCC/coastDat-2_TRIM-NP-2d

Outlook:

Various new data from regional earth system models for long time periods at high temporal and spatial resolution is expected in future. Regarding the changes in environmental conditions, specific questions about individual processes can be investigated with the assistance of coastDat data sets and added value for the society can be generated. DKRZ's smart data management and long-term archiving at WDCC are key to meeting the challenges.