Pacific Northwest
NATIONAL LABORATORY

Marine Energy Data Pipelines

May 11, 2021

Carina Lansing
Maxwell Levin
Chitra Sivaraman

U.S. DEPARTMENT OF ENERGY BATTELLE
PNNL is operated by Battelle for the U.S. Departm

U.S. DEPARTMENT OF ENERGY | Office of ENERGY EFFICIENCY & RENEWABLE ENERGY
WATER POWER TECHNOLOGIES OFFICE

Introduction

Data Pipeline Team:

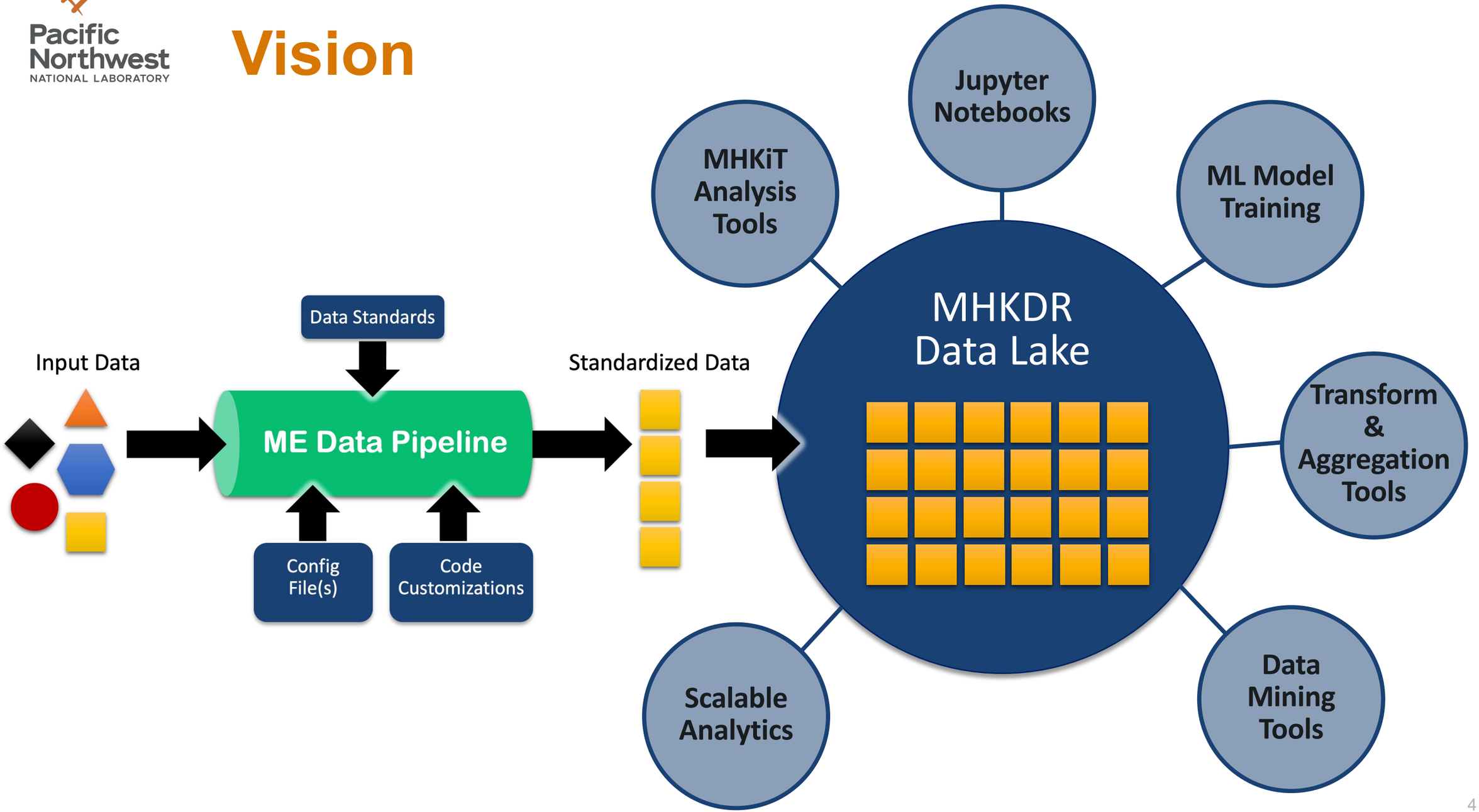
- Chitra Sivaraman – Project Lead
- Carina Lansing – Chief Architect
- Max Levin – Lead Software Developer

- Funded by EERE's Water Power Technologies Office
- Part of a larger MHKiT team and project led by NREL and Sandia
- Sponsored by the Portal and Repository for Information on Marine Renewable Energy (PRIMRE)
- 1st Webinar: Data Standards – April 20th, 2021
- 2nd Webinar: ME Data Ingestion Pipeline – May 11th, 2021

Outline

- Vision for data accessibility & interoperability
- Data standards recap
- Overview of tsdat data pipelines
- Live demonstration
- Roadmap
- Summary

Vision


Jupyter Notebooks
MHKiT Analysis Tools
ML Model Training
MHKDR Data Lake
Transform & Aggregation Tools
Scalable Analytics
Data Mining Tools
Data Standards
Input Data
Standardized Data
ME Data Pipeline
Config File(s)
Code Customizations

Data Standards Recap

A standardized dataset has:

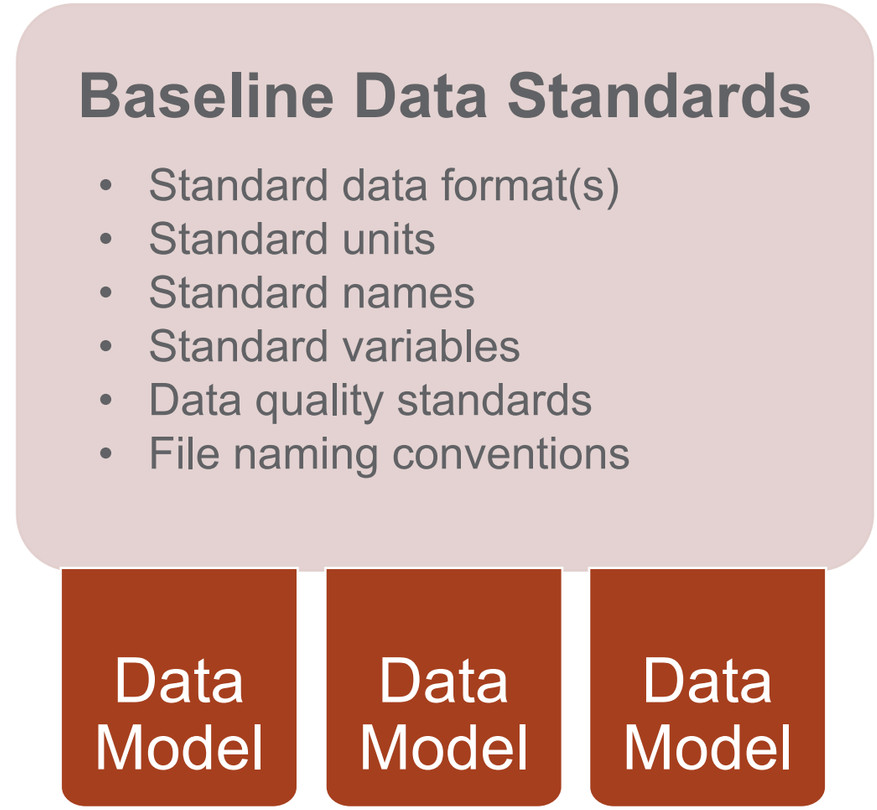
- Metadata (descriptive, provenance, quality control)
- Standard units
- Standard names for variables
- Standard format & file names

Standardized datasets can be extended by **Data Models** which have:

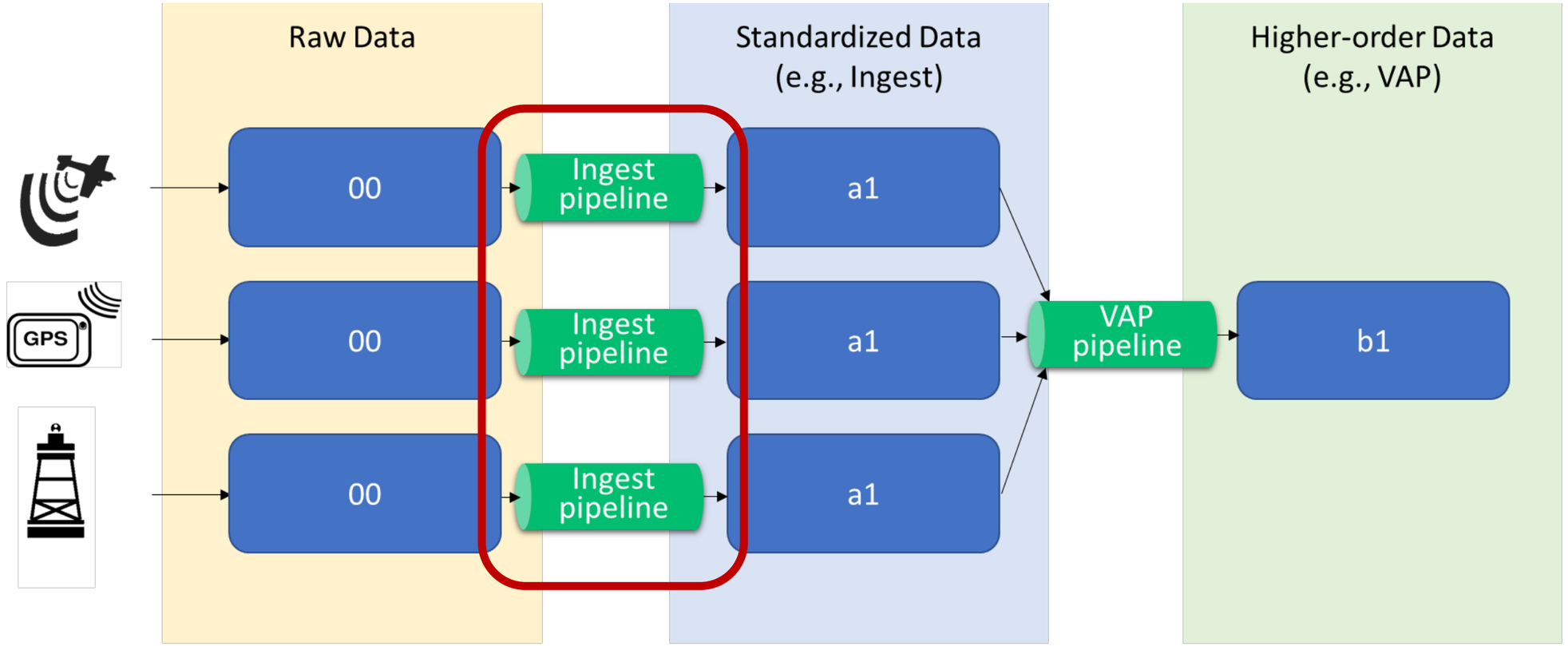
- Sets of specific variables that should be included
- Sets of specific calculations that should be included
- Sets of specific metadata that should be included

Data Standards document

Data Standards Webinar recording and slides



Types of Pipelines

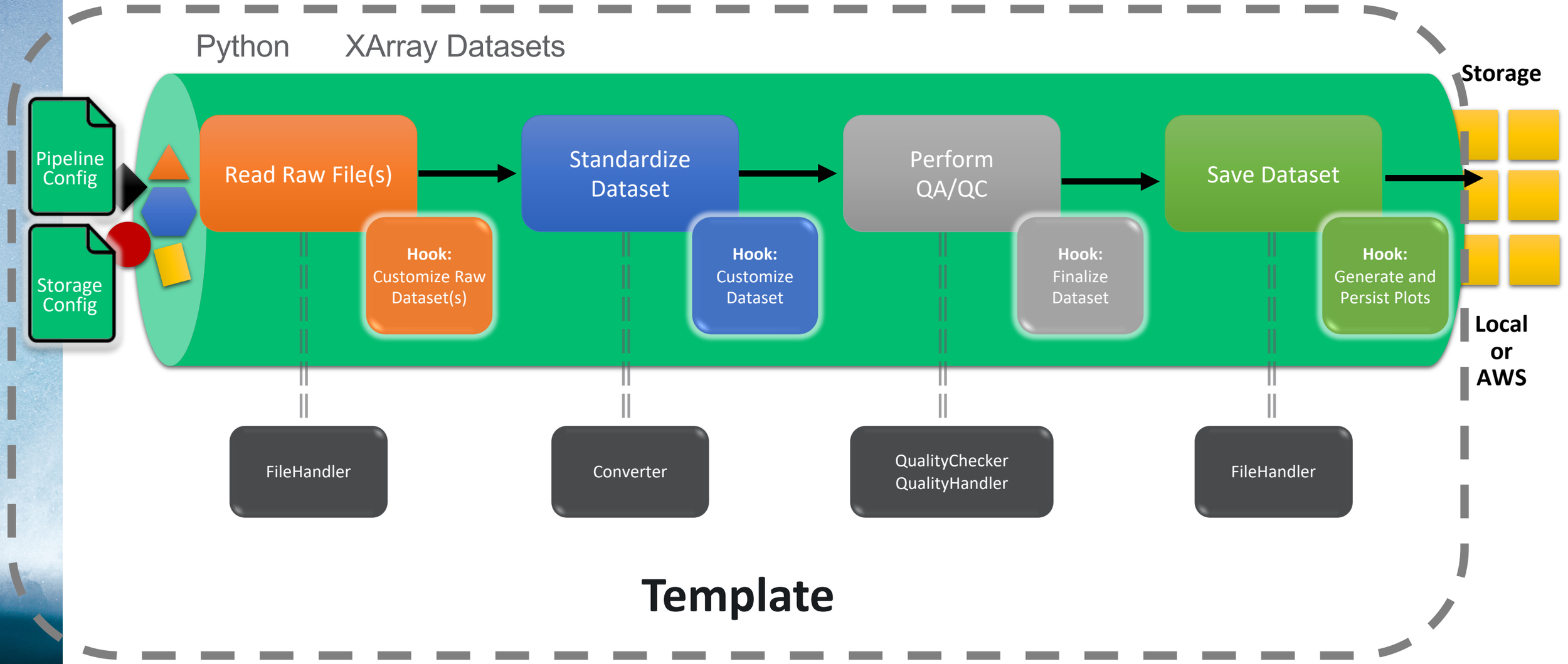


No changes to data produced by instruments or models

Data have been converted to standard format, and quality checks and corrections have been applied

Data have been combined and/or algorithms have been applied to translate the instrumental measurements to the geophysical quantities needed for scientific analysis

Tsdat Ingest Pipeline

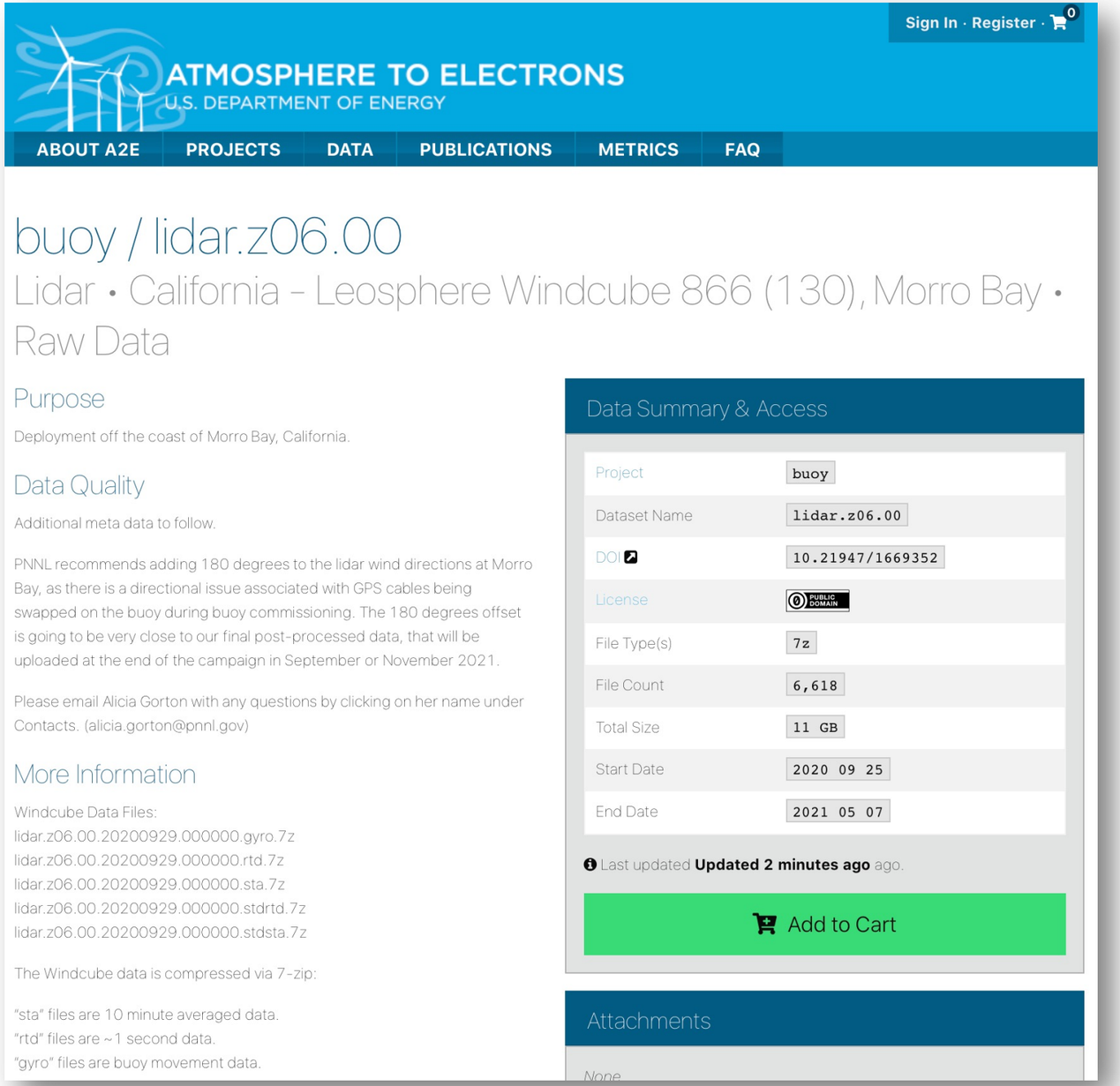


Live Demo

Pacific Northwest NATIONAL LABORATORY

Demo Overview

- Brief example of a tsdat ingest
  - Getting started, customizable components, running the pipeline
- Data from the A2E program
  - Buoy located in Morro Bay, CA
  - Lidar measuring wind vectors
  - “sta” 10-min averaged data
- Goals
  - embed metadata for ease-of-use
  - perform quality checks / controls
  - visualize processed data



Tsdat - Getting Started

Prerequisite: Python 3

Step 1 Get a template from GitHub `https://github.com/tsdat/template-repositories`

Step 2 Install tsdat library

```
pip install tsdat
```

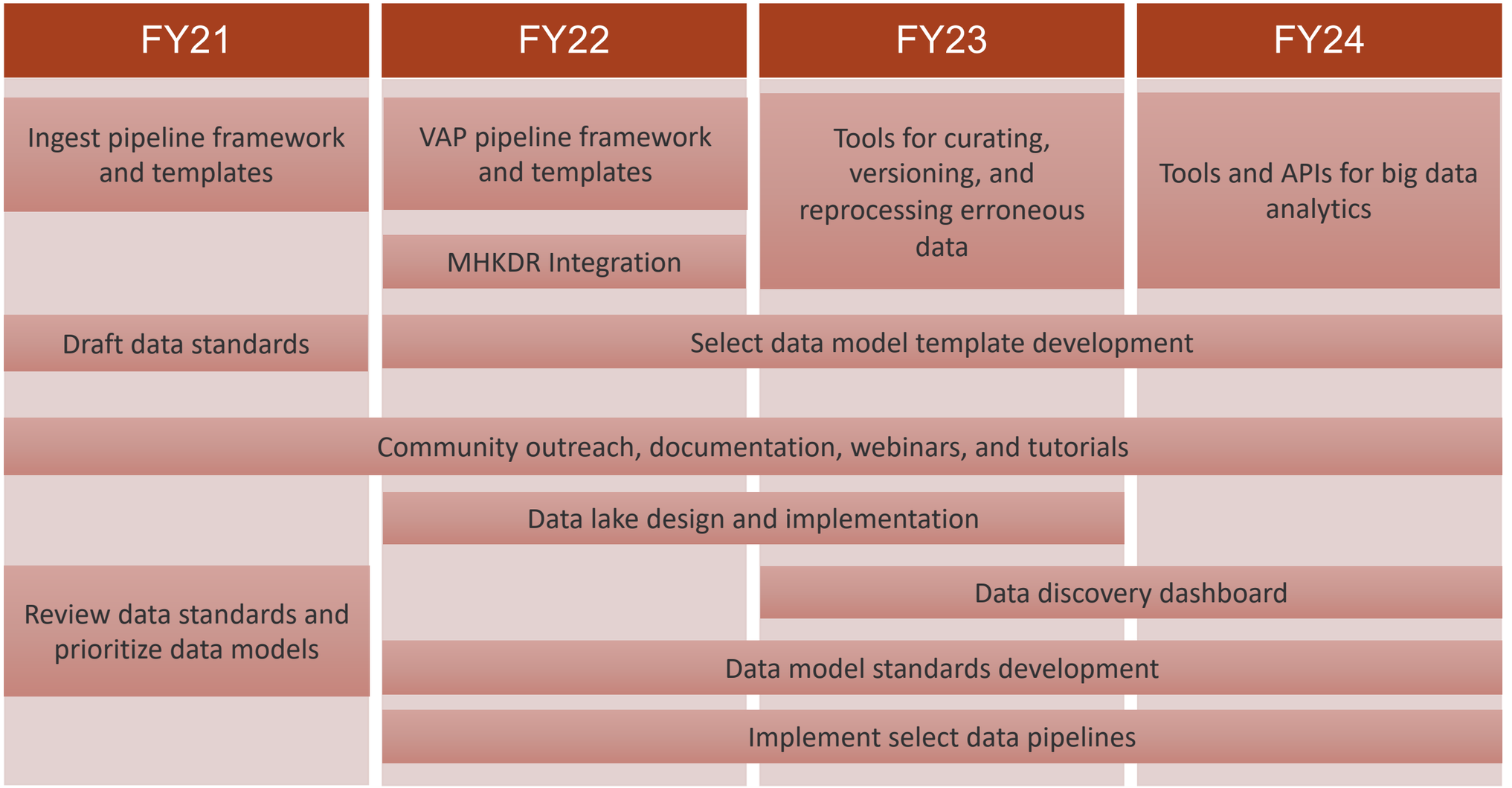
Step 3 Configure your template

Step 4 Run the pipeline

```
python3 run_pipeline.py
```

Questions & Answers

Roadmap


FY21
FY22
FY23
FY24
Ingest pipeline framework and templates
VAP pipeline framework and templates
Tools for curating, versioning, and reprocessing erroneous data
Tools and APIs for big data analytics
MHKDR Integration
Draft data standards
Select data model template development
Community outreach, documentation, webinars, and tutorials
Data lake design and implementation
Review data standards and prioritize data models
Data discovery dashboard
Data model standards development
Implement select data pipelines

Pacific Northwest
NATIONAL LABORATORY

Summary

- Tsdat code & templates: https://github.com/tsdat
- Tsdat documentation: https://tsdat.readthedocs.io/

- We are recruiting early adopters, so please reach out to us if we can assist!

  tsdat@pnnl.gov

Pacific Northwest
NATIONAL LABORATORY

Thank you