PROJECT CLOUD-READY DATA
A REAL-WORLD COMPARISON OF TileDB TO ZARR

PETER KILLYICK and KEVIN DONKERS • INFORMATICS LAB RESEARCHERS with THE TILEDB DEVS
@_DPeterK_ • @kaedonkers
ABOUT ME

Peter Killick • @DPeterK • www.informatics-lab.co.uk
ABOUT THE MET OFFICE
ABOUT THE LAB
THE ENVIRONMENTAL FUTURES AND BIG DATA IMPACT LAB
PROJECT

CLOUD-READY DATA
AIMS

Peter Killick  ·  @_DPeterK_  ·  www.informaticslab.co.uk
WHAT WE DID

tiledb-netcdf

NC → [tile]DB → Zarr
TILEDDB FEATURES
u-wind (t: 24, z: 72, y: 1600, x: 2560)
v-wind (t: 24, z: 72, y: 1600, x: 2560)
w-wind (t: 24, z: 72, y: 1600, x: 2560)

u-wind, v-wind, w-wind
USING TILEDB
An adapter to convert NetCDF files to TileDB or Zarr arrays.

**Usage Examples**

Here we'll demonstrate using this library to convert NetCDF files to TileDB and Zarr, and read the results using Iris and Xarray.

**Converting to TileDB**

[https://github.com/informatics-lab/tiledb_netcdf](https://github.com/informatics-lab/tiledb_netcdf)
ZARR IN COMPARISON
$ la my-zarr

. 
...
.zarray
.zattrs
0.0.0
1.0.0
2.0.0
3.0.0
4.0.0
5.0.0
6.0.0
7.0.0
8.0.0
9.0.0
10.0.0
PROJECT FINDINGS
FINDINGS

● Both usefully store NetCDF data
● Both allow domain-specific processing
● Both present data conversion challenges
● Neither fully represent metadata
● TileDB provides a richer storage format
● Zarr provides a lightweight, extensible API
THE NEW LIMITING FACTORS
FIND OUT MORE

- [https://medium.com/informatics-lab/analysis-ready-data-47f7e80c80a42](https://medium.com/informatics-lab/analysis-ready-data-47f7e80c80a42)
- [https://medium.com/informatics-lab/create-zarr-from-pp-files-ffa6b7972d6f](https://medium.com/informatics-lab/create-zarr-from-pp-files-ffa6b7972d6f)
RECAP

- Introductions
- Project Cloud-Ready Data
- **TileDB** for Cloud-Optimised Data
- **Zarr** in Comparison
- Project Findings
THANK YOU!

Anything else? Get in touch:
www.informaticslab.co.uk • @informatics_lab • hello@informaticslab.co.uk