OcéanIA
AI, Oceans and Climate Change

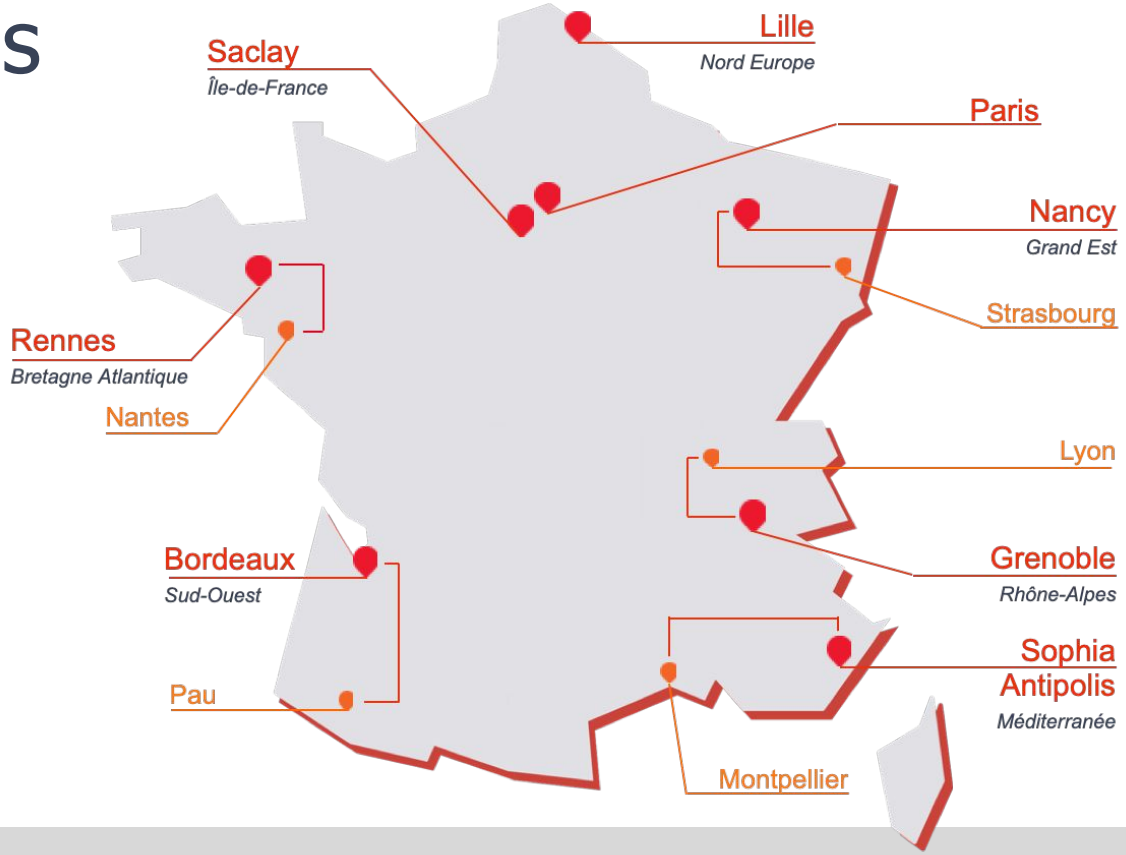
Nayat Sánchez Pi
Director Inria Research Center in Chile

NEURAL INFORMATION PROCESSING SYSTEMS

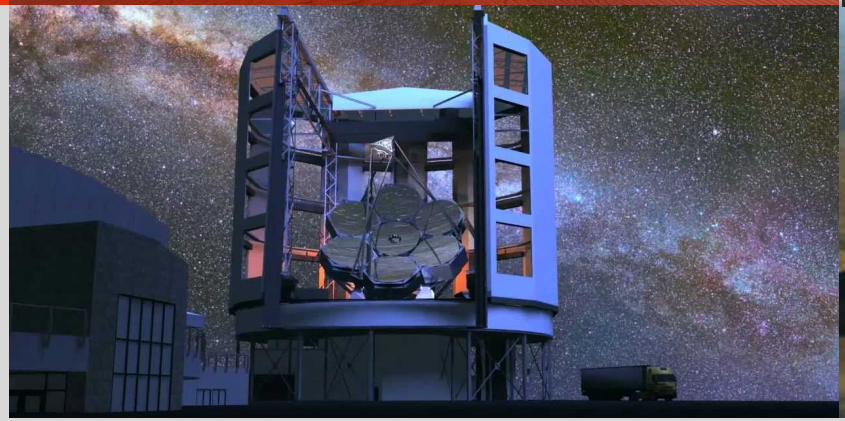
LXAI

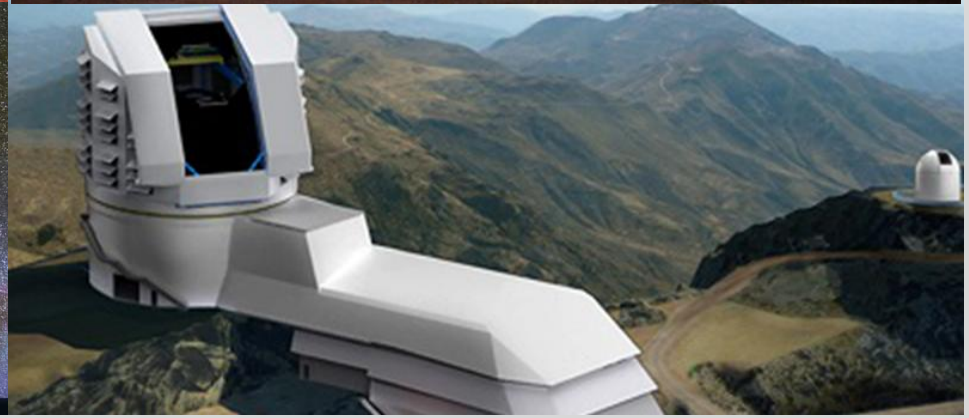
December 7th, 2020

Inria Centres



Diversity in Natural Resources: Atacama desert

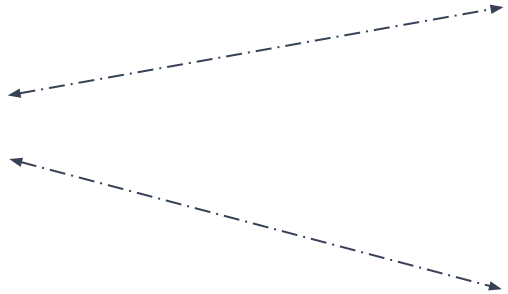




Diversity in Natural Resources: Antarctica

Challenging enough to AI/ML

AI/ML



- ➔ Small, big and/or high dimensional data
- ➔ Transfer learning
- ➔ Causal Learning
- ➔ Active Learning
- ➔ Explainable AI

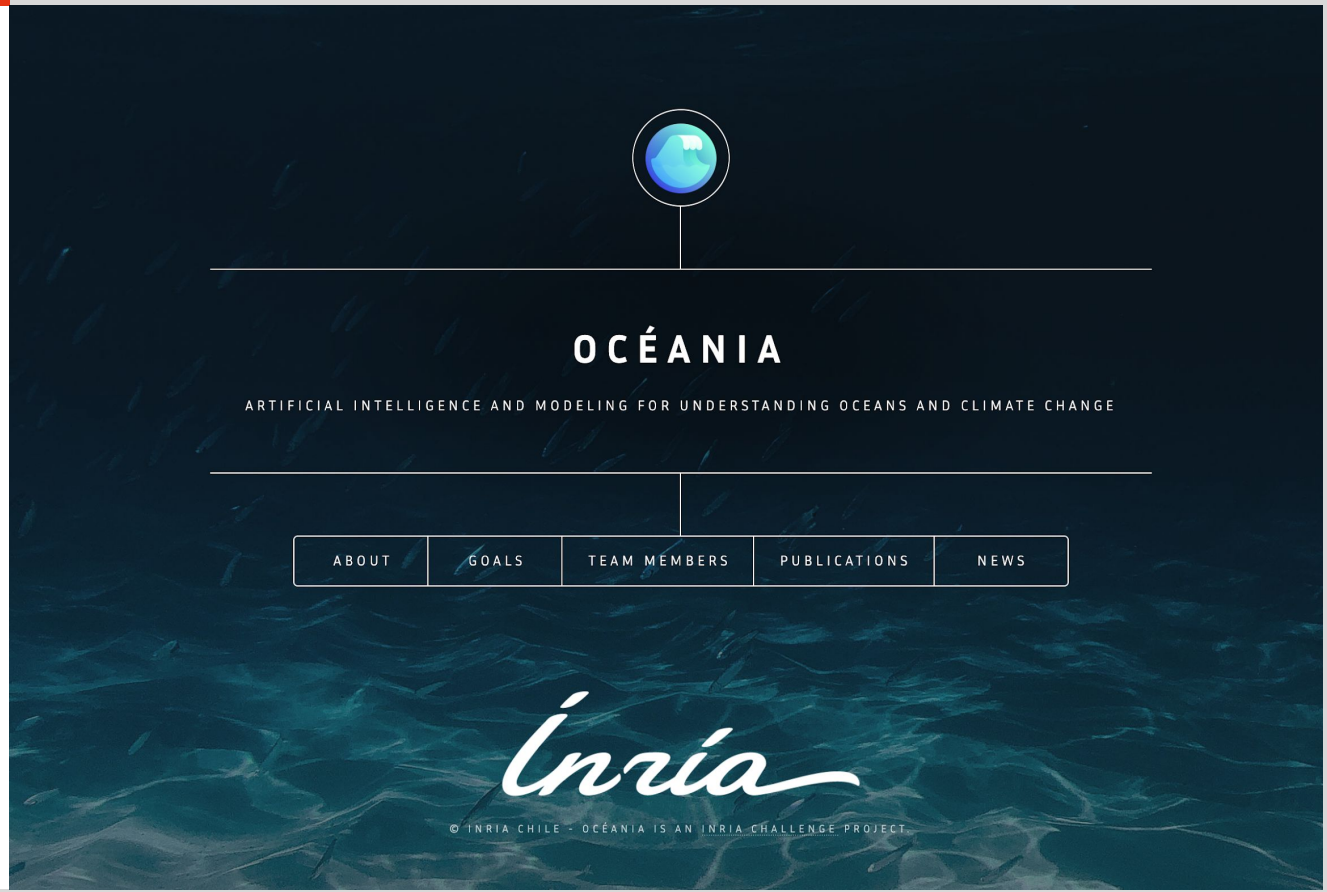
- ➔ Inherent complex context
- ➔ Convergence of physical and life sciences through Big and Data

Inria Challenge

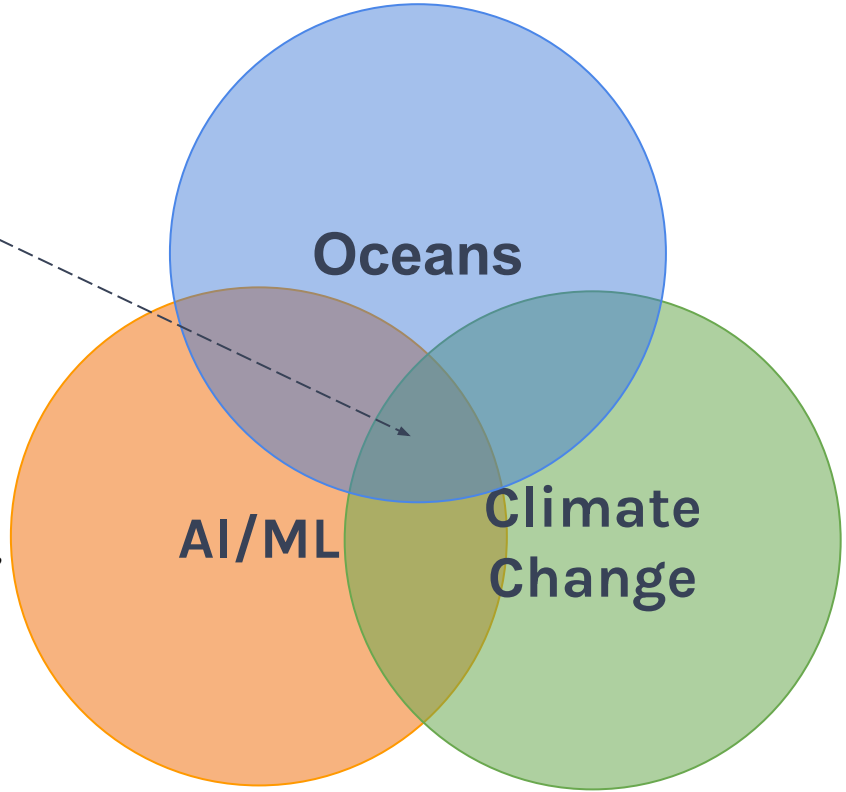
OcéanIA:

Artificial Intelligence and Modelling for Understanding Climate Change

https://oceania.inria.cl/



#OcéanIA



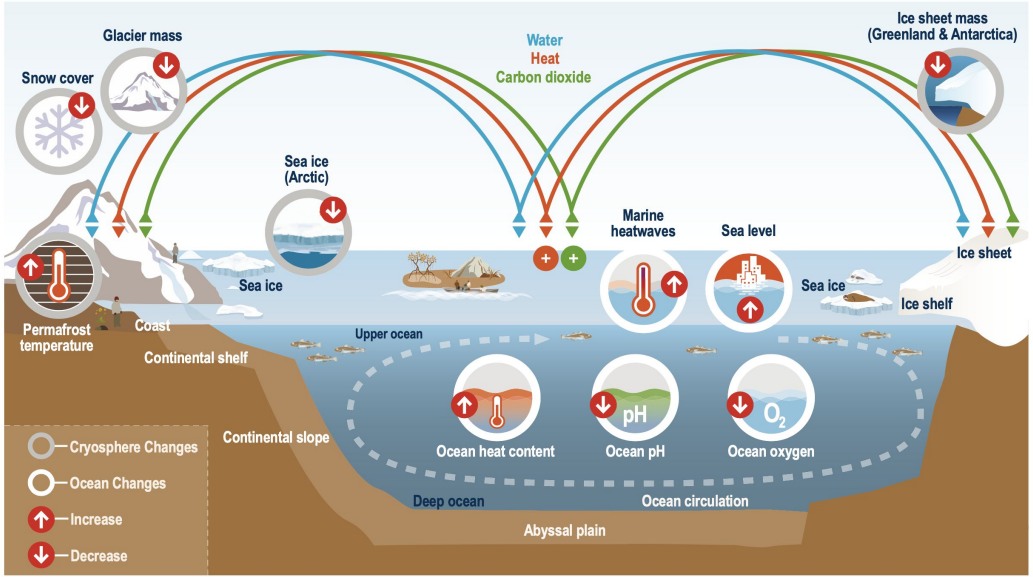
ENVIRONMENT

Nature has provided a lot of inspiration for the AI/ML area:

- ↳ neural networks,
- ↳ learning theories,
- ↳ evolutionary computing, etc.

It is about time that we return the favor!

Climate Change and the Ocean



Source: H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, E. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)]. In: *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* [H.- O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.).

Environment

- **TARA ARCTIC (2006-2008) :** First Arctic drift since Nansen (1893)
- **TARA OCEANS (2009-2013) :** First global study of the planktonic ecosystem
- **TARA MEDITERRANEAN (2014) :** Study of the impact of plastic on the marine ecosystem in the Mediterranean
- **TARA PACIFIC (2016-2018) :** Study of the adaptive capacity of coral reefs to climate change
- **MICROPLASTICS (2019) :** First study of river sources of microplastics on a European scale

11 missions including 5 major expeditions
100,000 microscopic marine species discovered
Over 150 million marine genes discovered
Almost 200,000 viruses characterized
Over 150,000 samples collected

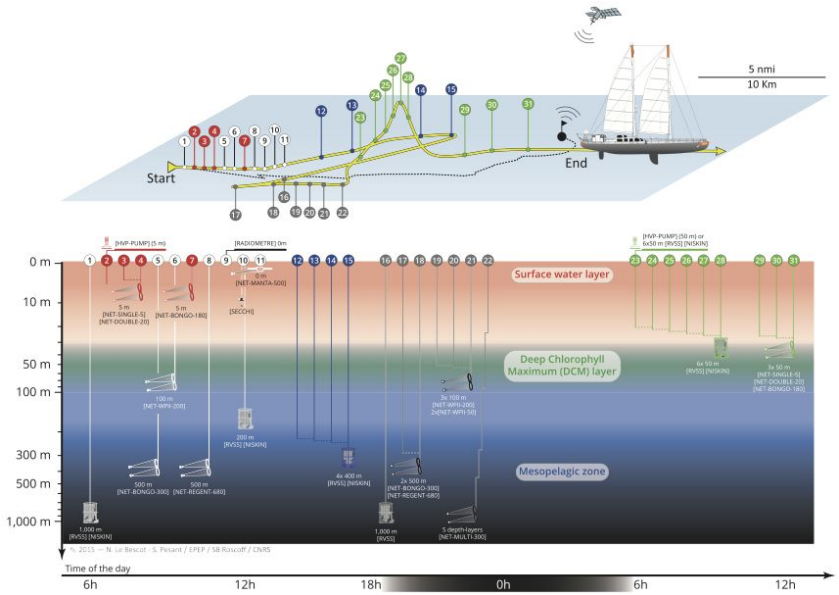
TARA OCEANS

Expedition
Feb - May 2021

Mission Microbiome
-CEODOS

Understand the ocean environment, climate change, its impact and mitigation.

Why Tara?



Source: Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Trouble, R., Dimier, C., & Searson, S. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, 2(Lmd), 1–16. https://doi.org/10.1038/sdata.2015.23

Schooner Tara:

- ↳ >40,000 plankton samples collected
- ↳ 210 sampling stations at three depths
- ↳ >60 terabases of DNA and RNA sequenced
- ↳ ~7 million images captured

➜ Tara Oceans adopts the principle of **open access** and **early release** of raw and validated data sets.

➜ Solid, diverse and international scientific community.

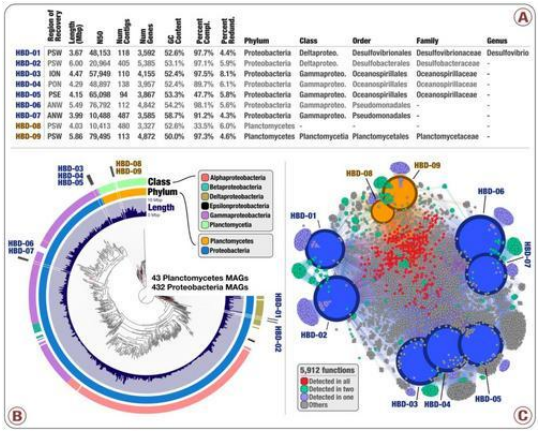
OcéanIA

An opportunity and a challenge to state-of-the-art AI/ML

CHALLENGES

- Small & high dimensional data
- Data Engineering & HPC
- Transfer Learning
- Multiobjective decision making
- Data & AI Governance
- Interpretability Explainable AI
- Causal inference
- Structure and graph-based NN
- Energy-aware Green AI/ML

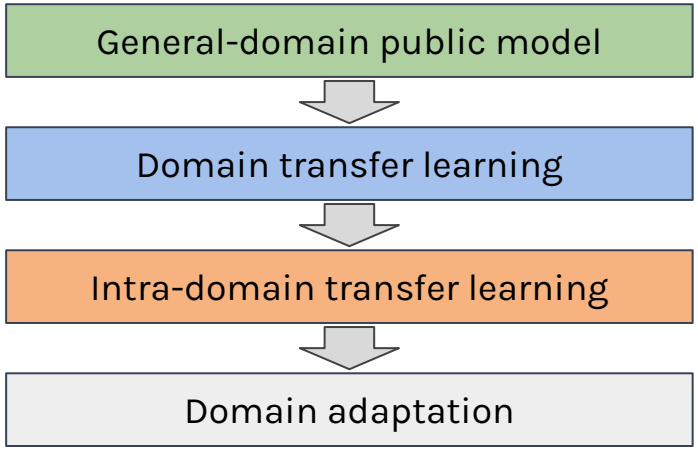
Small but high-dimensional heterogeneous data



- ➜ Tara expeditions gather lots of data - from a marine biology point of view.
- ➜ Highly heterogeneous: DNA barcodes, images, environment variables.
- ➜ Samples includes many species at the same time.
- ➜ Grouped by sample location, interested in networks and graph-based information.
- ➜ In spite of efforts it is not always consistent and it is always evolving: i.e. new hardware.
- ➜ Data from Tara allows exploration of the relationship between marine ecosystem functioning and biodiversity.

Source: Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T., Rappé, M. S., MacLellan, S. L., Lücker, S., & Eren, A. M. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nature Microbiology, 3(7), 804–813. https://doi.org/10.1038/s41564-018-0176-9

Model reuse, transfer learning and domain adaptation

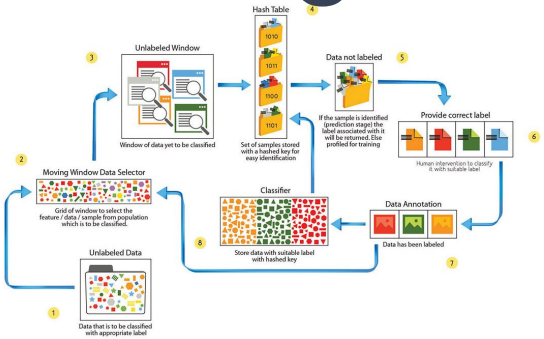


A three-step work hypothesis:

- ↳ Is it possible to adapt existing (computer vision or graph) models to out domain?
- ↳ How to transfer a domain model to other biomes, locations or across species?
- ↳ How to cope with variations across species and sensing hardware?

- **Transfer learning** addresses the issue of how to adapt and re-purpose the internal representations of a model that has been trained on a similar problem.
- **Domain adaptation** is the capacity to cope with changes in the environment because of the evolution of the system and/or the need to particularize a general model to a particular instance.

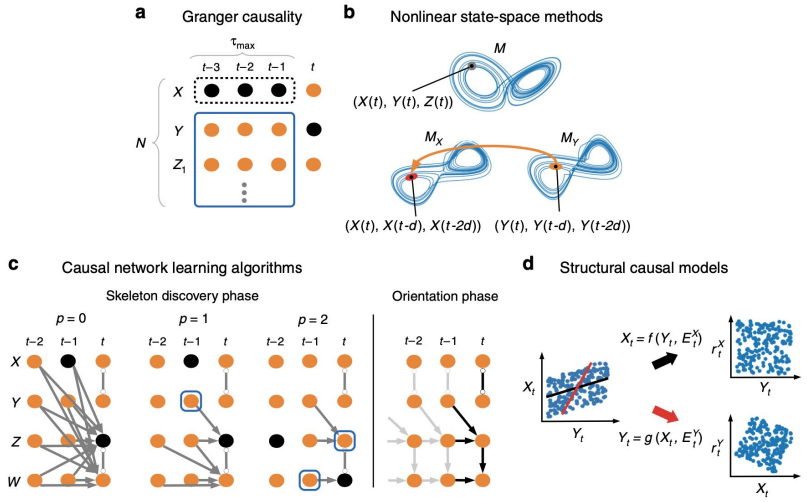
Active, few-shot and multi-task learning



Source: https://brandidea.ai/activeLearning.html

- ➜ Limited data and/or high uncertainty,
- ➜ Direct sampling to the areas of the domain where they are most necessary.
- ➜ Guiding sampling using **active learning.**
- ➜ **Few-shot learning** methods to produce actionable products with minimal data.
- ➜ **multi-source** or **multi-task learning** ensembles training signals of related tasks.
  - ↳ enables the model to generalize better on the main task.
  - ↳ effectively increases the sample size that is being used for training.
  - ↳ biases the model to prefer representations that are useful for other tasks -> transfer learning!

Causality and Explainable AI

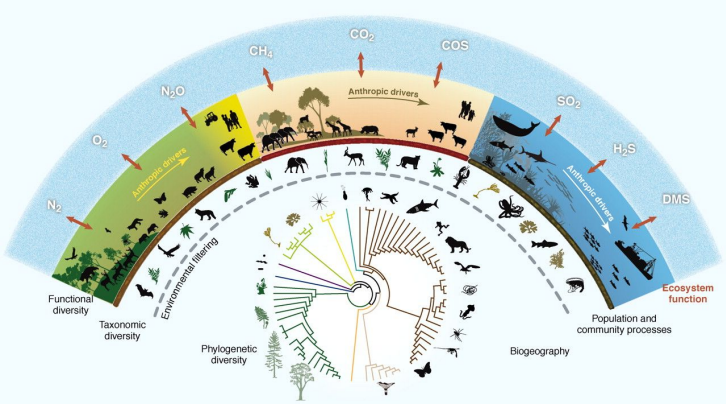


Source: Runge, J., Bathiany, S., Bollt, E. et al. Inferring causation from time series in Earth system sciences. Nat Commun 10, 2553 (2019). https://doi.org/10.1038/s41467-019-10105-3.

- ➜ Produce explainable models, while maintaining performance (prediction accuracy),
- ➜ research support tools that combine explainability and causality for new scientific discoveries and theories by making surrogate human-readable models, and
- ➜ enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.

Essential for policy making.

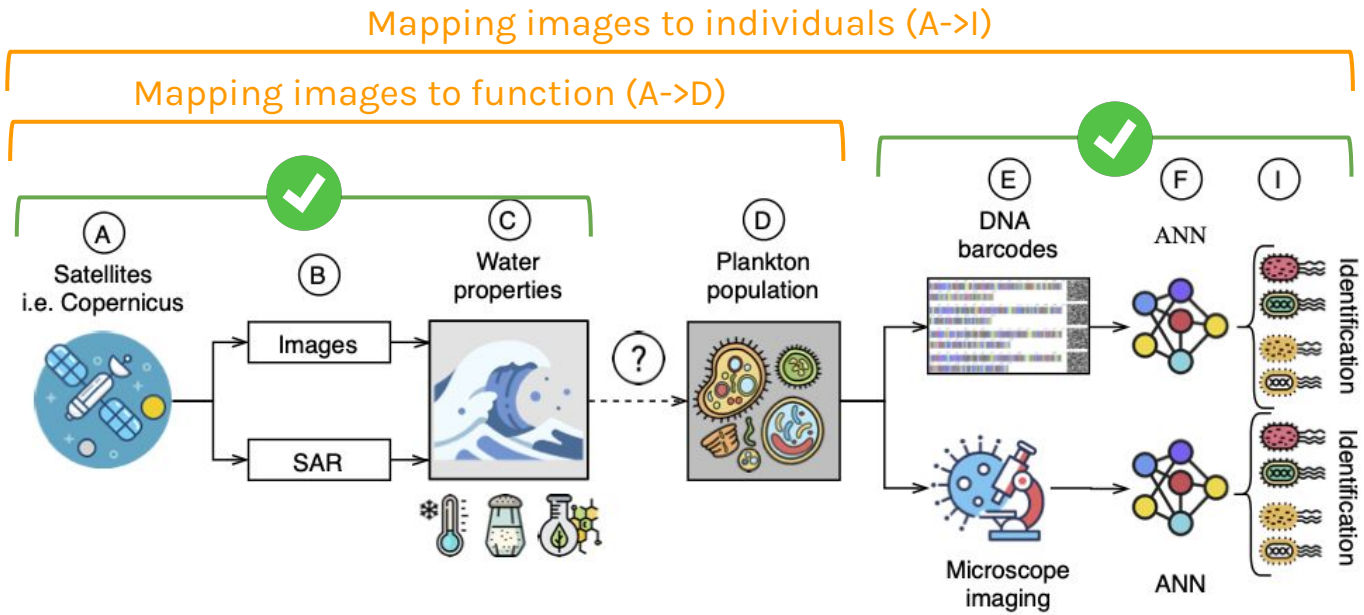
Biodiversity and ecosystem functioning



Source: Naeem, S., Duffy, J. E., & Zavaleta, E. (2012). The Functions of Biological Diversity in an Age of Extinction. *Science*, 336(6087), 1401 LP – 1406. https://doi.org/10.1126/science.1215855

- ➜ Biodiversity supports functions like primary productivity and carbon fixation and sequestration, etc.
- ➜ Understanding this is fundamental: science and policy making.
- ➜ Data from Tara allows exploration of the relationship between marine ecosystem functioning and biodiversity.
- ➜ How variations on biodiversity impact those functions?
- ➜ How changes in temperature (or other variables) impact biodiversity and functions?
- ➜ Understand causality and circular causality among different levels of biodiversity, ecosystem functioning.

Understanding plankton communities using AI, ML, and vision



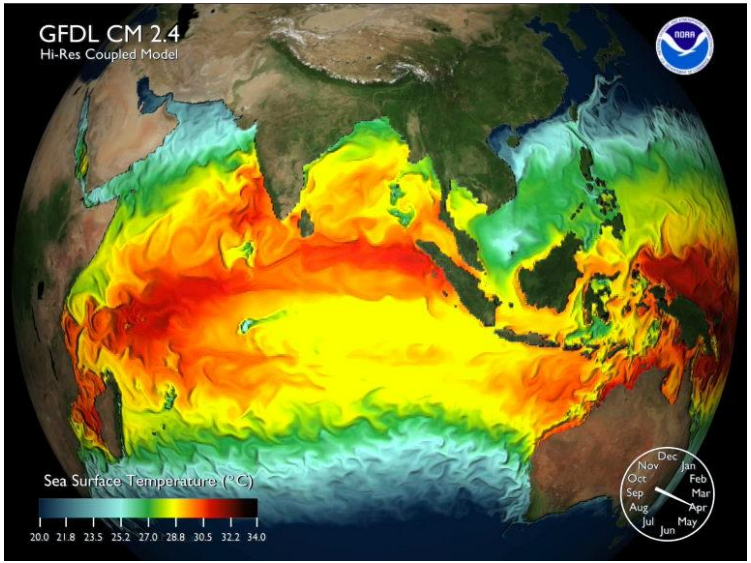
Anomaly detection and explainable AI for automatic plankton ID

Source: Wikipedia

- ➜ Identifying plankton as a supervised problem: already addressed.
- ➜ Estimated that more than 70,000 species unknown.
  - ↳ unsupervised or active learning approaches.
- ➜ Tara sampling includes high-res microscope.
- ➜ Identify **unknown** or **out of context** species.
- ➜ Why an organism represents an interesting specimen?
- ➜ Methods involved:
  - ↳ **transfer learning** and **domain adaptation**,
  - ↳ **(un/self)supervised object detection** and **segmentation**
  - ↳ **causal inference** to understand context.
  - ↳ **explainable AI**: i. e. hint what parts of the organism that determining its selection.

Field experts want a research tool not a black-box.

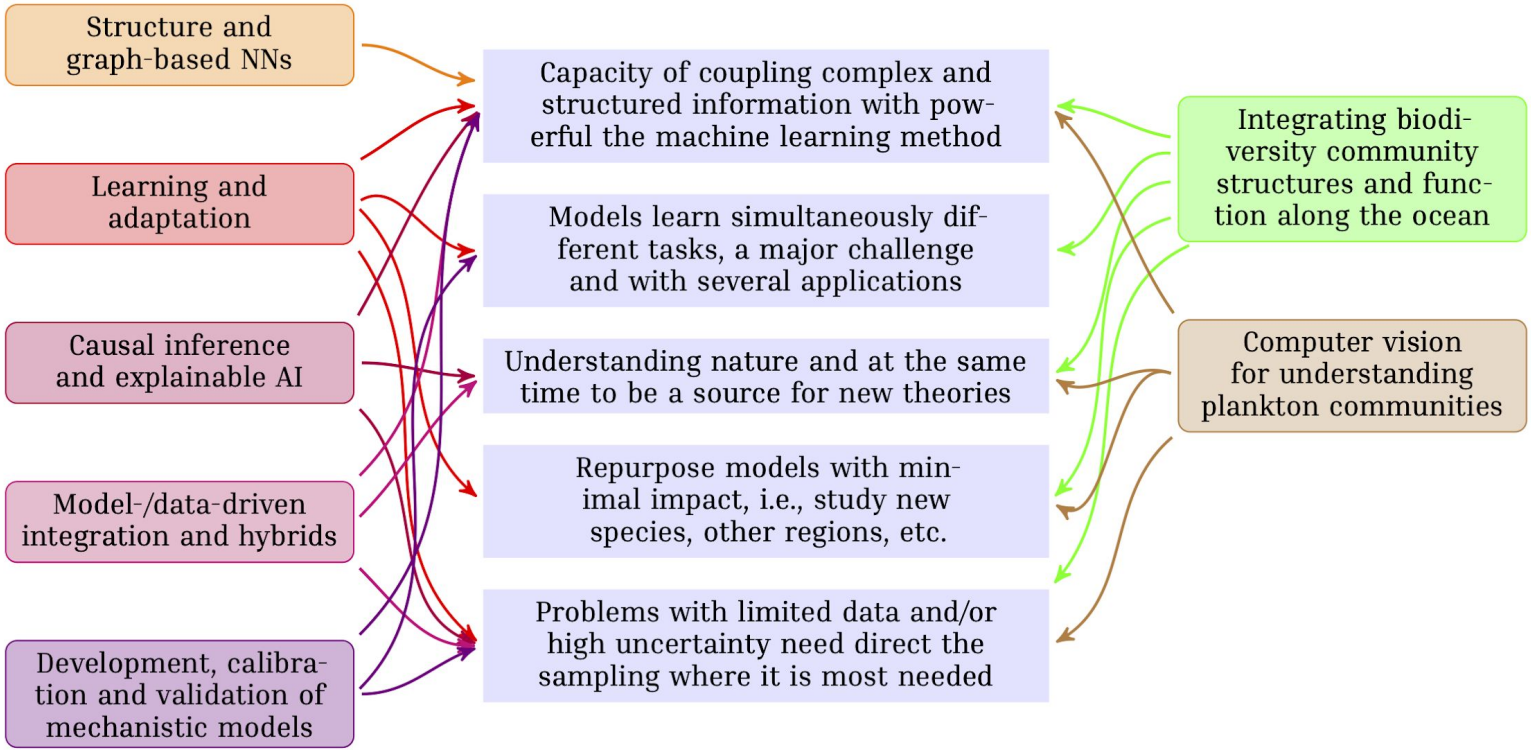
Integrating model-based (i.e. PDEs) and data-based (i.e. ML) approaches



Source: https://www.gfdl.noaa.gov/hires_indian_sst/

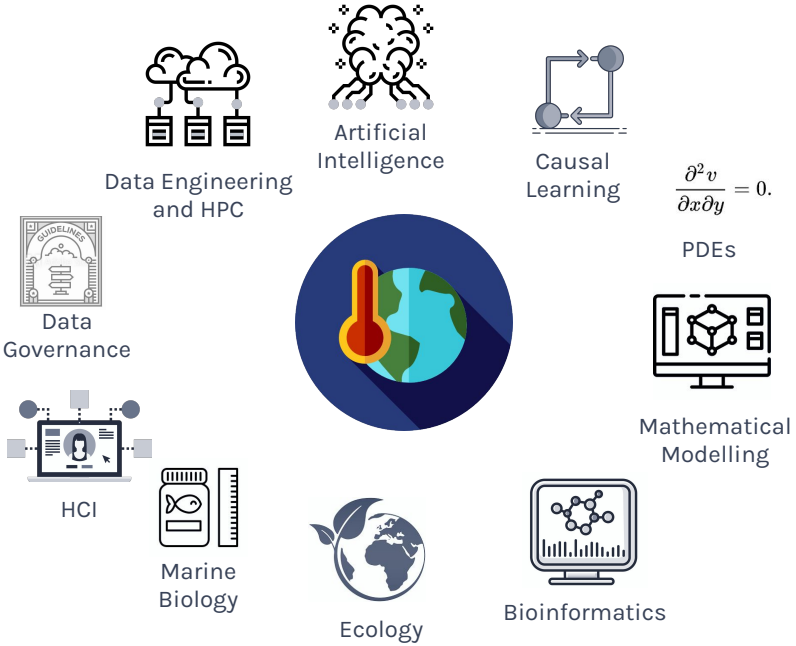
Ocean Circulation Models studying dynamics:

- ➜ high-res models are computationally inviable,
- ➜ current resolution of models is not sufficient,
- ➜ requiere large viscosity and diffusion coefficients smooths out features such as jets and mesoscale eddies.
- ➜ oceanic turbulence at small scales, which play an important role.
- ➜ Planned actions:
  - ↳ Learning PDEs from Data
  - ↳ Understanding learning dynamics
  - ↳ Hybrid models: Combining PDE solvers and DNNs (improved explicability).


Structure and graph-based NNs
Learning and adaptation
Causal inference and explainable AI
Model-/data-driven integration and hybrids
Development, calibration and validation of mechanistic models
Capacity of coupling complex and structured information with powerful the machine learning method
Models learn simultaneously different tasks, a major challenge and with several applications
Understanding nature and at the same time to be a source for new theories
Repurpose models with minimal impact, i.e., study new species, other regions, etc.
Problems with limited data and/or high uncertainty need direct the sampling where it is most needed
Integrating biodiversity community structures and function along the ocean
Computer vision for understanding plankton communities

The OcéanIA team

https://oceania.inria.cl/



Inria Challenge

LEADER

Inria Chile

France

Inria

ANGE TAU

BIOCORE

cnrs GO-SEE Federation

LS2N COMBI

Chile

Universidad de Chile

CMM Center for Mathematical Modeling

Pontificia Universidad Católica de Chile

Others

TARA OCEANS

Copernicus

Support

Cooperación Regional Francesa para América del Sur

Summary

- Oceans and climate change are intimately related:
  - Carbon capture, impact of change of temperatures, etc.
- Oceans are the last 'unknown':
  - Understanding the role of oceans in climate change is not only important but also challenging for modern AI, ML and applied math.
- A way to address current hot topics like explainability, bias, etc. on a different domain.
- Tara as source of data that can be crossed with other sources.
- I have described the approaches we are to follow in the next 4 years.
- OcéanIA has just started!
  - **We are hiring!**

Thank you! Merci !
Obrigada! ¡Gracias!

Follow us: www.inria.cl and oceania.inria.cl