Advancing Geospatial Data Analysis with XDGGS

Justus Magin, Benoit Bovy, Jean-Marc Delouis, Anne Fouilloux, Alejandro Coca-Castro, Ryan Abernathey, Peter Strobl, Alexander Kmoch, Tina Odaka
Corresponding author: tina.odaka@ifremer.fr

Funded by the European Union   Destination Earth   IMPLEMENTED BY   EUMETSAT   esa   ECMWF
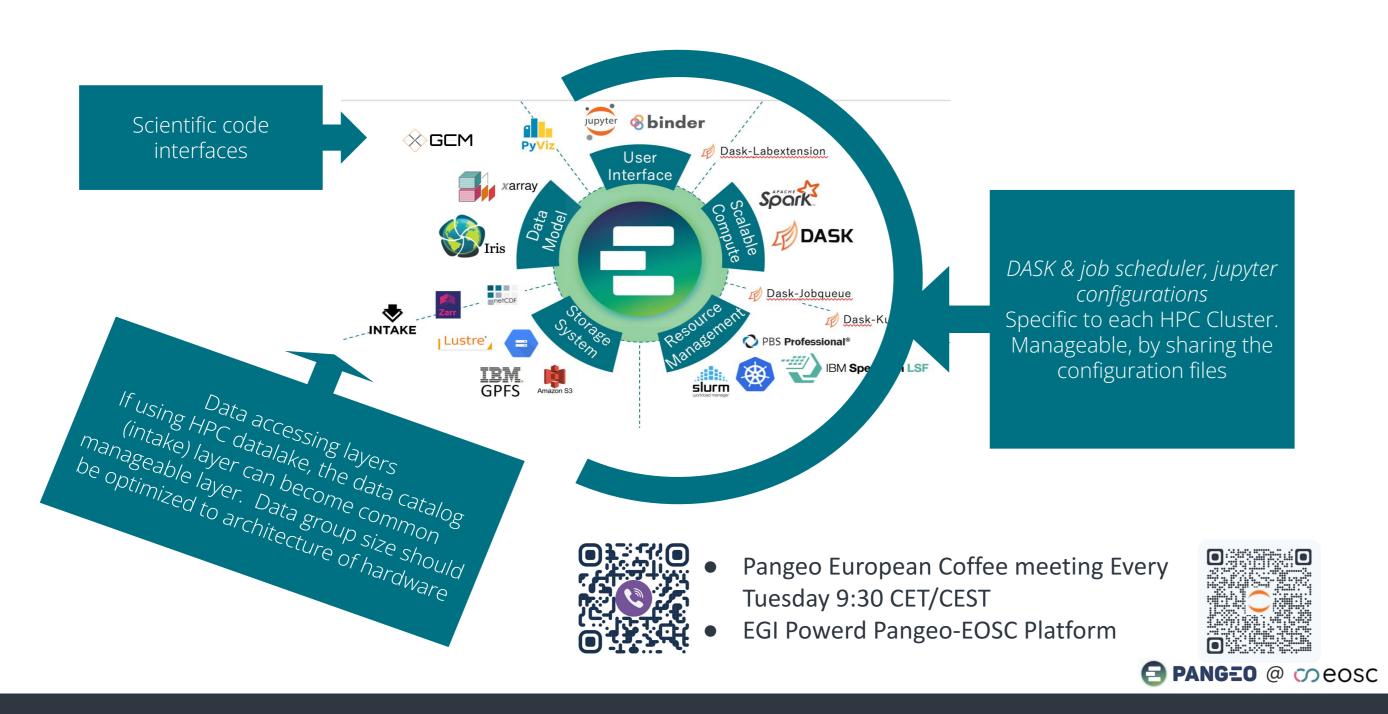
***XDGGS** is an extension for **Xarray**, spearheaded by the Pangeo community, that introduces a high-level API that allows the use of Discrete Global Grid Systems (**DGGS**).*
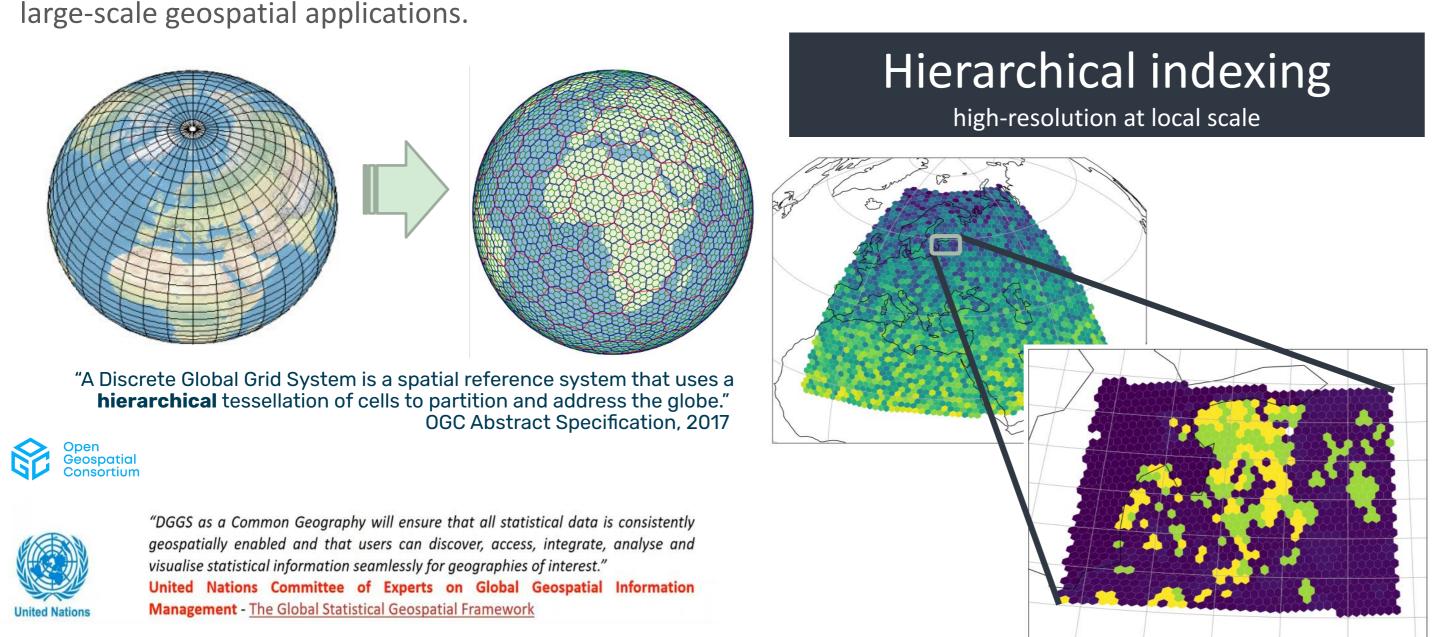
Pangeo in the world and in Europe

PANGEO is a **global community initiative** for Big Geoscience Data with active **developers**, **scientists** and **users**.

From laptops to cloud and HPCs.

Scientific code interfaces

DASK & job scheduler, jupyter configurations Specific to each HPC Cluster. Manageable, by sharing the configuration files

Data accessing layers If using HPC datalake, the data catalog (intake) layer can become common manageable layer. Data group size should be optimized to architecture of hardware

- Pangeo European Coffee meeting Every Tuesday 9:30 CET/CEST
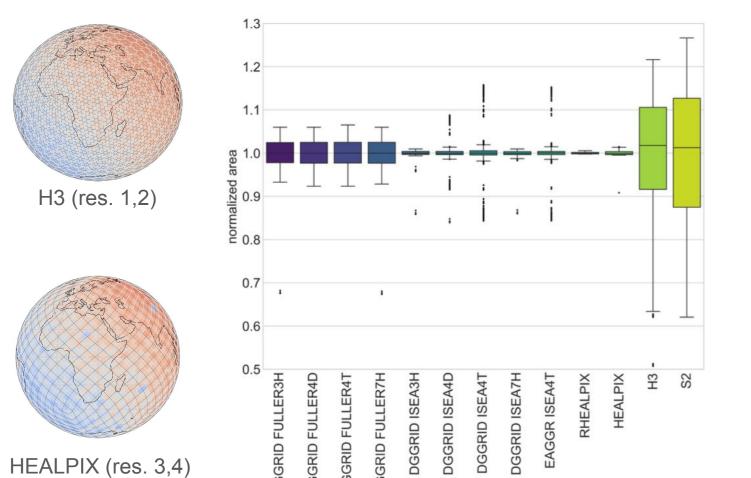- EGI Powerd Pangeo-EOSC Platform

PANGEO @ eosc

What are Discrete Global Grid Systems (DGGS)?

- Traditional geospatial representations on two-dimensional planes often suffer from distortions in area, distance, and angles, particularly at global scales.
- Discrete Global Grid Systems (DGGS) provide a solution by using a hierarchical grid structure, reducing distortions and enabling more accurate spatial data analysis.
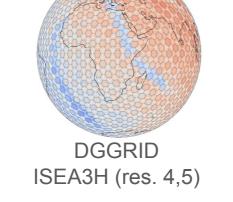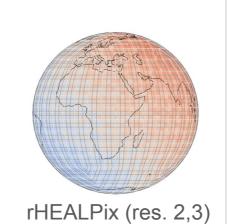- DGGS is essential for integrating and aggregating multi-resolution datasets, making it valuable for large-scale geospatial applications.

Hierarchical indexing
high-resolution at local scale

"A Discrete Global Grid System is a spatial reference system that uses a **hierarchical** tessellation of cells to partition and address the globe."
OGC Abstract Specification, 2017

"DGGS as a Common Geography will ensure that all statistical data is consistently geospatially enabled and that users can discover, access, integrate, analyse and visualise statistical information seamlessly for geographies of interest."
United Nations Committee of Experts on Global Geospatial Information Management - The Global Statistical Geospatial Framework

DGGS: Equal-area assessment

Boxplots of normalized area values for the cells of open-source DGGS:
Kmoch, A., Bovy, B., Magin, J., Abernathey, R., Coca-Castro, A., Strobl, P., Fouilloux, A., Loos, D., Uuemaa, E., Chan, W. T., Delouis, J.-M., and Odaka, T.: XDGGS: A community-developed Xarray package to support planetary DGGS data cube computations, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-4/W12-2024, 75–80, https://doi.org/10.5194/isprs-archives-XLVIII-4-W12-2024-75-2024, 2024.

H3 (res. 1,2)

HEALPIX (res. 3,4)

DGGRID ISEA7H (res. 2,3)

DGGRID ISEA3H (res. 4,5)

rHEALPix (res. 2,3)

Current DGGS libraries support:
- HEALPIX (via healpy)
- H3 (via h3-py)
- DGGRID ISEA7H (via DGGRID and dggrid4py)
- Other well-known open-source DGGS types that could be added in the future are rHEALPix (via rhealpixdggs), Google S2 (via s2geometry), or DGGRID ISEA3H
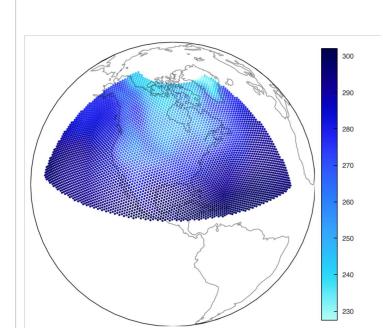
Learn more

A community of developers, scientists and users

- More info & links here: GitHub - xarray-contrib/xdggs: Xarray extension for DGGS
https://github.com/xarray-contrib/xdggs
- Try it : **pip install xdggs**
- Join us for bi-weekly 30 minute developer / community meetings on Monday 14:00 CET / CEST.
https://github.com/xarray-contrib/xdggs/discussions/45
- Related projects
  - https://github.com/IAOCEA/xarray-regridding
  - https://github.com/IAOCEA/healpix-convolution
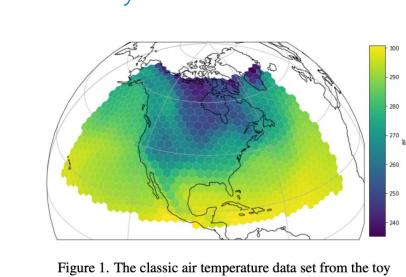
Add an event

XDGGS in action

Pangeo teams up with OGC and OSGeo

- XDGGS is an Xarray extension developed through a sprint at the ESA BiDS'23 conference, led by the Pangeo community in collaboration with OSGeo and the OGC.
- The OGC DGGS SWG works on standardizing concepts and definitions around DGGS. The cooperation ensures reference systems and grids are usable across libraries and APIs.
- XDGGS provides a unified, high-level API that simplifies the use of DGGS types like HEALPIX, H3, or DGGRID ISEA7H, using DGGS-specific cell identifiers as an index within Xarray.
- By integrating DGGS with Xarray, XDGGS enables efficient geospatial workflows, reducing complexity and enhancing Xarray's data processing capabilities.
- This extension promotes FAIR data practices, improving data access and interoperability for large-scale geospatial applications.
- XDGGS is user-friendly, abstracting the complexities of various DGGS libraries, and makes DGGS accessible for planetary-scale data analysis.
- XDGGS supports geospatial analysis for gridded data like global climate models, oceanography, satellite imagery, and other raster datasets.
- Demonstrations through executable notebooks highlight XDGGS's ability to handle complex geospatial data, providing a valuable tool for scientists and developers working with global datasets.
- Detailed design:
https://github.com/xarray-contrib/xdggs/blob/main/design_doc.md

Xarray + DGGS = XDGGS

Figure 1. The classic air temperature data set from the toy weather data tutorial in Xarray, converted and plotted via a hexagonal DGGS.

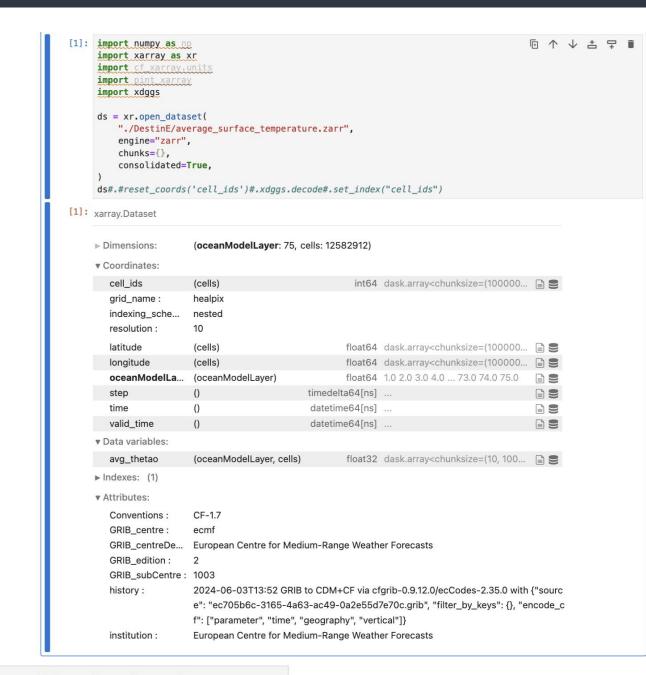Open Geospatial Consortium   PANGEO   OSGeo

XDGGS and DestinE : Healpix

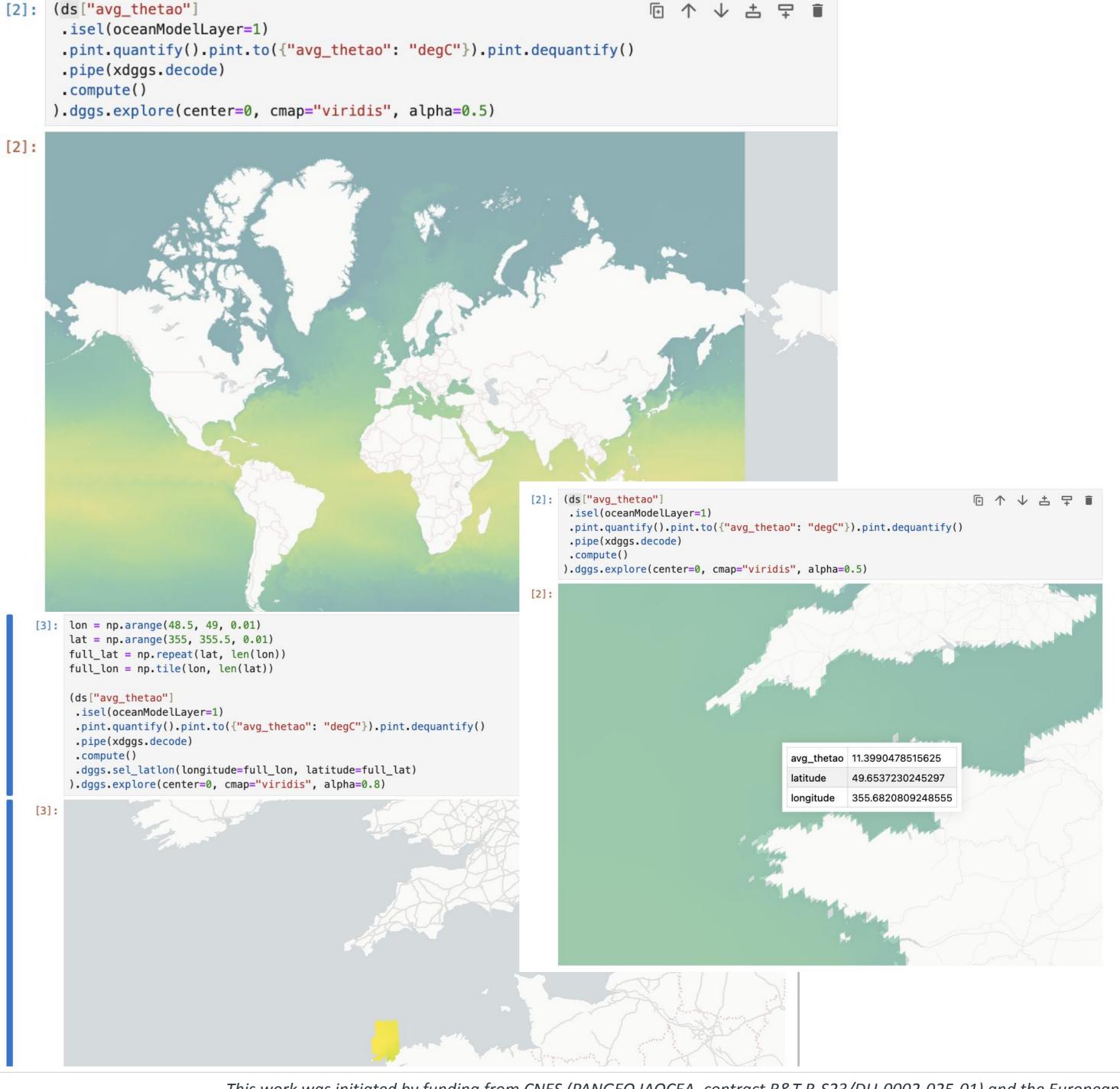- Try out the new example here opening zarr format of healpix data.

https://github.com/tinaok/xdggs_examples/blob/main/visualisation/xdggs-explore-demo.ipynb

- Transforming DestinE Climate Twin data to zarr

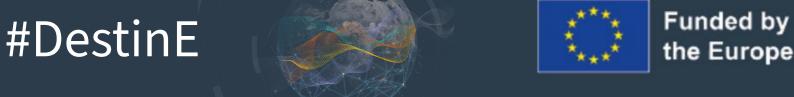https://destination-earth.github.io/DestinE_ESA_GFTS/ClimateDT-zarr-2D-3D.html

DestinE Climate Twin in Healpix with Xarray, Zarr, longboard and XDGGS