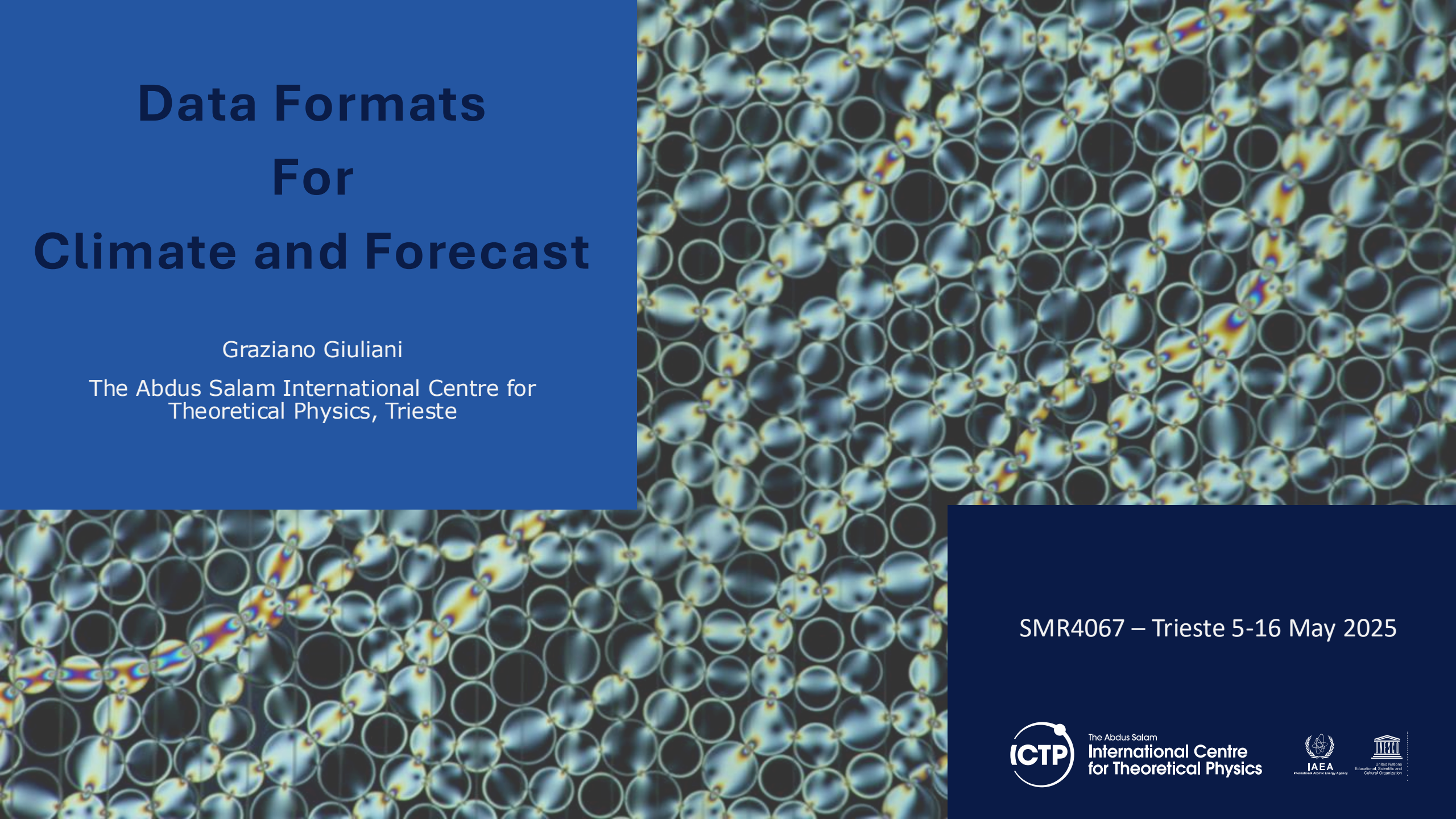
Data Formats For Climate and Forecast

Graziano Giuliani

The Abdus Salam International Centre for Theoretical Physics, Trieste



SMR4067 – Trieste 5-16 May 2025

The Abdus Salam International Centre for Theoretical Physics

SECTION TITLE

The big problem

Predicting the future

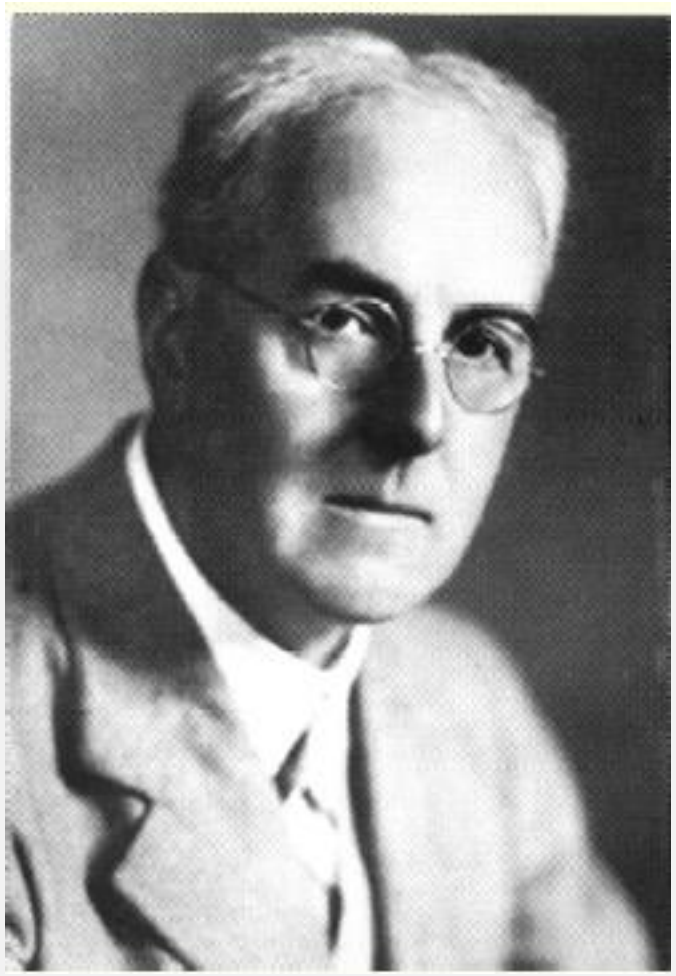
- Science
- Strategic advantage in War
- Public Service
- Profit
- Fear

The Gods of weather

Botticelli – La nascita di Venere

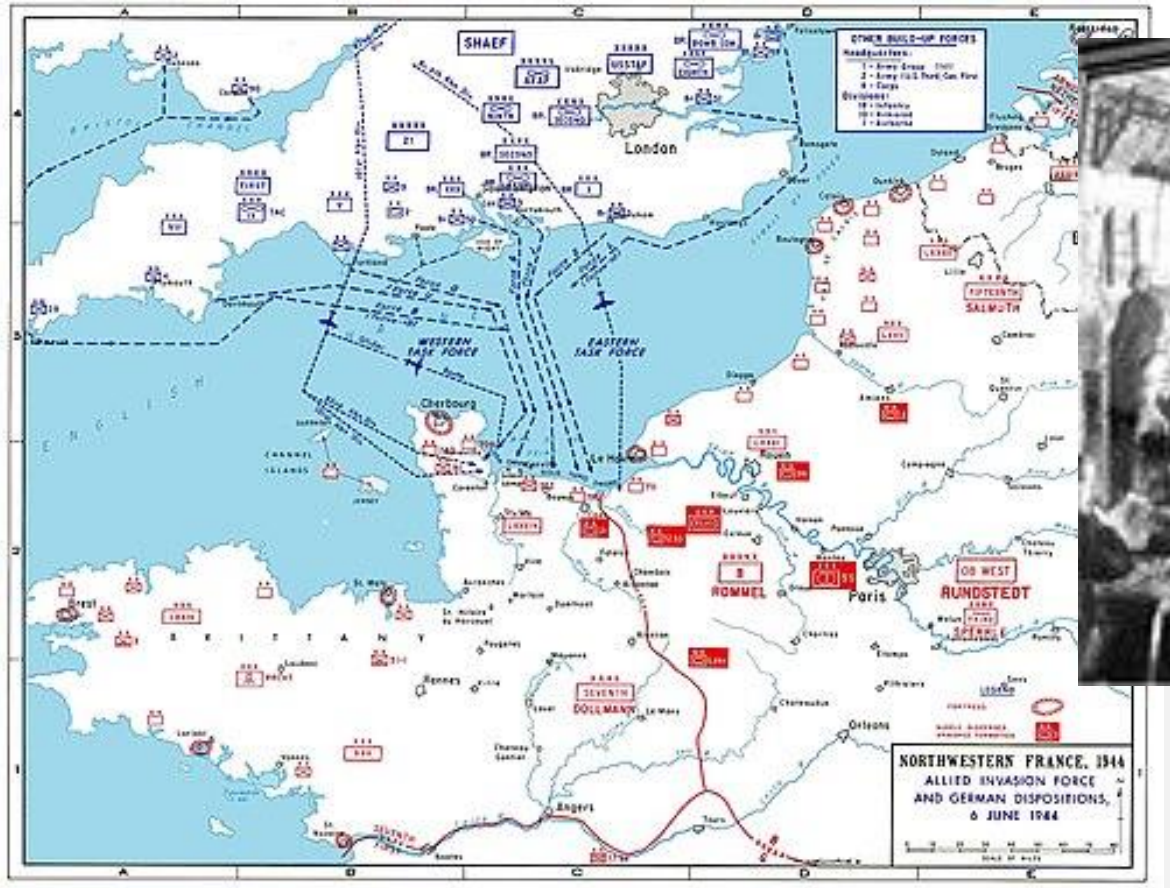
The Father of Weather Forecast

The Eskdalemuir Observatory - BBC



Lewis Fry Richardson - Wikipedia

D-Day Weather map room



Source - Wikipedia

The Joint Weather Bureau

John Von Neumann - Wikipedia

Dr. George Cressman, JNWP director, studying numerical display. IBM photo.

Organization and Operation of the Joint Numerical Weather Prediction Unit

EDWIN B. FAWCETT; LCDR. WILLIAM E. HUBERT, USN; AND LCDR. ALBERT L. STICKLES, II, USN

INTRODUCTION

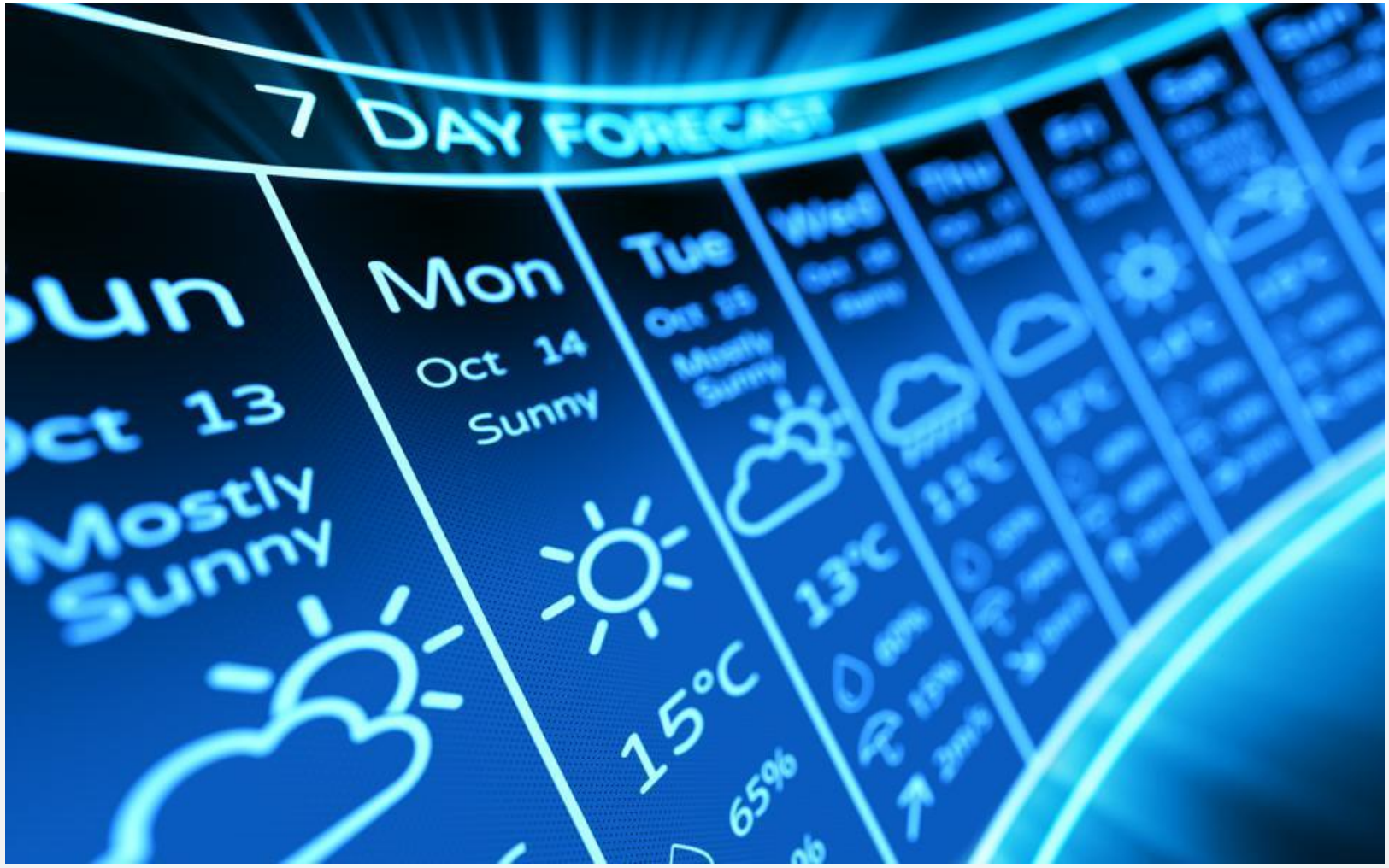
ON the first of July 1954 a group of 15 meteorologists gathered at the Central Office of the Weather Bureau in Washington, D. C. They were the nucleus of the Joint Numerical Weather Prediction Unit. The Weather Bureau, Naval Aerological Service, and Air Weather Service of the Air Force had decided to combine forces in a revolutionary approach to forecasting. Much research work had already been done in this country and abroad on numerical weather prediction using high-speed electronic computers developed during and since World War II.* The time had arrived to put the experimental techniques into practice. It was felt that a great deal could be learned by utilizing the new methods on a daily operational basis.

The beginning of the new Unit was humble indeed. The Weather Bureau very kindly provided four rooms in the Annex at its Central Office pending completion of new quarters at Suitland, Maryland. The IBM Type

* Several successful numerical forecasts were made at the Princeton Institute for Advanced Study under the sponsorship of the Office of Naval Research.

J. G. Charney - Wikipedia

Weather for Profit



The Fear for the Future

Takens Theorem

F. Takens (1981). "Detecting strange attractors in turbulence". In D. A. Rand and L. S. Young (ed.). Dynamical Systems and Turbulence, Lecture Notes in Mathematics, vol. 898. Springer-Verlag. pp. 366–381.

"…From the analysis of a time series , where is an observable sampled at the discrete times , it is possible (if we know that the system is deterministic and is described by a finite dimensional vector, and Mis large enough) to determine the proper variable ,and then determine the main features of the dynamics, e.g. the Lyapunov exponent and D…"

Floris Takens - Wikipedia

Mark Kac - Wikipedia

Kac's Lemma

Extends Poincare's theorem:

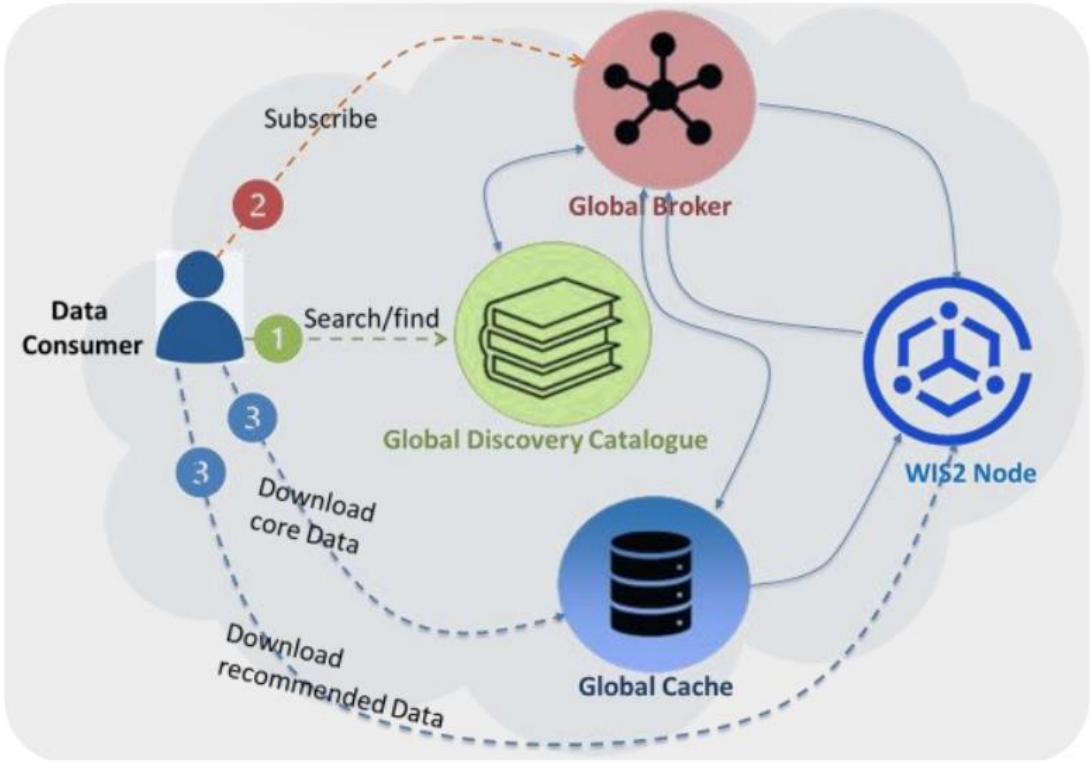
"...in a measure space the orbit of almost all the points contained in a set A of such space, whose measure is M(A), return to A within an average time which is inversely proportional to M(A)…"

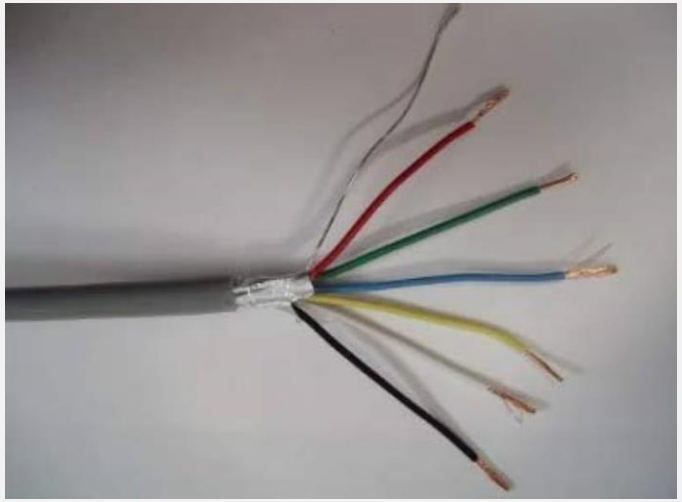
The UN backed WMO

WMO and data

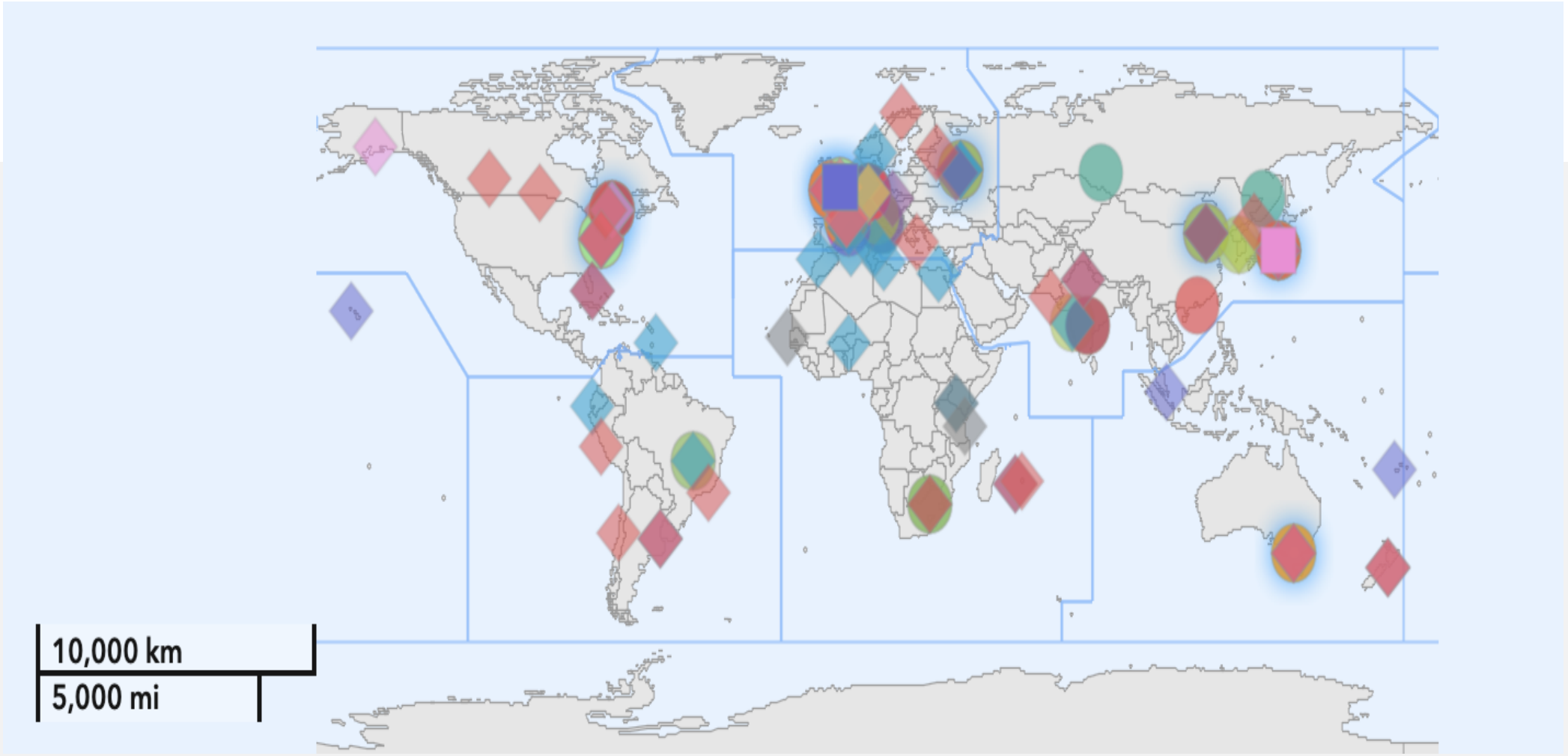
WORLD METEOROLOGICAL ORGANIZATION

WIS 2.0





WMO Integrated Processing and Prediction System



ECMWF

ecmwf/**eccodes**

ECMWF's GRIB and BUFR decoding/encoding library

46 Contributors

1 Used by

215 Stars

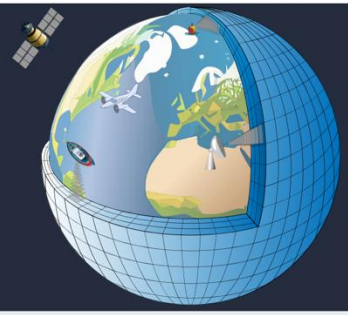
91 Forks

World Climate Research Program

WCRP
World Climate Research Programme

ESMO
ESMO Website
Earth System Modelling and Observations

Coordinating modelling and observations to understand our climate



CMIP Coupled Model Intercomparison Project
A WCRP project

Home

CMIP

Coupled Model Intercomparison Project

CMIP is a project of the World Climate Research Programme (WCRP) providing climate projections to understand past, present and future climate changes. CMIP and its associated data infrastructure have become essential to the Intergovernmental Panel on Climate Change (IPCC) and other international and national climate assessments.

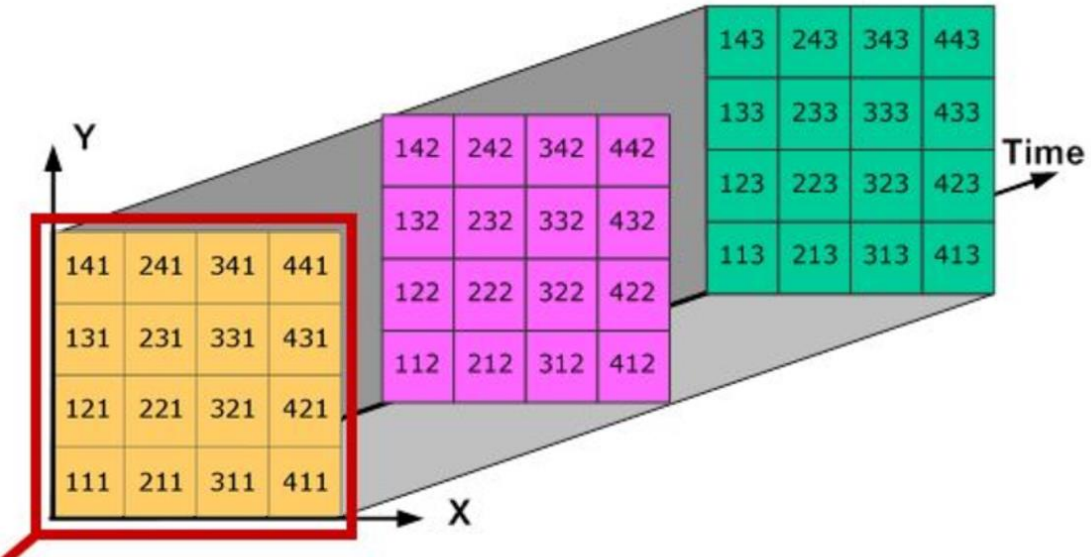
Watch introduction video

WGORC logo

CF NetCDF

```
netcdf mynetcdf {
dimesions:
        x=4;
        y=4;
        time=UNLIMITED;
variables:
        float x(x);
        float y(y);
        int time(time);
        float temperature(time,x,y);
data:
        x = 10, 20, 30, 40;
        y = 110, 120, 130, 140;
        time = 31, 59, 90;

Temperature = 111, 211, 311, 411, 121, 221, 321,
        421, 131, 231, 331, 431, 141, 241, 341, 441;
}
```



time = 1

x = 1 to 4

y = 1 to 4

Pangeo Xarray

xarray

Pangeo is first and foremost a community promoting open, reproducible, and scalable science. This community provides documentation, develops and maintains software, and deploys computing infrastructure to make scientific research and programming easier.

One of the supported tools is Xarray, a library for N-D labeled arrays and datasets in Python. It is interoperable with the scientific Python ecosystem including NumPy, Dask, Pandas, and Matplotlib.

The netCDF4 (HDF5) is the “default” data type in Xarray. And with the use of CF-array it readily understands the CF conventions. And it can be extended with operators for climate like in xCDAT.

And with cfgrib can read WMO format data.

Where we are not doing well

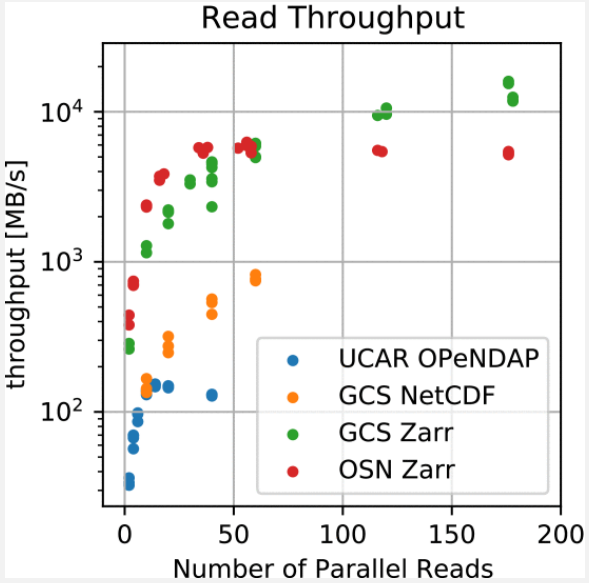
Yes, we have a lot of data. Somewhere.

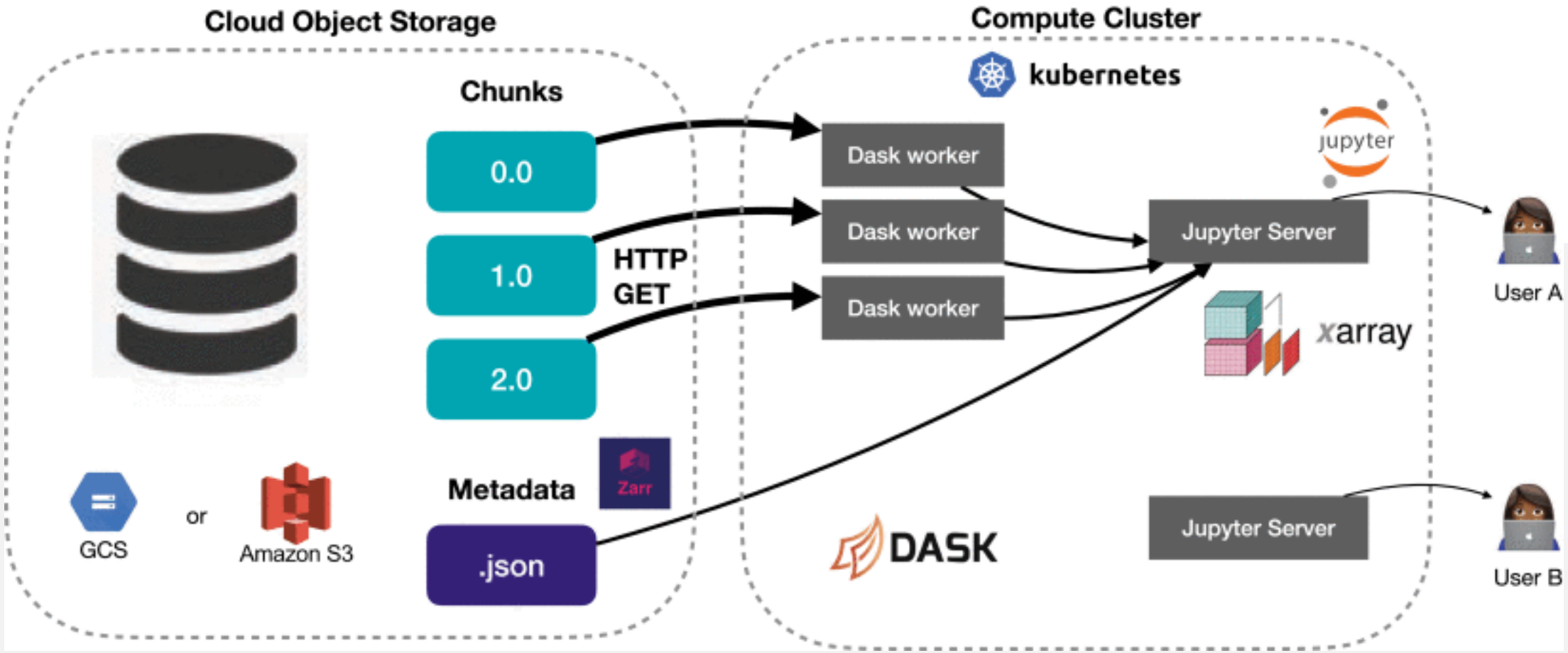
- We STILL have to download data if we are not the data producer. And we still are limited to the POSIX file standard.
- The data are typically setting on disk storage of a few data centers: only “first world” researcher have fast physical access to the data. And are distributed mostly on a file by file basis.
- We can use compression, but any data transfer and access to the data has a cost. We must reduce data access cost.
- A BIG percentage of the data analysis is repeated and repeated and repeated. Very few institution provide server side computing support and data caching.
- The current model to get fast data access is to duplicate and distribute.

How to access data

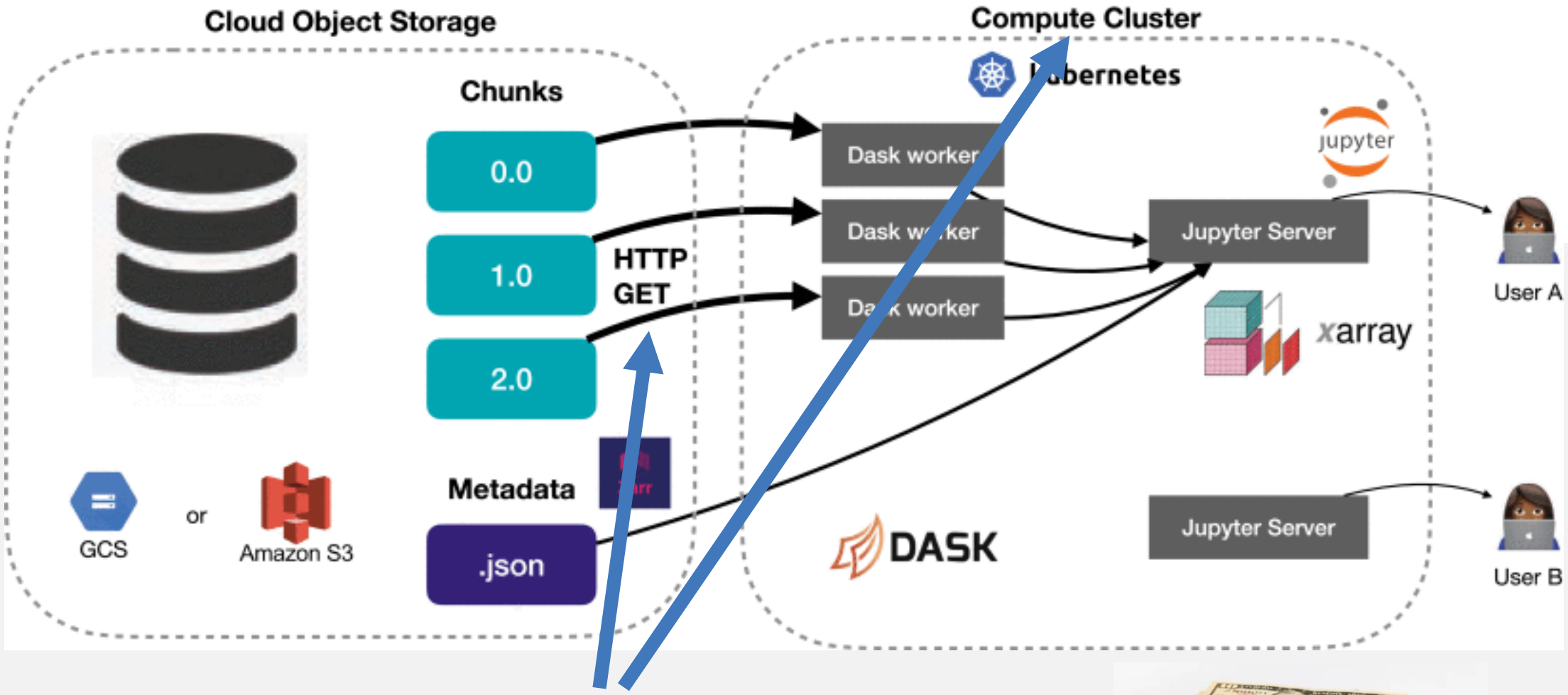
In the past some effort has been put on providing some basic tools for a web based data access protocol:

- OpenDAP – Open Data acces protocol. It was great, but current implementation don’t scale to petascale applications and massively parallel distributed processing in mind. Only basic sub-setting service available from server.





Can you spot the problems?



Can you spot the problems?

Article Type: Research Article

Exascale Computing and Data Handling: Challenges and Opportunities for Weather and Climate Prediction

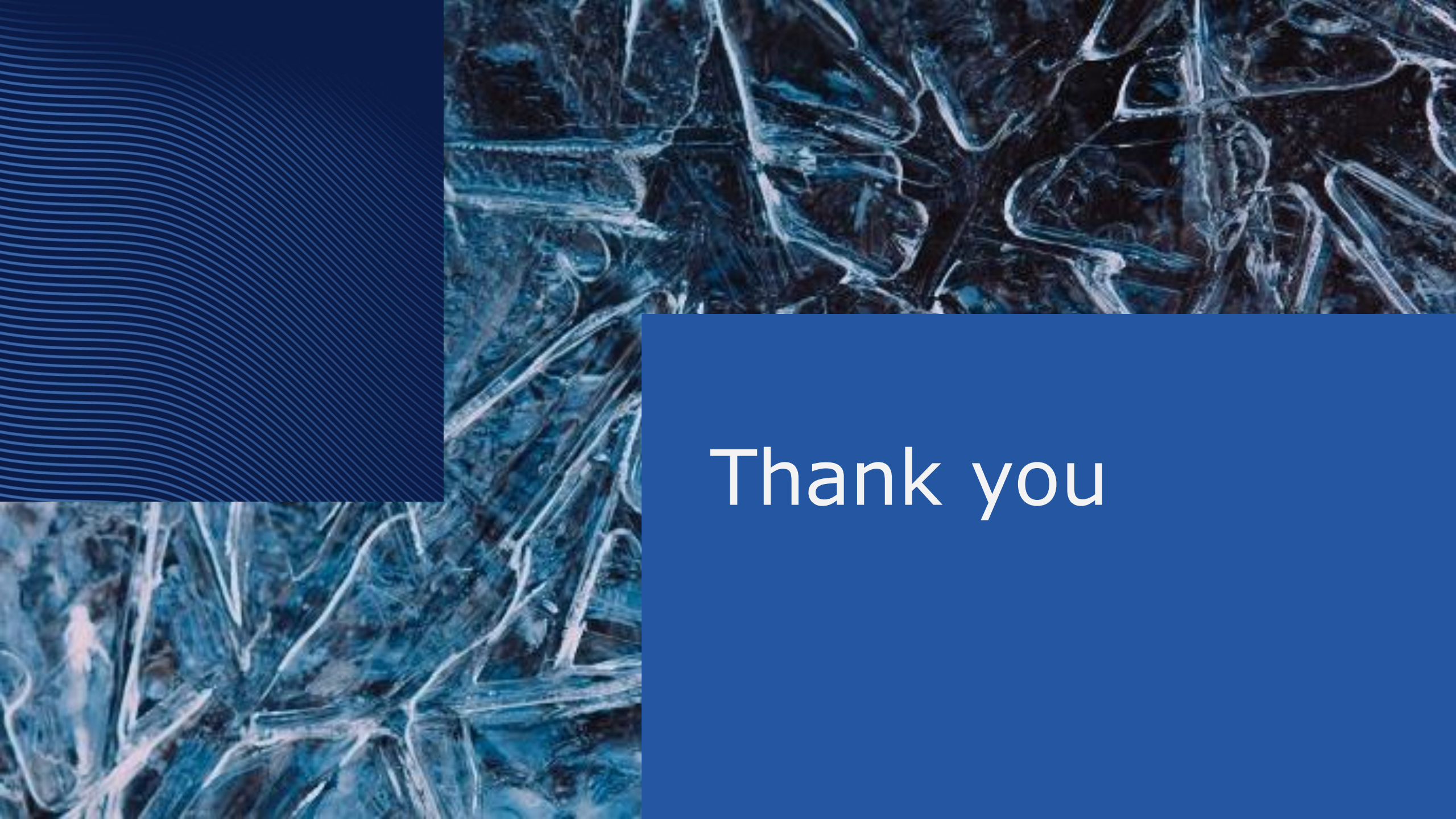
Mark Govett, Bubacar Bah, Peter Bauer, Dominique Berod, Veronique Bouchet, Susanna Corti, Chris Davis, Yihong Duan, Tim Graham, Yuki Honda, Adrian Hines, Michel Jean, Junishi Ishida, Bryan Lawrence, Jian Li, Juerg Luterbacher, Chiasi Muroi, Kris Rowe, Martin Schultz, Martin Visbeck, and Keith Williams

Online Publication: 11 Dec 2024

Print Publication: 01 Dec 2024

DOI: https://doi.org/10.1175/BAMS-D-23-0220.1

Page(s): E2385–E2404



Thank you