



**W** UNIVERSITY *of* WASHINGTON



eScience Institute

## Interoperating ocean sonar data of heterogeneous sources using **echopype**

---

**Wu-Jung Lee** | Applied Physics Laboratory

**Valentina Staneva** | eScience Institute

**Kavin Nguyen** | Department of Physics

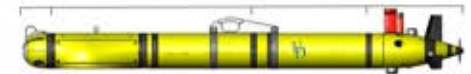


PICES 2019, Victoria, BC, Canada  
October 17, 2019



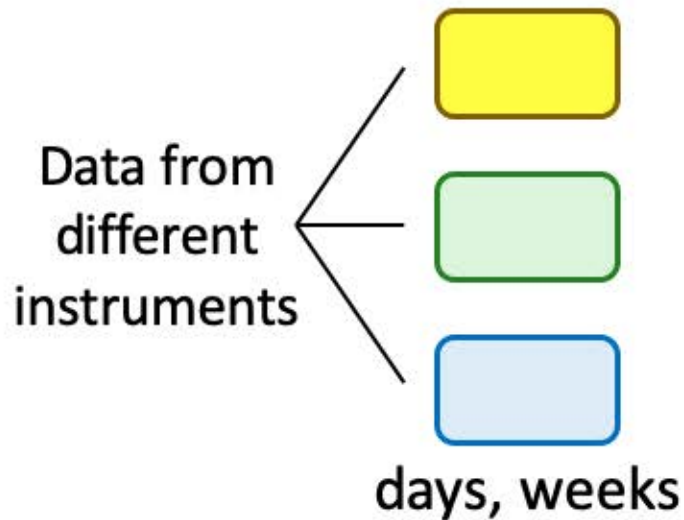
# The acoustic data deluge

- We have become very good at collecting data
- Continuous data collection on numerous platforms
- Excellent opportunities to study marine ecosystems at scales never before possible

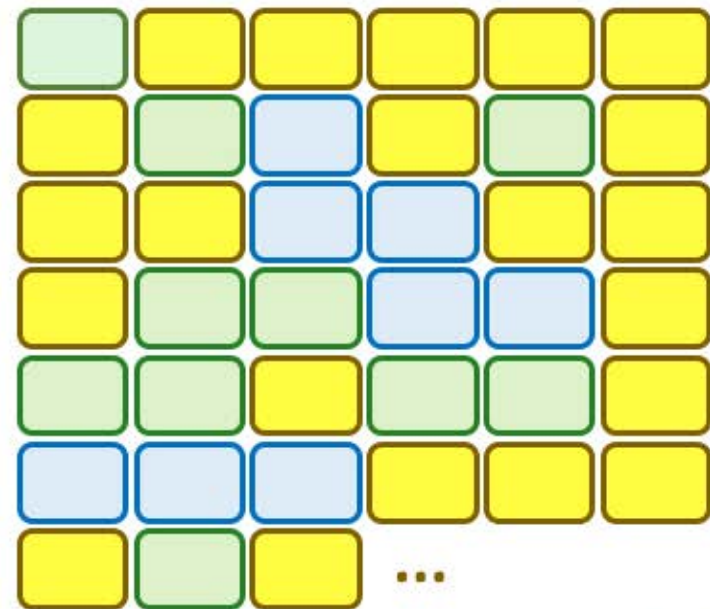


# These new data bring many new challenges

**It used to be...**



**But now...**



- Significantly increased data volume
- Heterogenous instrument sources



# Challenges of the Big Acoustic Data

- Interoperability

- Many manufacturers, many sonar models
- Proprietary software or open-source software written in proprietary languages (e.g., Matlab)



**SIMRAD**



**HIPS & SIPS**

**Sonar4**

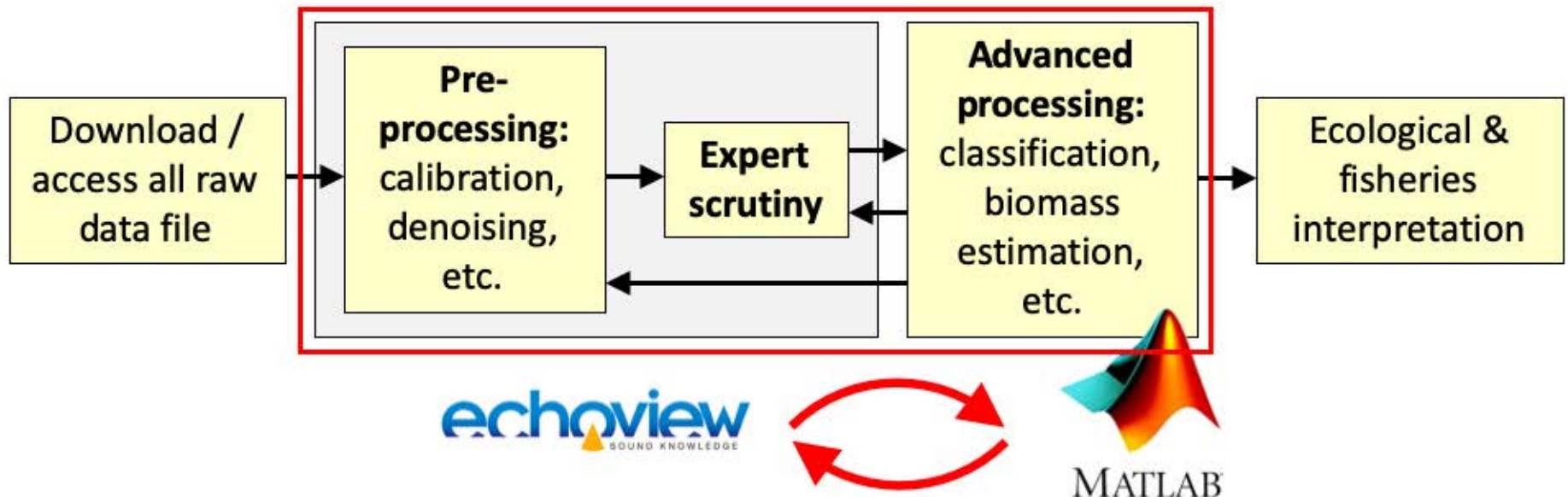
**Sonar5-Pro**

# Challenges of the Big Acoustic Data

## ■ Scalability

- Current analysis workflow is labor-intensive
- No support for parallel computation with random-access file formats

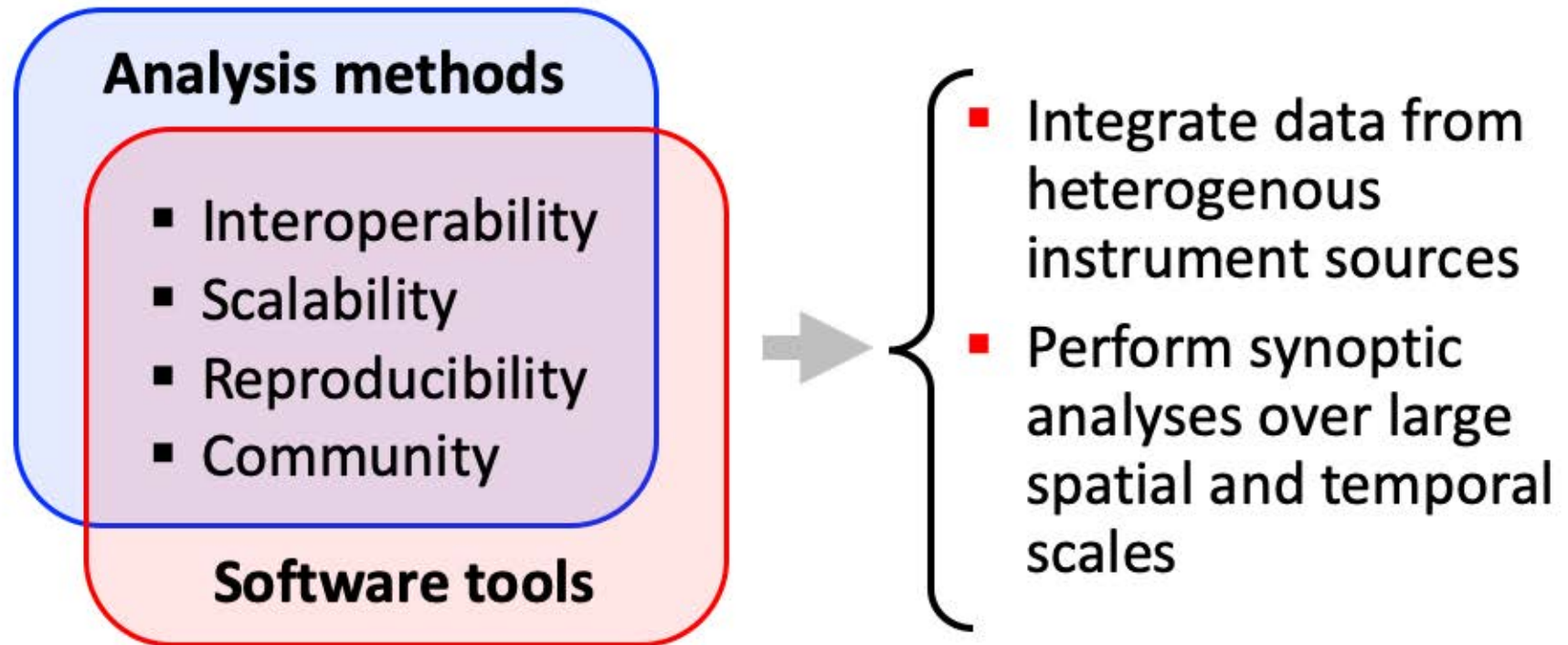
### A typical fisheries sonar data analysis pipeline



# Challenges of the Big Acoustic Data

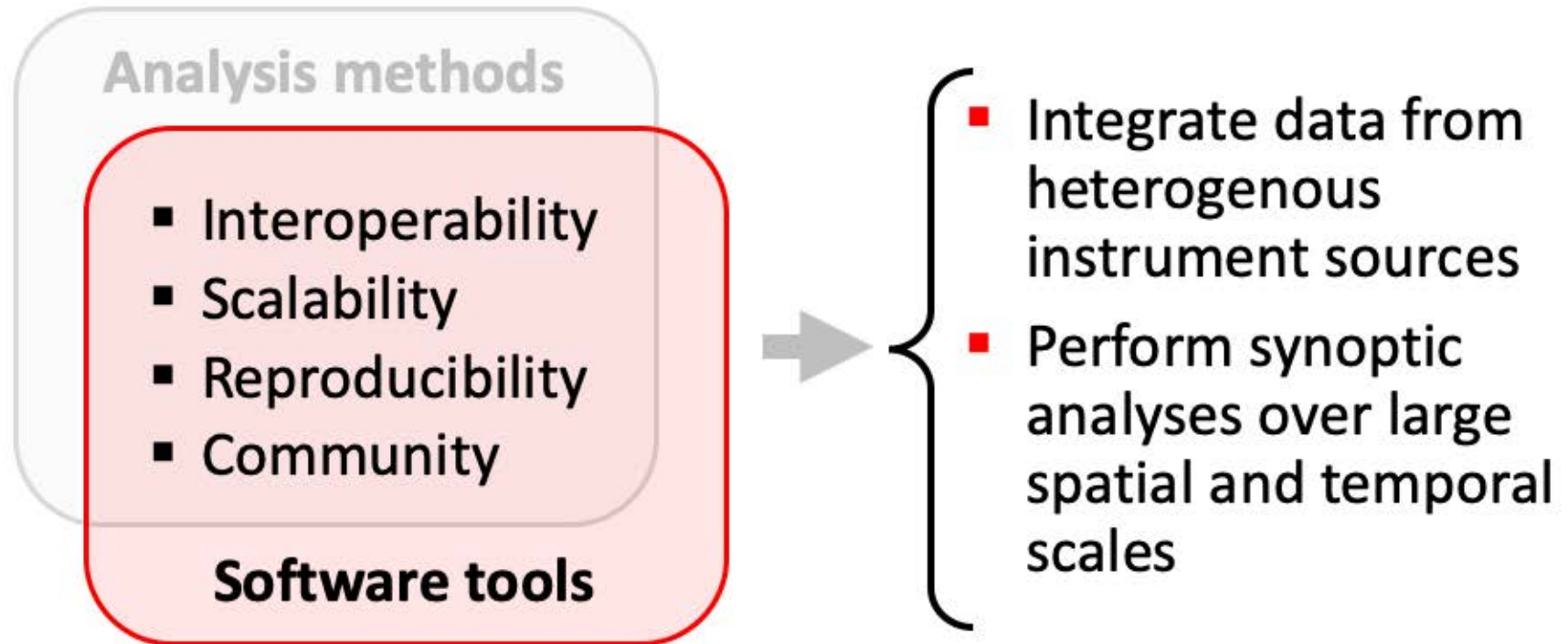
- Interoperability
- Scalability
- Reproducibility
  - Currently mostly GUI-based: good for exploration but hard to reproduce
- Community
  - Workflow transformation
  - Community-driven development
  - Data convention

# Vision



- Long-term goal: make ocean sonar data an integrated component of standard oceanographic data sets

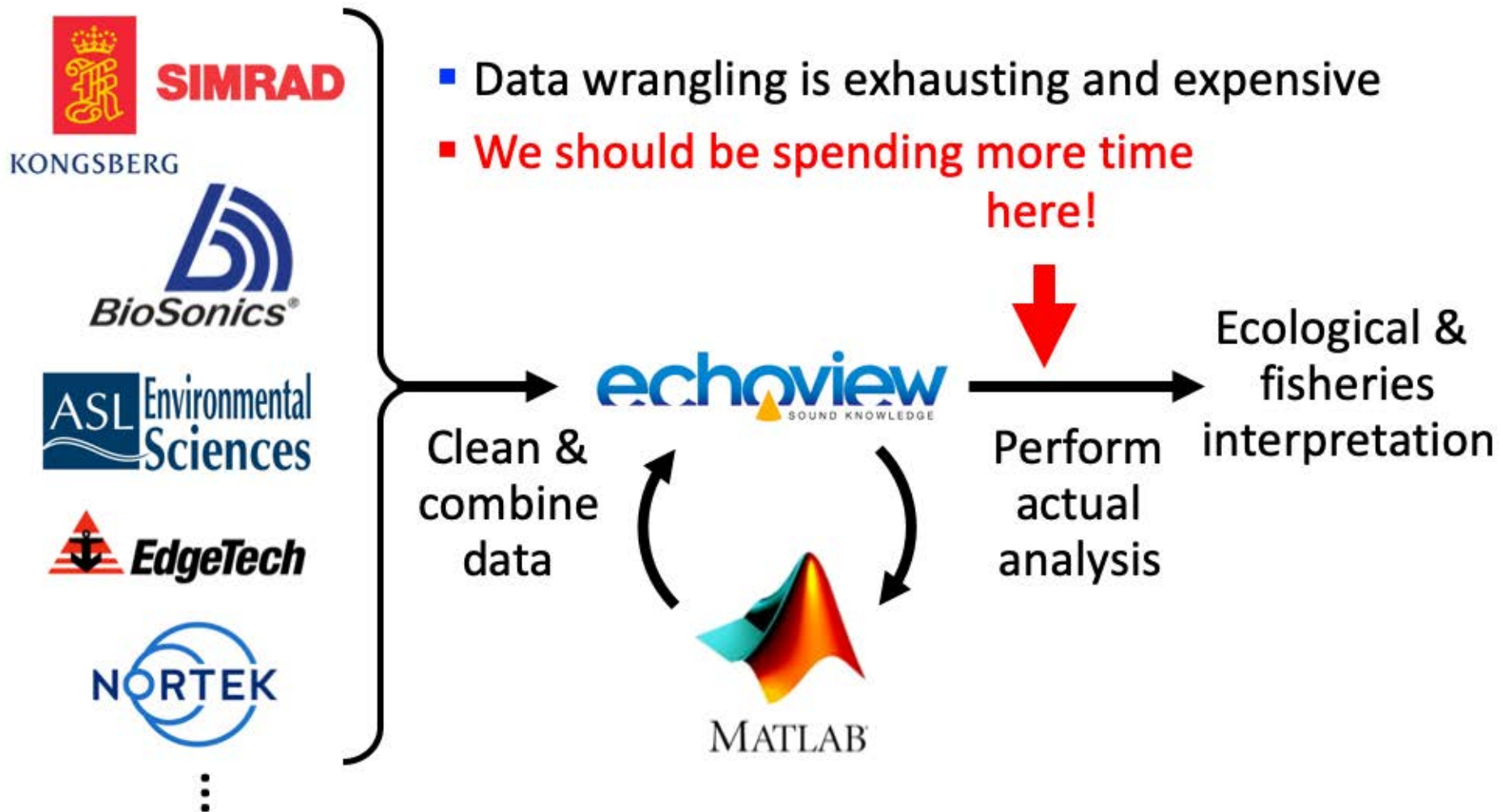
# Vision



- Long-term goal: make ocean sonar data an integrated component of standard oceanographic data sets



# Current common workflow requires extensive data wrangling

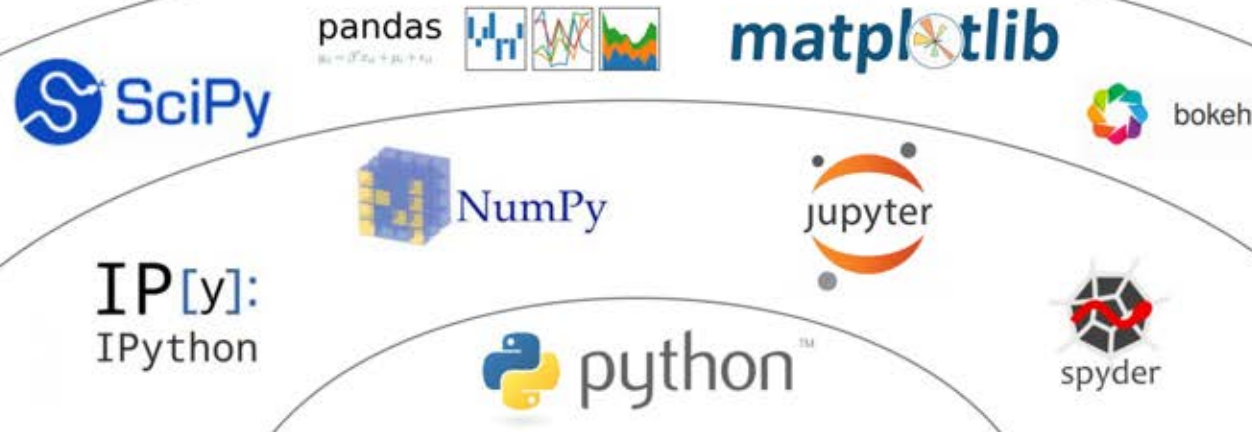


# We can do better

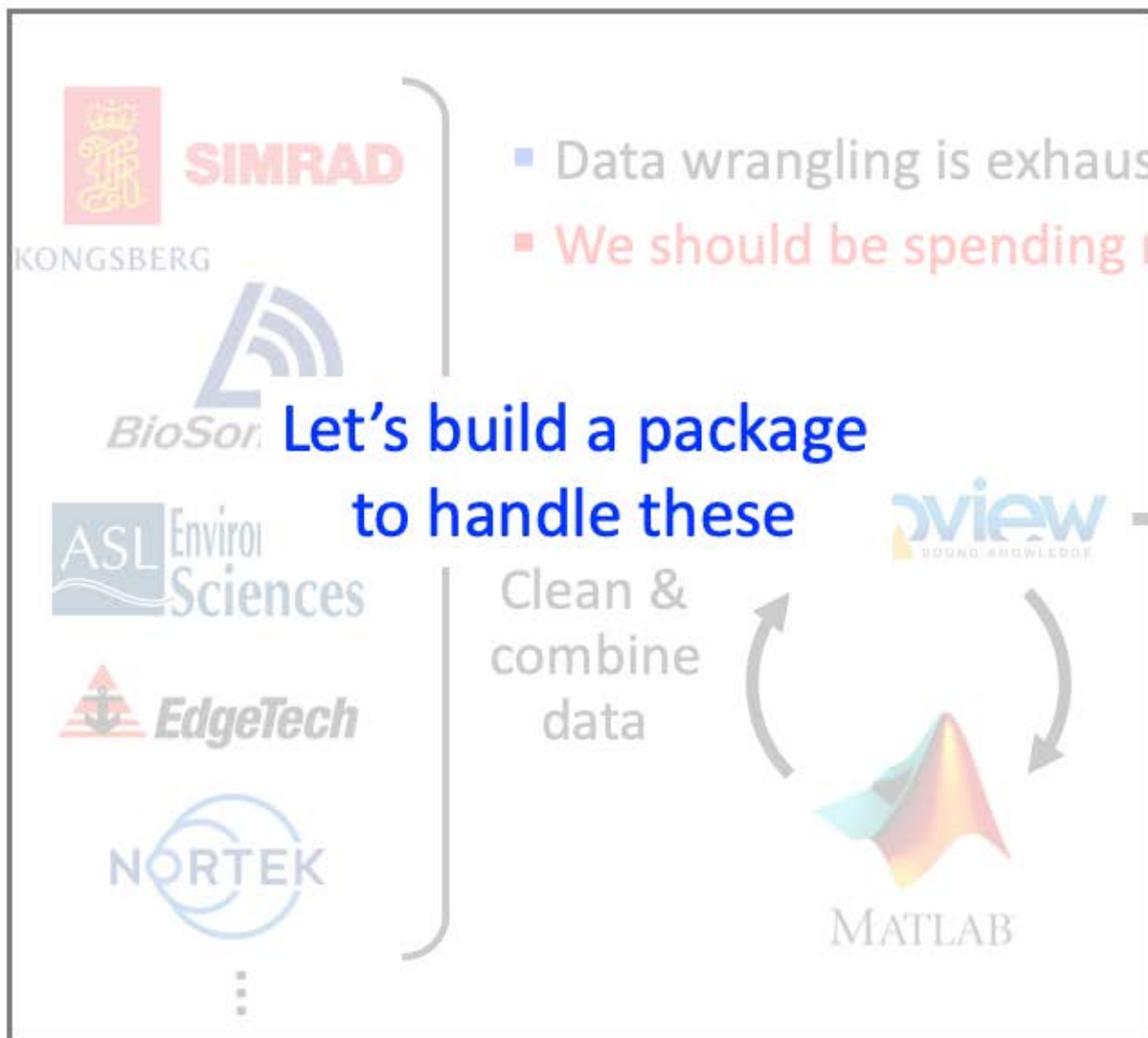
- Leverage the scientific Python ecosystem!



And many,  
many more...



# Current common workflow requires extensive data wrangling



- Data wrangling is exhausting and expensive
- We should be spending more time here!

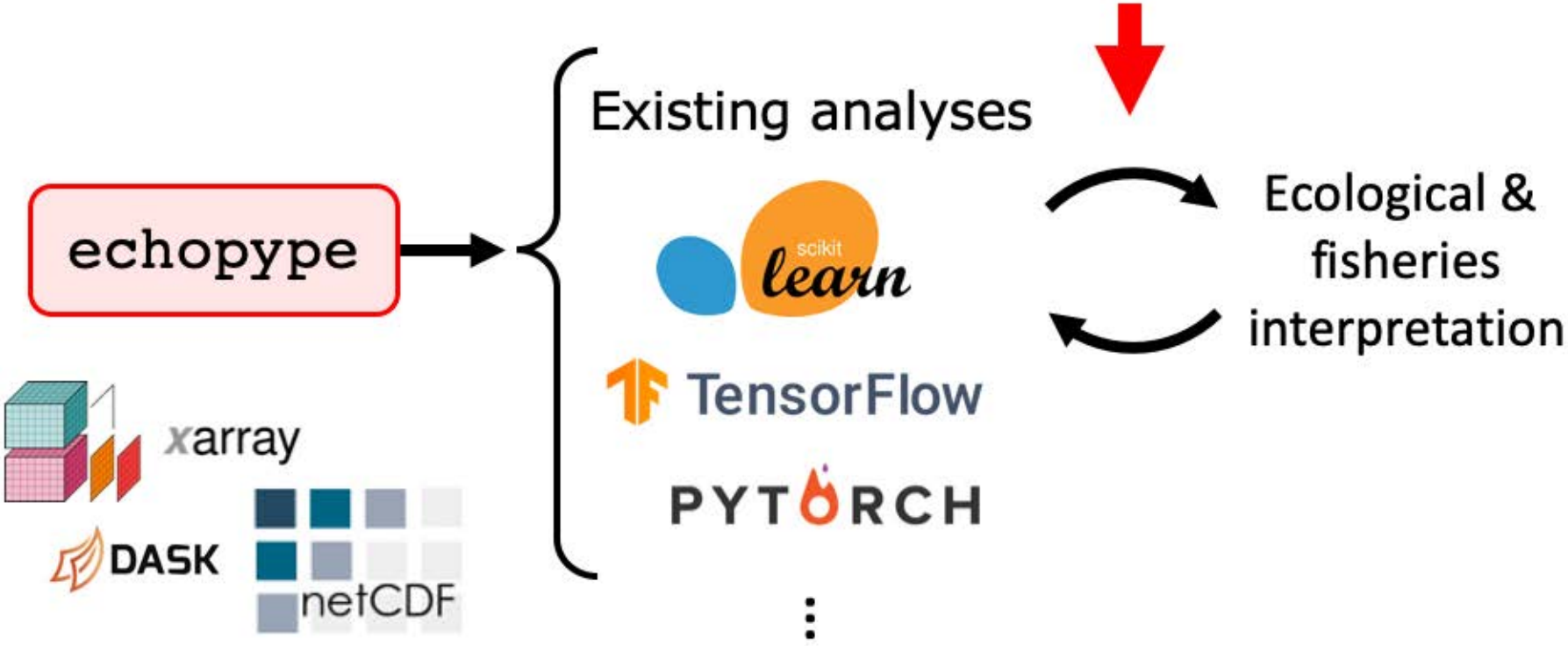


Perform actual analysis

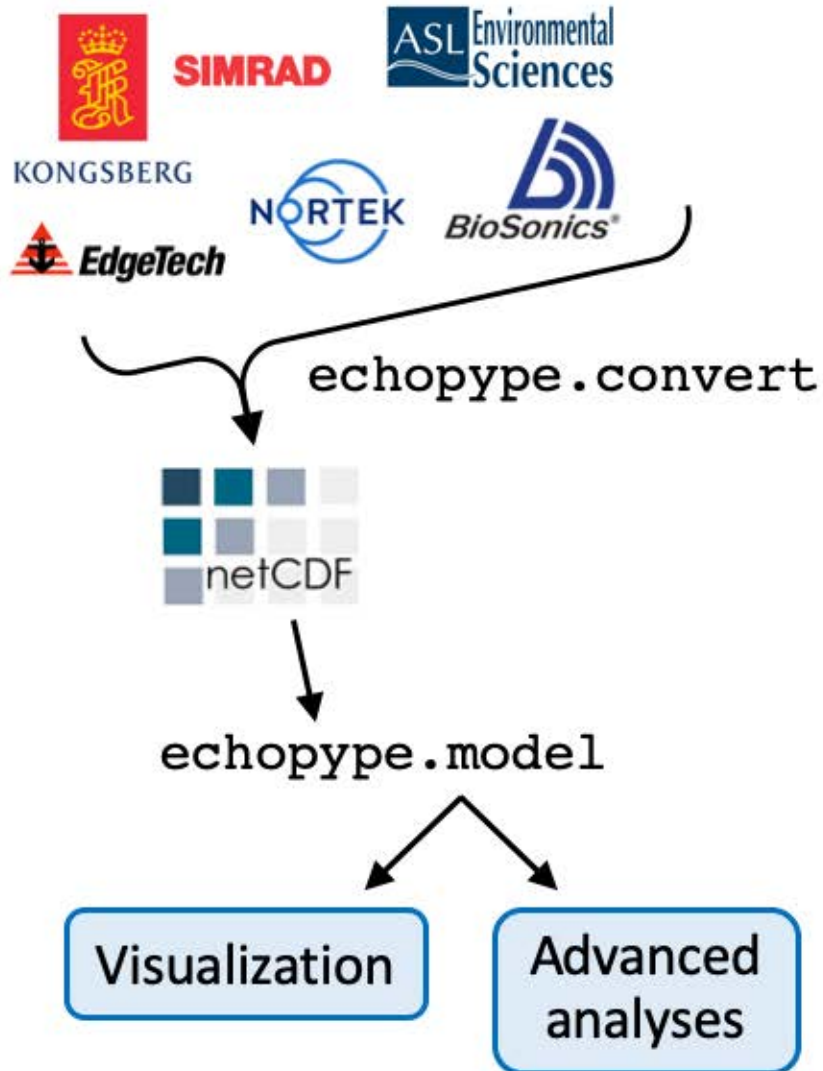
Ecological & fisheries interpretation

# Our proposed workflow

- Simplify data wrangling, so that we can spend more time here!



# echotype at a glance



- Follow ICES SONAR-netCDF4 for raw data storage when possible
- Take advantage of existing libraries
- Uniform interface to facilitate use and further development
- Calibration
- Other pre-processing

# Data representation

- Raw sonar data are just arrays



## NumPy arrays

- Simple for Matlab users
- In-memory operations
- No labeled indexing for time, depth, frequency, etc.
- No metadata support



echopype

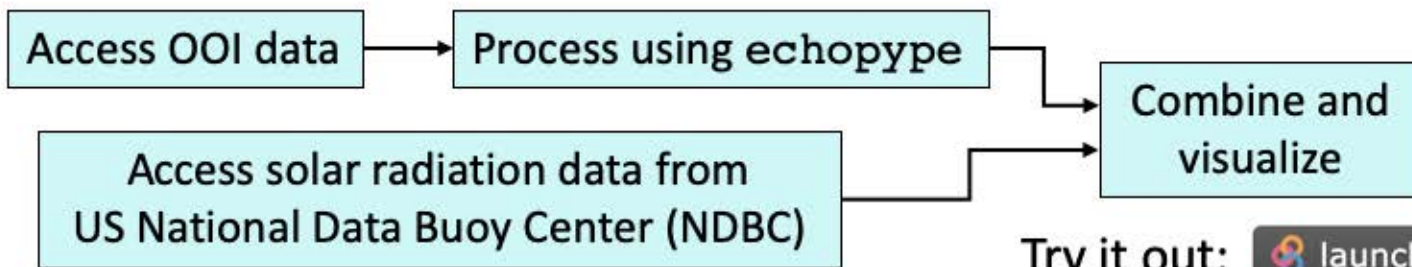


## NetCDF4 gridded arrays

- More complex structure
- Allows direct access for efficient data selection/subsetting
- Labeled support for scientific interpretation
- Libraries for scalability and visualization
- Conversion tools for efficient cloud storage

# Example notebook

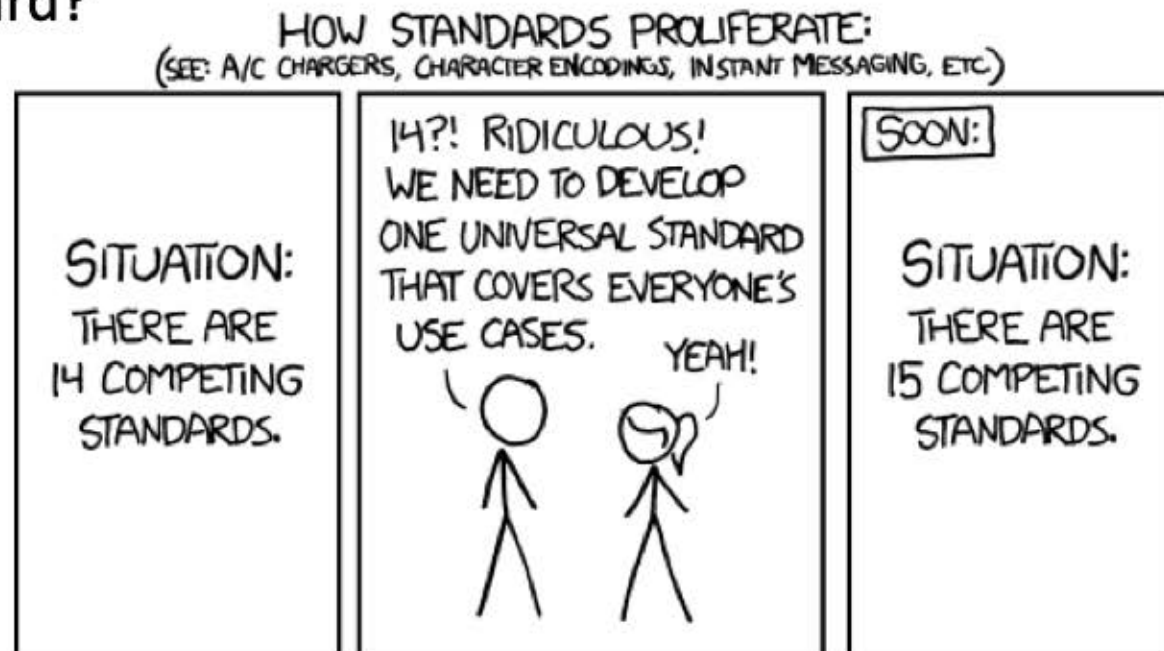
- Watching a solar eclipse using an OOI sonar
- OOI = Ocean Observatories Initiative



Try it out: [launch](#) [binder](#)

## Going forward 1: seek community input for sonar data convention

- echopype follows the SONAR-netCDF4 convention (ICES, May 2018) for raw data storage when possible
- Need convention for processed data to provide computational capability beyond raw data storage
- Generalizability vs efficiency?
- Yet another standard?





## Going forward 2: grow echotype

- Include more sonar data formats, currently echotype supports:
  - Simrad EK60 .raw files
  - ASL AZFP .01A files
- Add advanced analysis and visualization routines
- Streamline cloud deployment
- Engage community in testing and development



# THANK YOU!



- <https://github.com/OSOceanAcoustics/echotype>



Open-Source Ocean Acoustics

Home for open source tools and resources in ocean acoustics

▪ **Join us!**

- **Other contributors**

- Frédéric Cyr (DFO)
- Sven Gastauer (UCSD)
- Marian Peña (IEO Spain)
- Mark Langhirt (PSU)
- Erin LaBrecque (freelance)
- Emma Ozanich (UCSD)
- Aaron Marburg (APL-UW)

- **AZFP Matlab toolbox developer**

- Dave Billenness (ASL Env Sci)

- **pyEcholab developers**

- Zac Berkowitz (SOI)
- Rick Towler (AFSC)
- Chuck Anderson, Veronica Martinez, Carrie Wall (NCEI)

