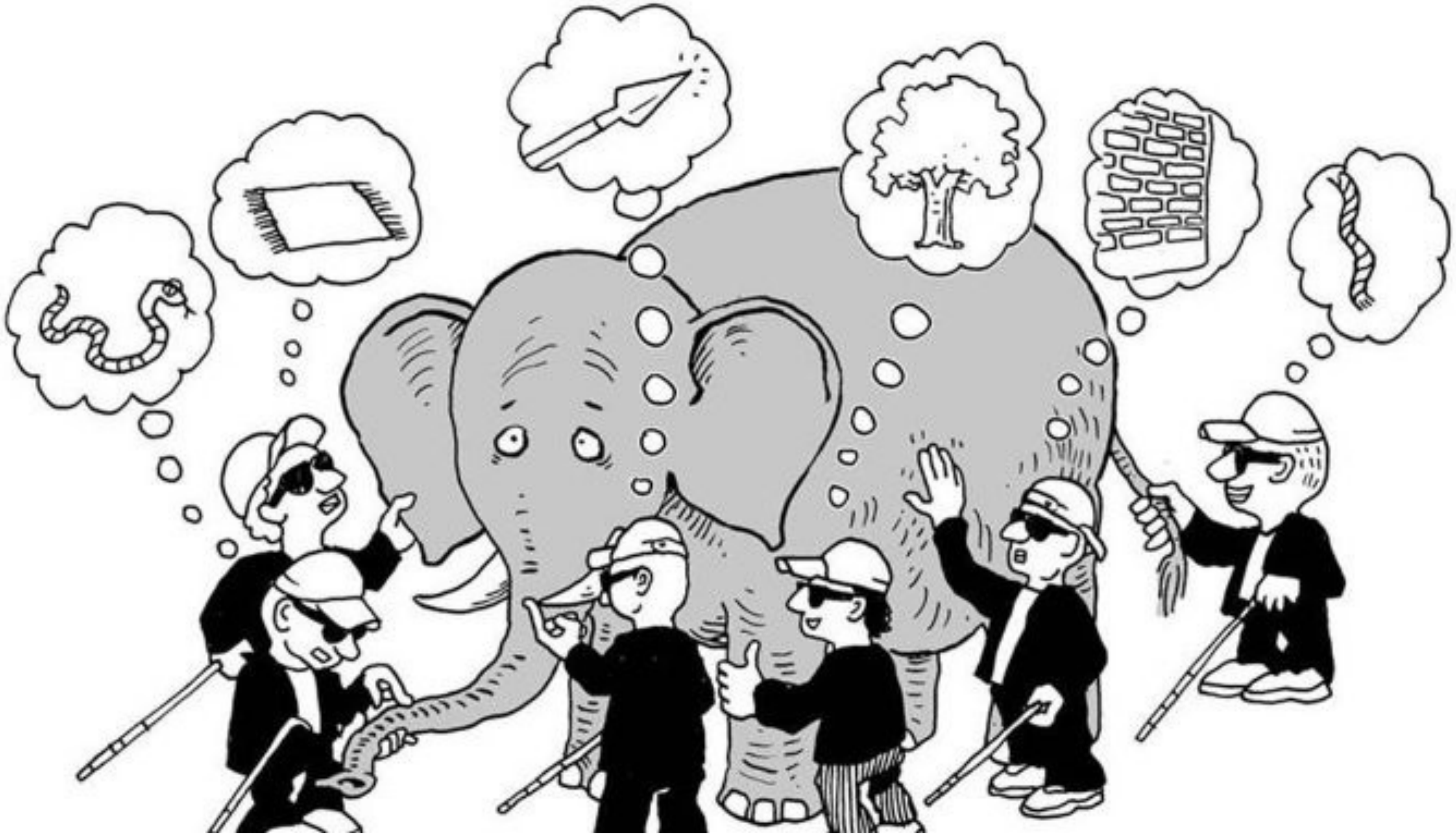
ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES

$\int \mathcal{M}^2 \, dt$

COMPUTATIONAL SCIENCE AND SCIENTIFIC MACHINE LEARNING TO ENABLE DIGITAL TWINS OF THE OCEAN

Patrick Heimbach
The University of Texas at Austin

What are Digital Twins?



(Parable of the Blind Men and an Elephant)

Role of Digital Twins – A DITTO perspective

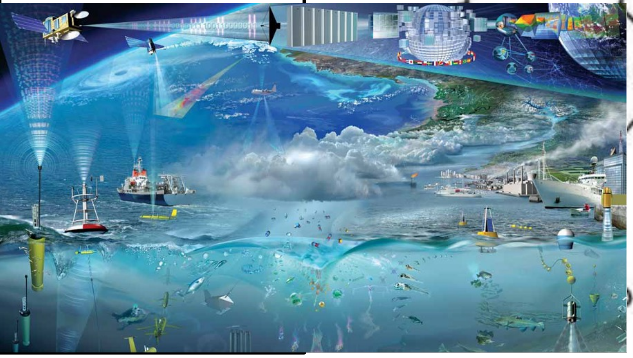
Role of Digital Twins – A DITTO perspective

“Digital Twins of the Ocean will bring together ocean data, models and digital information with those who are planning and regulating human ocean interactions. They will become an indispensable element of sustainable development of the ocean space”

– Martin Visbeck (GEOMAR)

What are Digital Twins? – A DITTO perspective

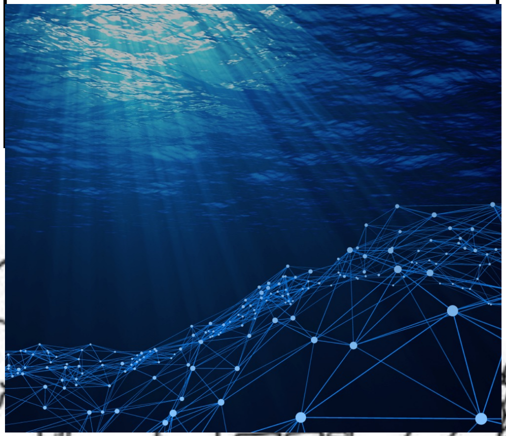
Observing system



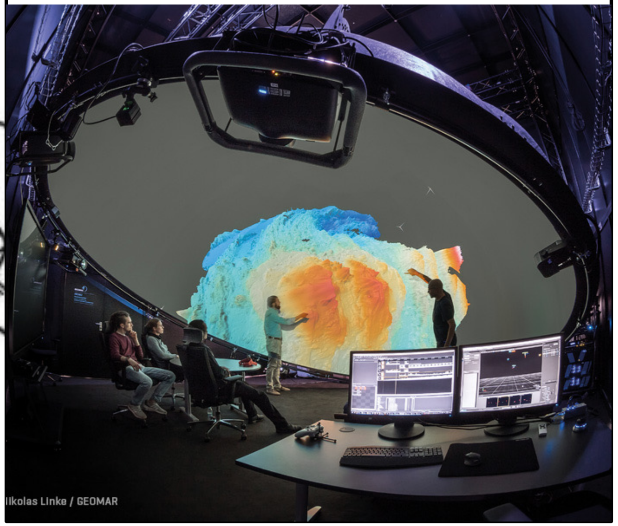
Data space

Data analytics

Prediction engine



Data & decision cyberinfrastructure



(Parable of the Blind Men and an Elephant)

2021
2030

DITTO
Digital Twins of the Ocean

Role of Digital Twins – An engineering perspective

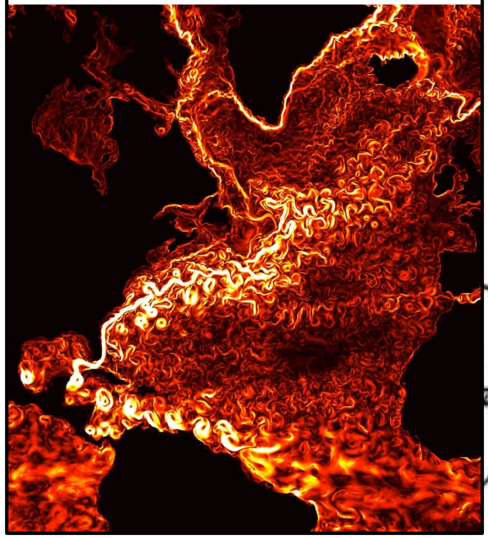
Role of Digital Twins – An engineering perspective

*A DT is an **in-silico replica/model** of a system subject to requirements:*

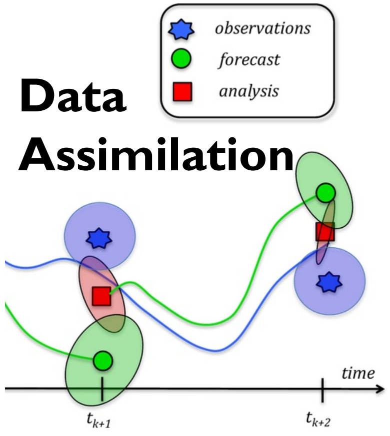
1. must continuously improve as it ***integrates new data*** and provides a ***dynamic digital history*** of the asset or entity
2. must be able to ***issue predictions*** about yet-unseen conditions and future states ***with quantified uncertainties***
3. must be able to support – with confidence – assessments of ***WHAT-IF scenarios*** that support critical decisions
4. must entail ***synergistic two-way coupling*** between the physical system, the ***data collection, and the user/social system***

What are Digital Twins? – From simulation to decision

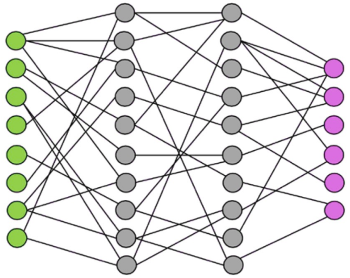
high-fidelity models



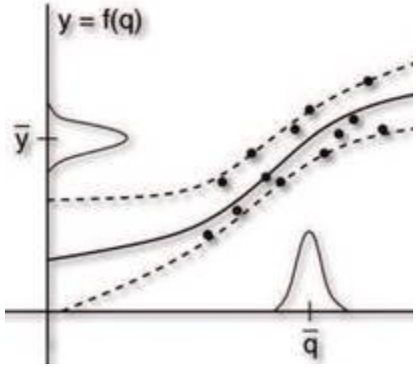
Data Assimilation



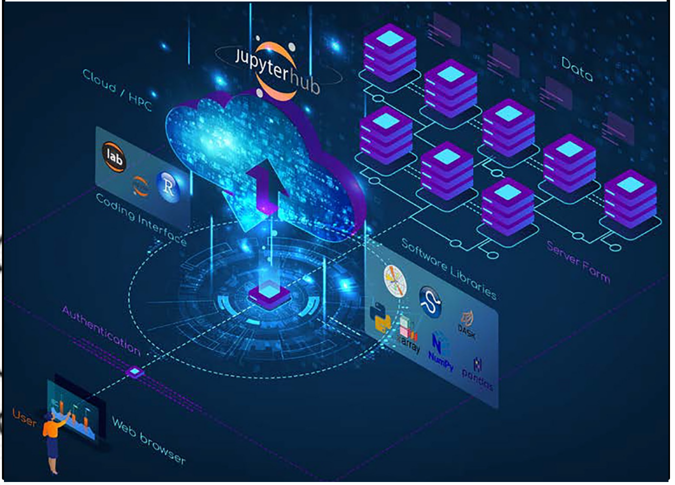
fast surrogate, reduced-order, ML models

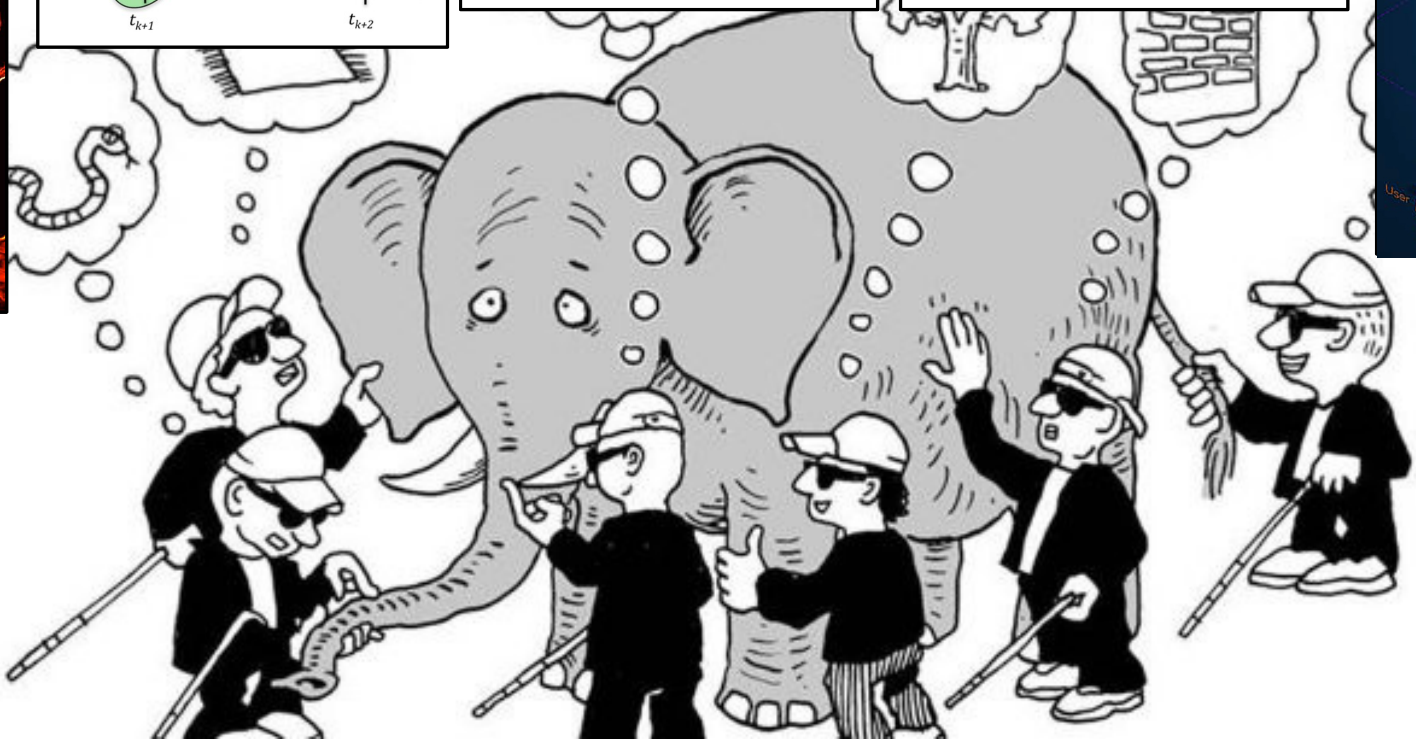


Uncertainty Quantification



Data cyberinfrastructure & decision support





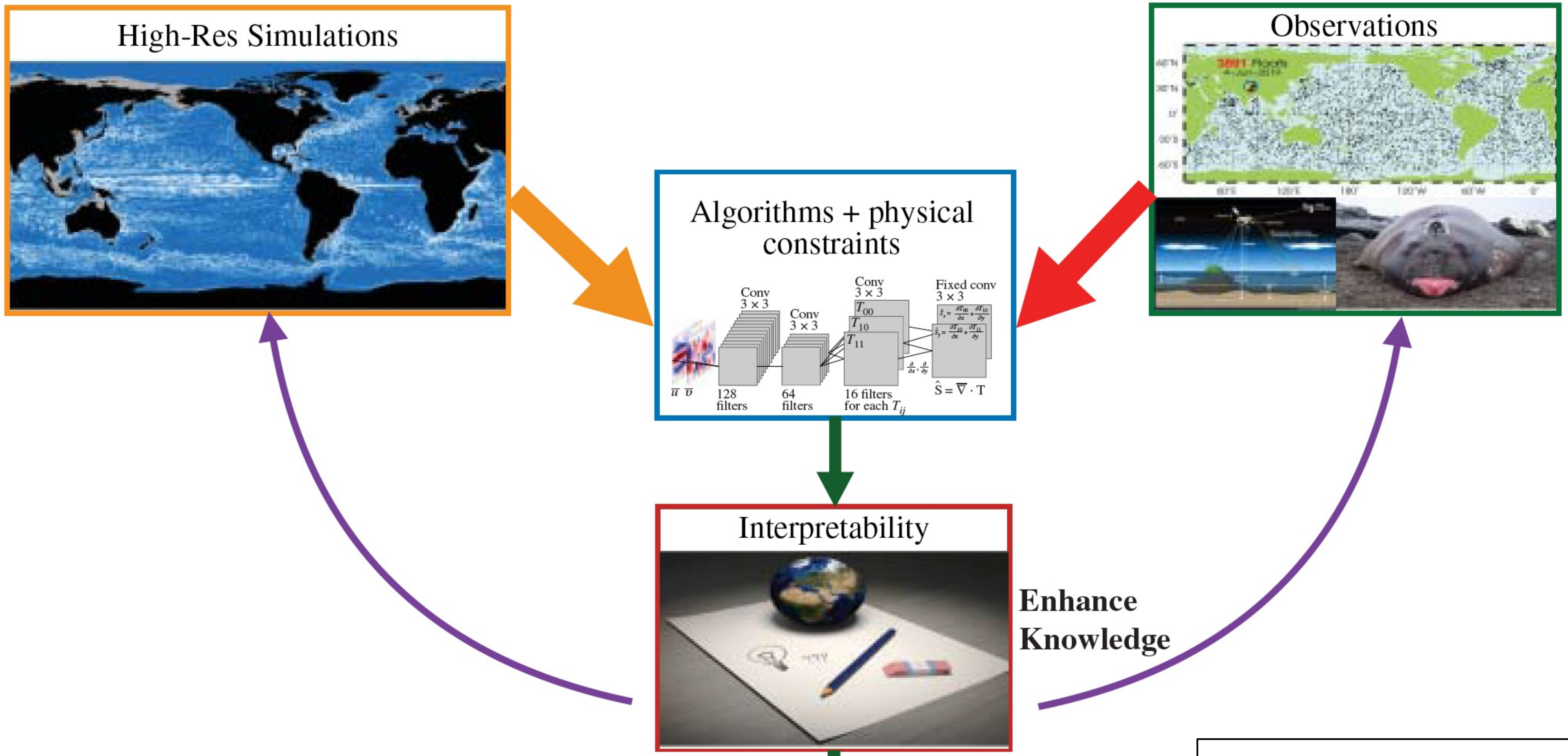
(Parable of the Blind Men and an Elephant)

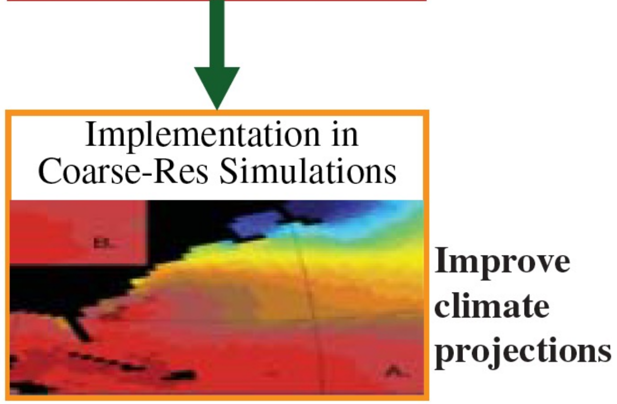
Role of Machine Learning to support Digital Twins

1. **Classification / anomaly detection**
2. **Regression (parameter calibration & state estimation)**
3. **Space- and/or time-dependent state prediction**
4. **Autonomous systems & active sampling**
5. **Emulation for uncertainty quantification**

Will (subjectively!) focus on 2. in the following

Physics-aware interpretable data-driven parameterizations

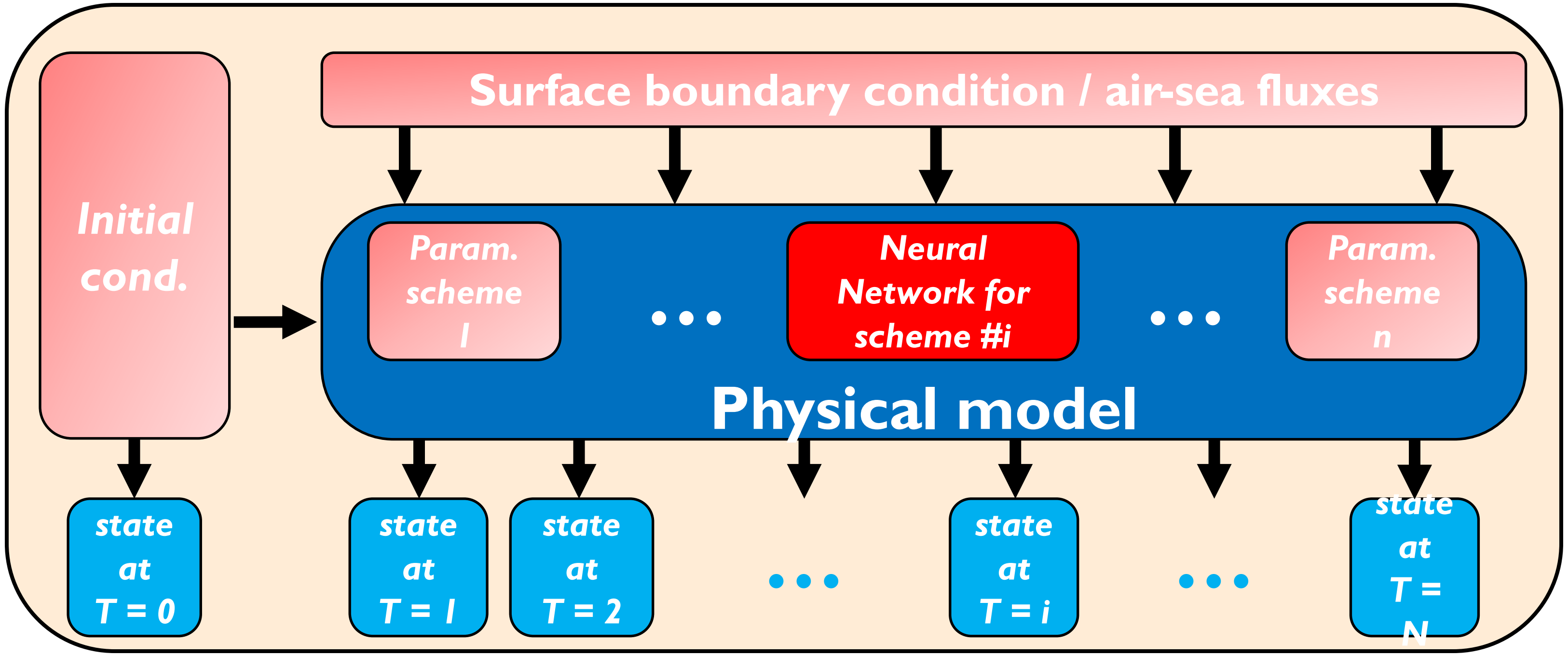




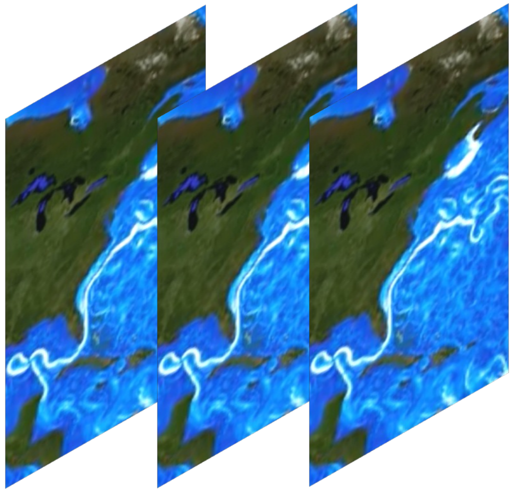
Zanna & Bolton (2021):
Deep Learning of Unresolved Turbulent Ocean Processes in Climate Models.

in: Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences

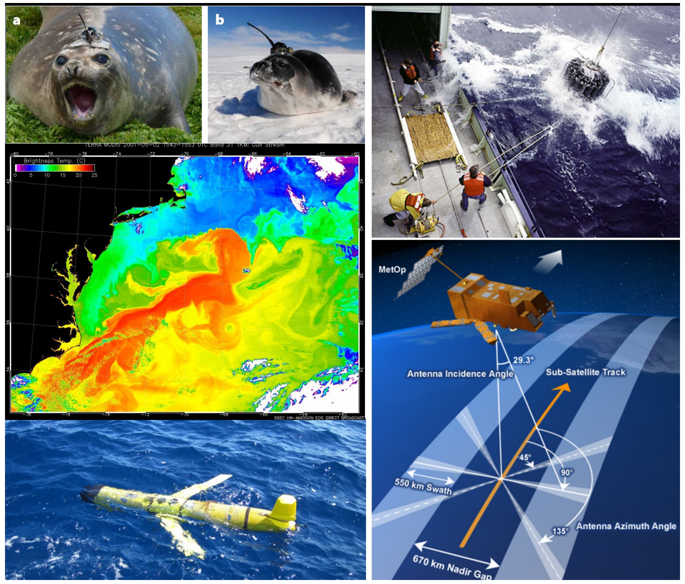
Seamless integration of scientific machine learning and inverse modeling



Training of the NN is part of "training" of the physical model on state variables

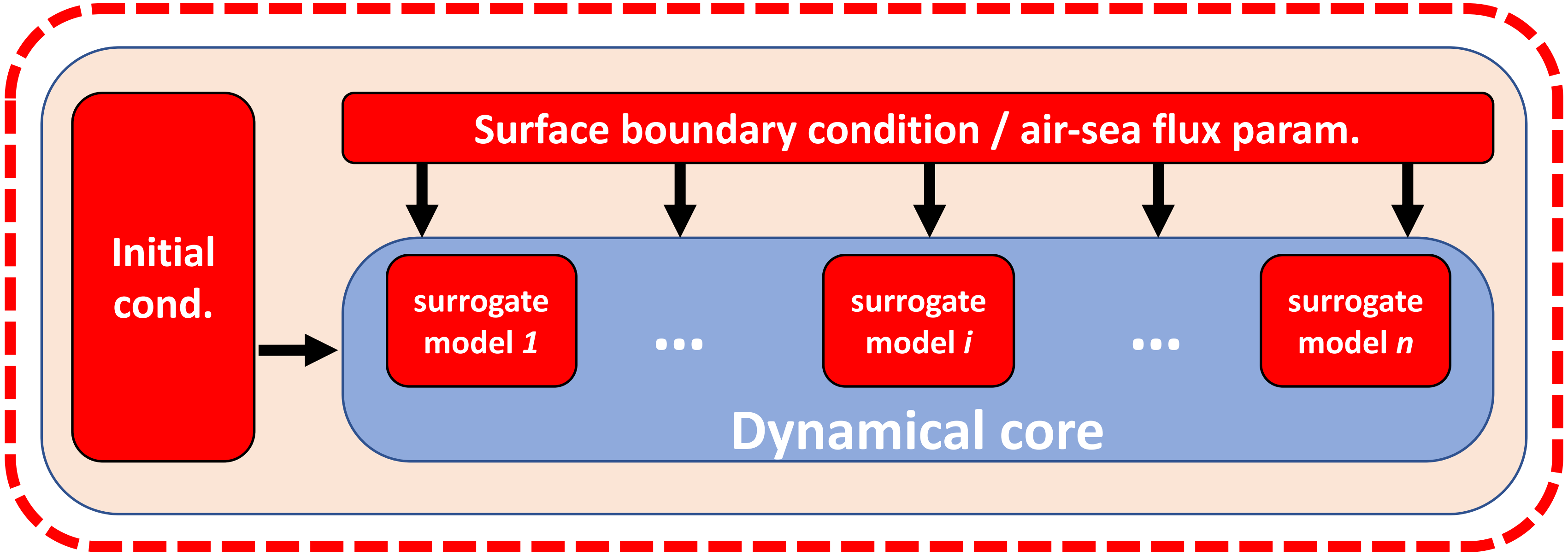


+



a posteriori / full-model / online / end-to-end learning

Differentiable programming for full-model / end-to-end learning with quantified uncertainty



Here: use of full-model ***differentiable programming*** to

- replace parts (or all?) of model by appropriate surrogates
- use all available observations to train/calibrate all uncertain variables
- combines inverse modeling and ML in *end-to-end learning*

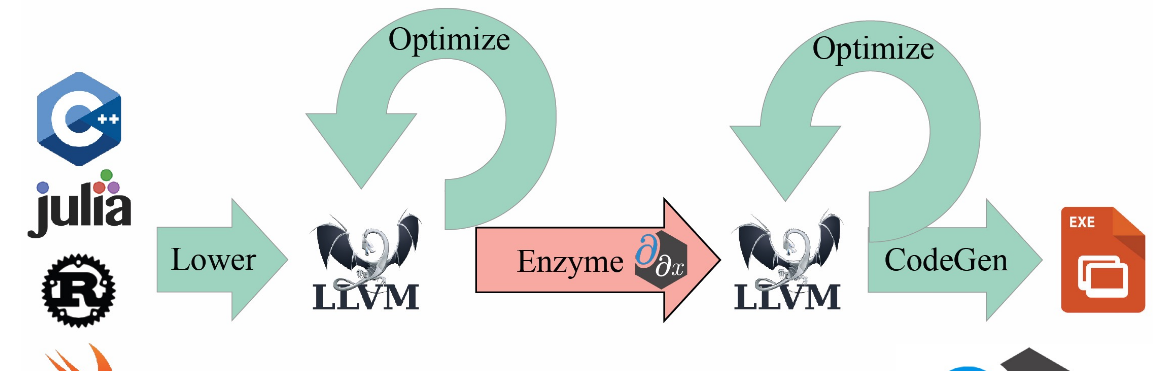
relies on general-purpose automatic differentiation (AD)

Differentiable programming for full-model / end-to-end learning with quantified uncertainty

Differentiating a GPU-enabled ocean model in Julia via the AD tool ***Enzyme.jl***

requires close collaboration between …

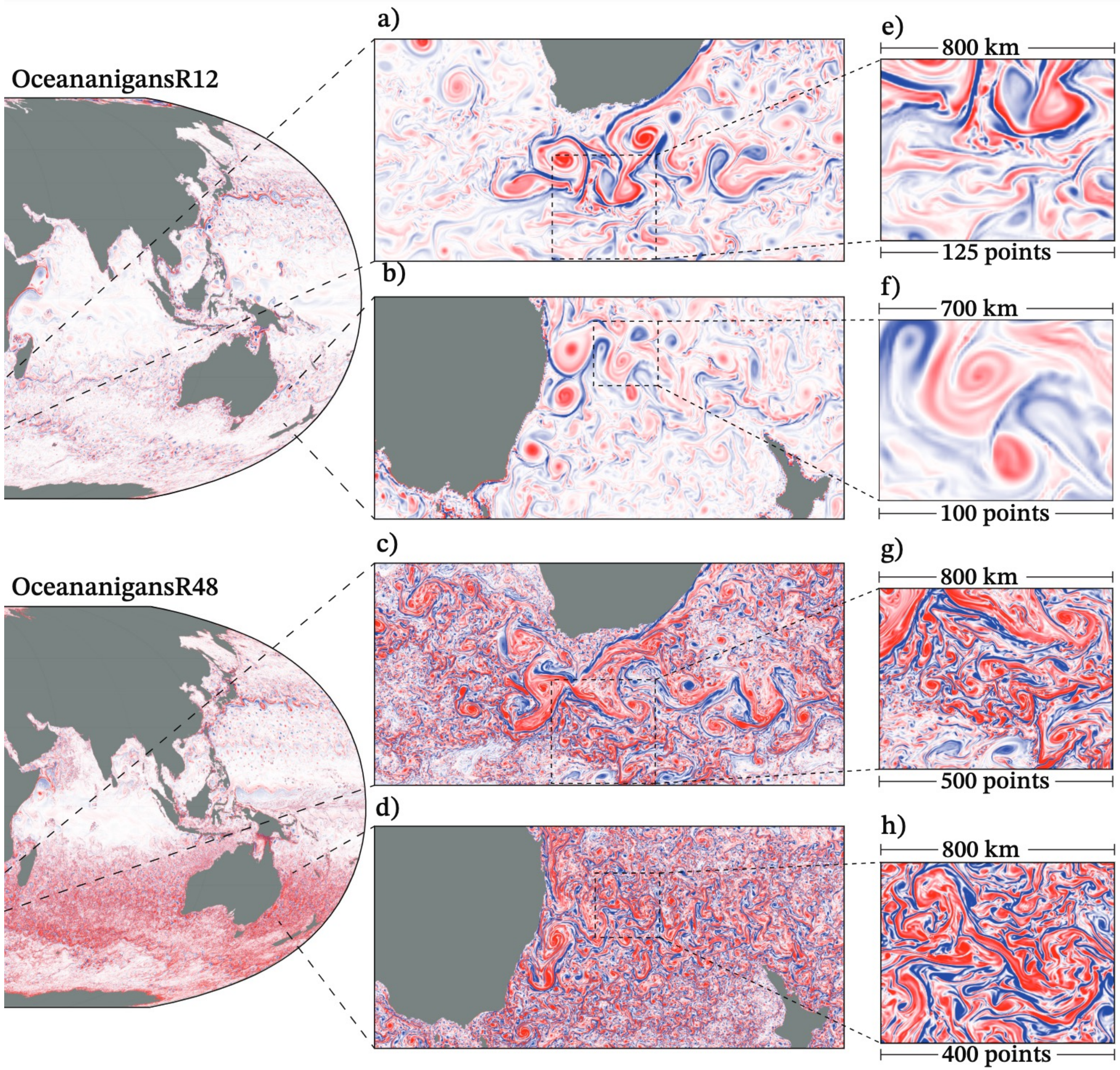
- **<u>computational science</u> (algorithms)**
- **<u>computer science</u> (compilers & compute architectures)**
- **<u>domain science</u> (model application)**



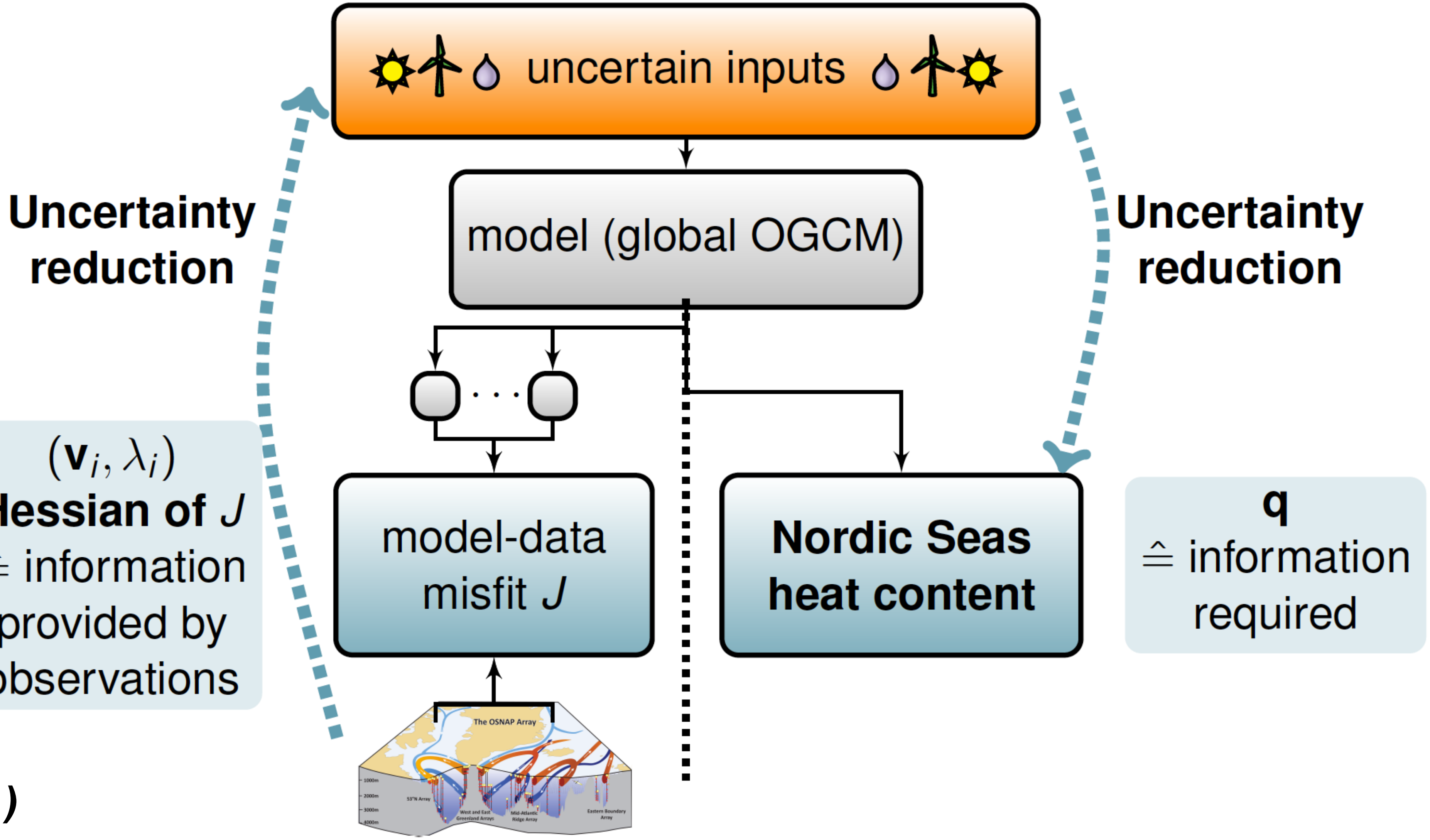
Moses, Churavy, et al., SC'21

DJ4Earth

Oceananigans.jl (from Silvestri et al., arXiv, 2023)



Role of Differentiable Programming
Targeted Sensing / Optimal Experimental Design



Loose et al. (JGR, 2020)

Loose et al. (JAMES, 2021)

Ghattas & Willcox (Acta Numerica, 2021)

OSNAP constraints on Nordic Seas heat content

= **uncertainty reduction** along

$$= \sum_i \frac{\lambda_i}{\lambda_i + 1} (\mathbf{v}_i \bullet \mathbf{q})^2$$

Making (scientific) data more usable and used for ML & DTs
A Scientific Data Commons (Courtesy Ryan Abernathey)

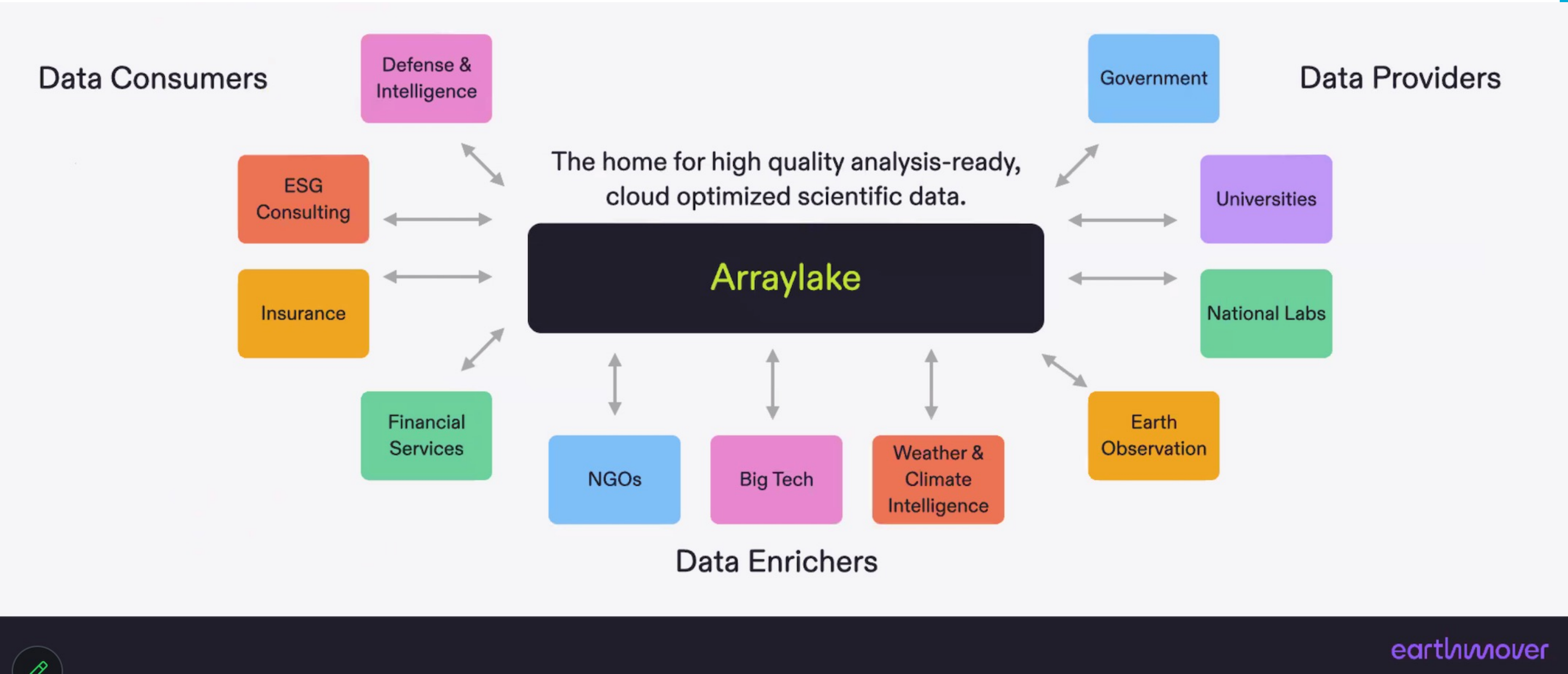
PROBLEM

Cloud Data Platforms Don’t Understand Scientific Data

Applications	**Business Data** Customers, products, orders, sales, etc.	**Science and Deep Learning** Sensor outputs and simulations. Geospatial, Weather, Climate, Biotech, Medical, Energy, Agriculture, Manufacturing. DL weight tensors and features.
Data Model	Tables	Arrays (aka “Tensors”)
Analysis APIs	SQL, pandas, DuckDB	NumPy, PyTorch, xarray
Storage Formats	Parquet, ICEBERG	HDF, netCDF, Zarr
Data Lakes / Warehouses	snowflake, Google BigQuery, databricks	🤷
Low-Level Object Storage	AWS S3, Google Cloud Storage, Azure Blob Storage	

earthmover

Making (scientific) data more usable and used for ML & DTs
A Scientific Data Commons (Courtesy Ryan Abernathey)



DOMAIN KNOWLEDGE

PREDICTIVE PHYSICS-BASED MODELING & SIMULATION

HUMAN-COMPUTER INTERACTIONS

OPTIMIZATION & CONTROL

HIGH-PERFORMANCE COMPUTING

EDGE COMPUTING

DATA ASSIMILATION

SURROGATE MODELING

UNCERTAINTY QUANTIFICATION

ARTIFICIAL INTELLIGENCE

MACHINE LEARNING

Digital Twins

Next-generation digital tools that move **Beyond Forward Simulation**

Courtesy Karen Willcox

(NASEM Study on Foundational Research Gaps and Future Directions for Digital Twins)