Exploring Lossy Data Compression in an Online Laboratory for Climate Science and Meteorology

Juniper Tyree[1], **Sara Faghih-Naini**[2],
Peter Dueben[2], **Karsten Peters-von Gehlen**[3], **Heikki J. Järvinen**[1]

ESiWACE3 WP 3.3: [1] INAR, University of Helsinki, [2] ECMWF, [3] DKRZ

Sharing is encouraged

EGU25, Session EOS4.6, 30.04.2025

EuroHPC Joint Undertaking

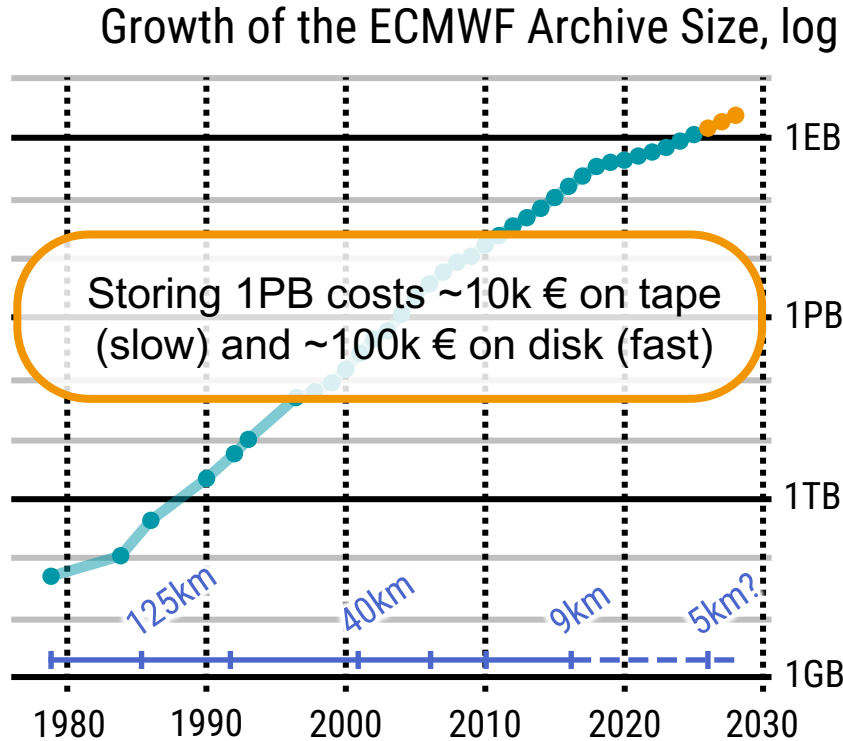**Open Science** is all about **sharing** and **collaboration**

We want others to discover, **try out**, and benefit from our work

But trying out something new can be a time-consuming **hurdle**

e.g. needing to install new software, which may have tricky dependencies

Trying something out for yourself is especially important if you have **doubts**

Example: Lossy Compression of CliMet model outputs in ESiWACE3

Growth of the ECMWF Archive Size, log

Storing 1PB costs ~10k € on tape (slow) and ~100k € on disk (fast)

1EB

1PB

1TB

125km   40km   9km   5km?

1GB

1980   1990   2000   2010   2020   2030

European Centre for Medium-Range Weather Forecasts archive size
Dataset provided by Mike Hawkins (ECMWF), Figure adapted from Milan Klöwer (Oxford)

We need **lossy compression**

But losing data accuracy is *scary*

To **convince yourself** that lossy compression is safe, you need to **try it out yourself**

Trying something out should be easy to **engage earlier** and **delay setup costs** after **convincing** users

Solution for the entire community

i.e. not just for our project (lossy compression) but also for your project, documentation, demos, …

The Online Laboratory for Climate Science and Meteorology

What's easier than opening a URL in your browser?

Serverless In-Browser Interactive Computing

jupyterlite

PYODIDE

WA

Code and data stay *local*

Internet icon by Freepik at
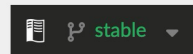https://www.flaticon.com/free-icon/internet_4861920

No setup, **no installation**, <1min to start

Reproducible and version-locked
Ensure your examples keep working

Supports many compiled scientific + especially Earth Science Python packages

Extra support for **accessing large datasets**
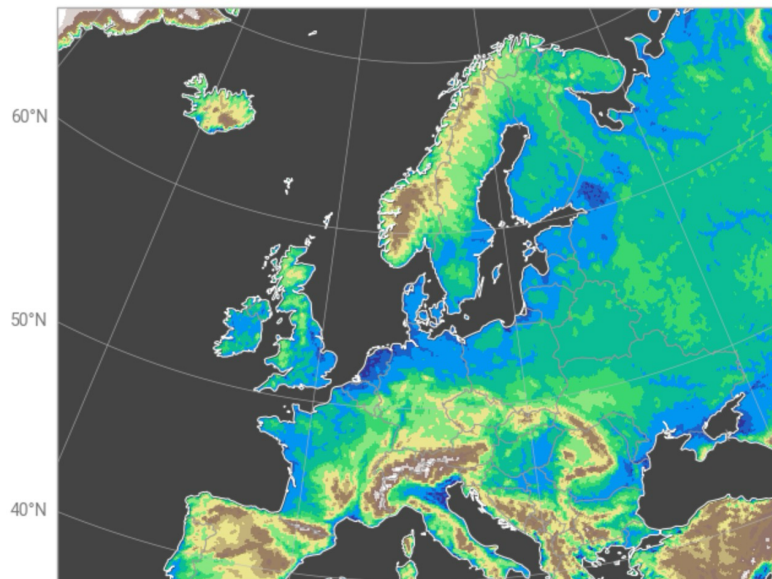
Ease of use: same code, same results

Example: Software documentation is non-interactive

Demo: Interactive Documentation in the Online Laboratory


```
[5]: # Compress the geopotential (m^2/s^2) using ZFP

z = CodecStack(
    Zfp(mode="fixed-accuracy", tolerance=100)    ← strict error bound
).encode_decode_data_array(data[0].to_xarray().z)
```

```
chart = earthkit.plots.Map(domain=[-15, 35, 32, 72]) •••
```
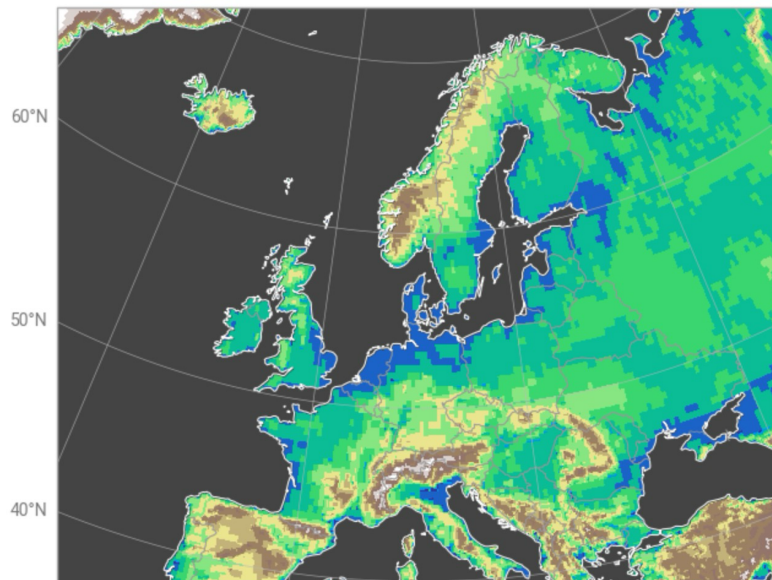
Demo: Interactive Documentation in the Online Laboratory

model-orography.ipynb

File   Edit   View   Run   Kernel   Settings

Python (Pyodide)

/ .sessions / 04-22-2025-14-11-02-790053d4 /

Name	Modified
model-orograph...	now
README.md	7d ago
requirements.txt	7d ago

[7]:
```python
# Compress the geopotential (m^2/s^2) using ZFP

z = CodecStack(
    Zfp(mode="fixed-accuracy", tolerance=10000)   ← loose error bound
).encode_decode_data_array(data[0].to_xarray().z)
```

```python
chart = earthkit.plots.Map(domain=[-15, 35, 32, 72]) •••
```


Simple   1   Python (Pyodide) | Idle

Mode: Edit   Ln 1, Col 1   model-orography.ipynb   4

How can **You** benefit from the Online Laboratory today?

docs.climet.eu

Interactive and Reproducible Open Science Laboratory

Share your existing **notebooks** from repos / gists / URLs

Customize the provided **packages** and **versions**

Ongoing development and support and new packages

Explore lossy data compression on CliMet model data

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

Exploring Lossy Data Compression in an
Online Laboratory for Climate Science and Meteorology

Thank You for Your attention!
Please reach out for help using the lab

Learn more about **Data Compression** in Session **ESSI2.13** in room -2.92 on **Thursday 16.15-18.00**

Find out more about the **Compression Laboratory** from **Poster X4.70** in hall X4 on **Thursday morning**

Juniper Tyree
juniper.tyree@helsinki.fi
iD 0000-0002-7923-9609
juntyr

EGU25, Session EOS4.6, 30.04.2025