Towards Structure-Aware Earth System Data Management

Jakob Lüttgau, Julian Kunkel, Bryan N. Lawrence, Sandro Fiore, Huang Hua
How to manage Earth System Data?
What to optimize? Write throughput on generation? Avoid transformation?
Data Representations

Different views to the same data. Suboptimal serialization on storage.

Formats

- **Raw**: Binary, optimized for transmission
- **Pre/Post**: Optimized for fast reading or locality
- **Out**: Optimized for fast writing

Post-Processing

- Single Value: Temperature Anomaly
  - Some average
- Images/Movies
  - CSV/Plots (x=time, y=CO2)

Domain Decomposition

Layout on Storage
Middleware for Earth System Data
Adaptively choose backends. Discriminate by data, metadata and access type.
Summary, Status and Outlook

- Architectures likely to become more heterogeneous
- Systems prohibitively complex for manual optimization

Status

- Reports and design documents publicly available: http://esiwace.eu (WP4, Deliverables 4.1 and 4.2)
- Prototype to demonstrate viability of adaptive tier selection

Outlook

- Open development of middleware, licensed under LGPL: http://github.com/ESiWACE
- Backends being developed for Object Storage and MongoDB
- NVM backends as hardware becomes more widely available
The ESiWACE project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 675191

Disclaimer: This material reflects only the author’s view and the EU-Commission is not responsible for any use that may be made of the information it contains