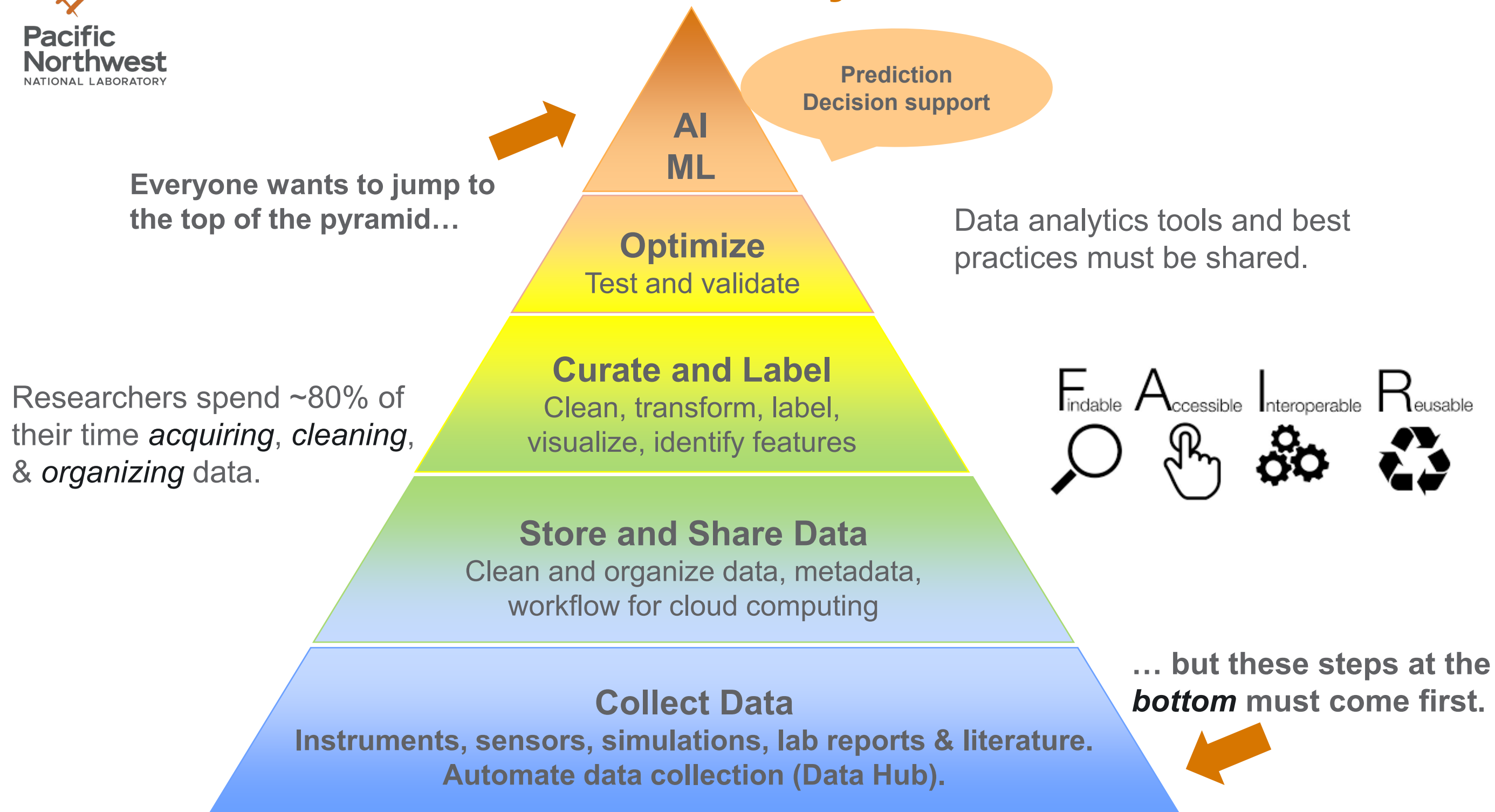


# The Data Pyramid



# Challenges when AI meets domain science

1. Small & expensive data.  
Limited data relevant to energy storage.

2. Data quality and uncertainty.

3. Complex processing-structure-property-performance relationships.

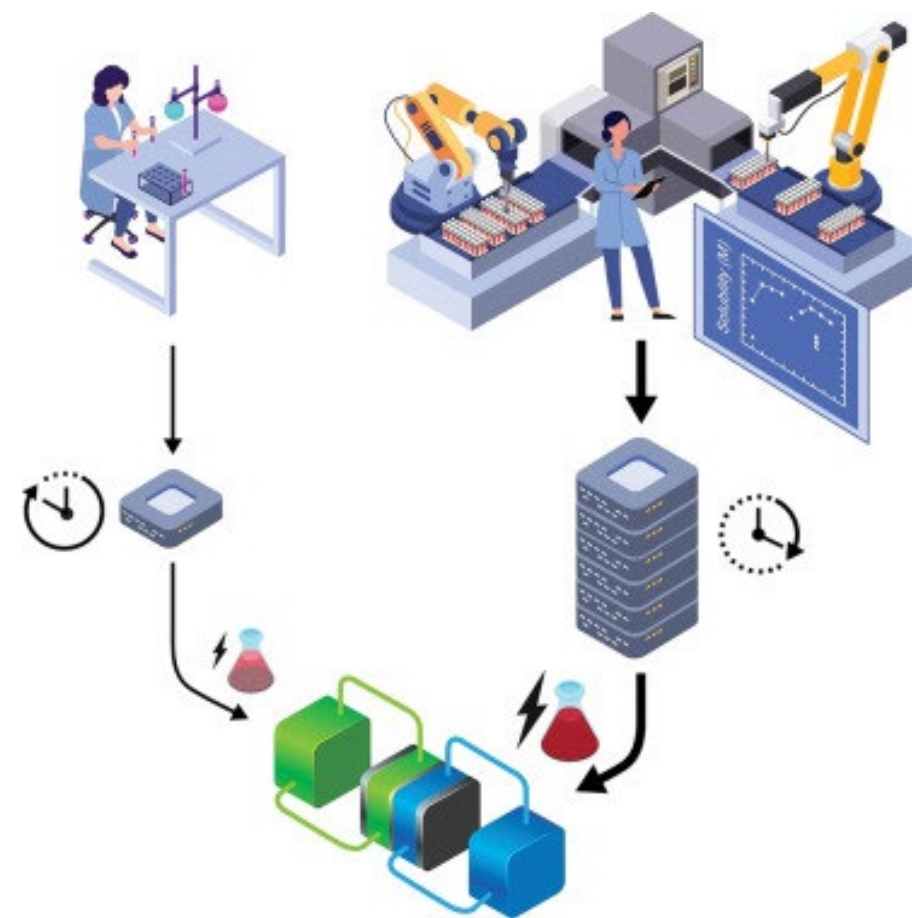
4. Use of black box models.

5. Sloppy use of ML + hype.

**Solutions: Data curation, data sharing, few shot ML, emphasize statistical learning and physical basis of results. Domain-specific foundation models.**

**Dimensions of data quality: Completeness, accuracy, usability, and standardization.**

M. Wenzlick, J. R. Bauer, K. Rose, J. Hawk, and R. Devanathan. Integrating Materials and Manufacturing Innovation 9 (2020) 89.



Y. Liang et al, High-throughput solubility determination for data-driven materials design and discovery in redox flow battery research. Cell Reports Physical Science 4 (2023) 10.