



Pacific Northwest
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

Signature Discovery Initiative

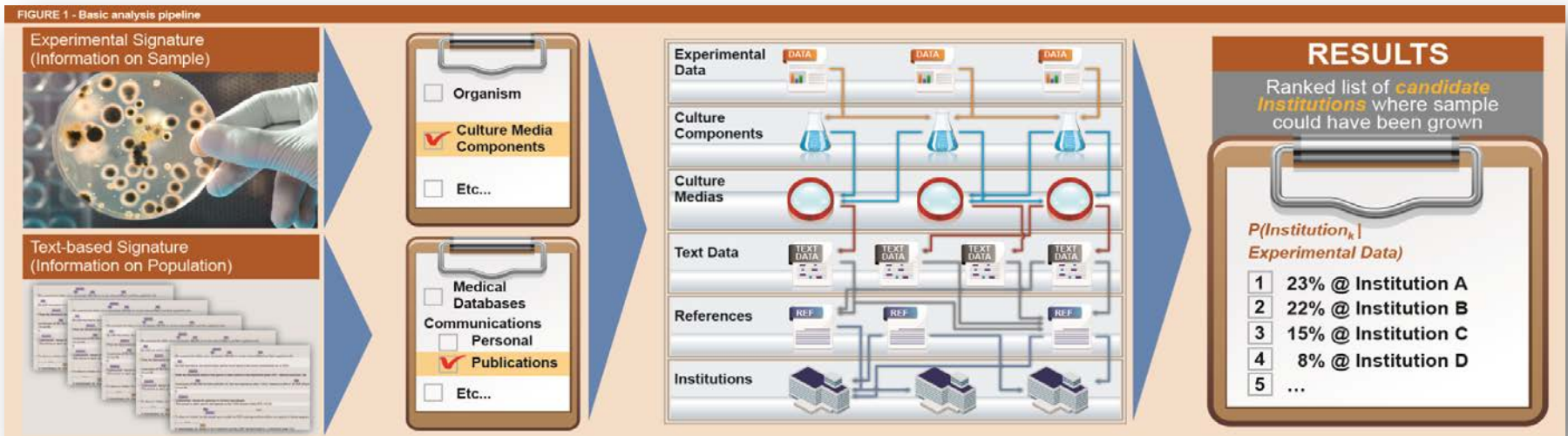
NIAC DAY @PNNL

Presented by:

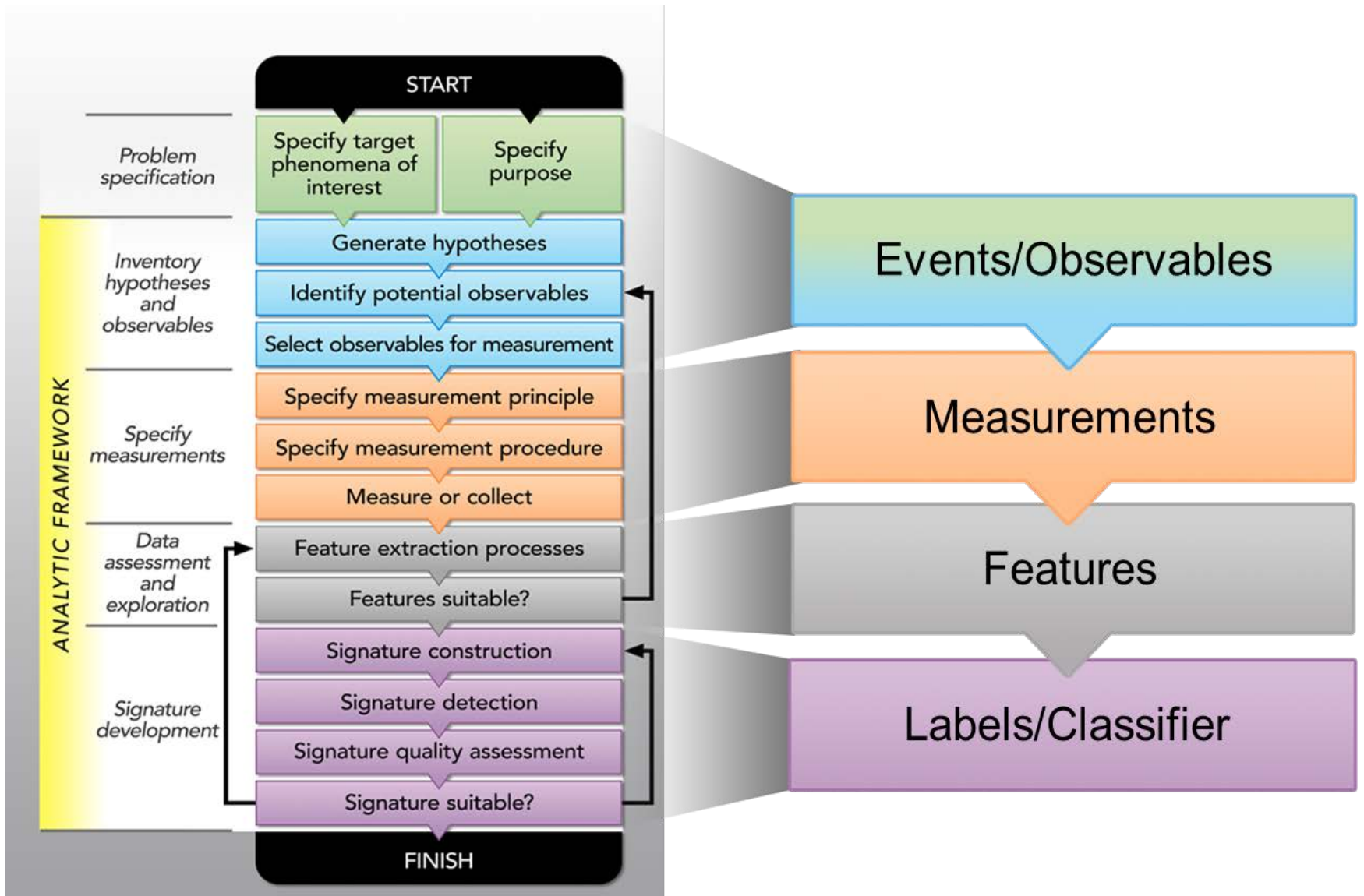
Nathan Baker (nathan.baker@pnnl.gov)

Introduction to Signature Discovery

- ▶ **Signature:** A distinguishing collection of features that detects or characterizes a phenomenon of interest:
- ▶ **Signature systems** consist of multiple components:
 - *Measurements* transform observations of events into data
 - *Feature extraction* maps data into features
 - *Classifiers* transform features into labels and confidence levels



Signature Discovery Process



SDI Initiative Challenges and Goals

Challenges

- Modern signatures often involve features from many different data types
- Signatures are most useful when they are easily interpreted by decision-makers

Key questions

- How do we best **select features** from multiple measurement sources to construct our signature?
- How do we **detect signatures** in “real world” environments?
- How do we **assess the quality** of a signature and compare different signatures?
- How do we **recognize change and adapt** signatures to dynamic phenomena?
- Is this process **generalizable across domains**?

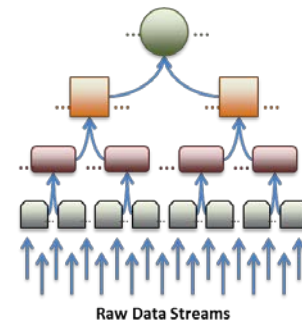
Fishing for features

- Greedy strategy to feature discovery and selection
- Application to survey-based disease signatures
- Application to microbe-based fuel production



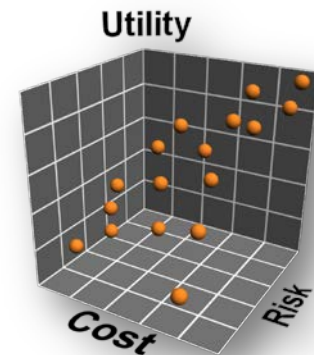
Hierarchical signature detection

- Multilevel signature detection
- Bandwidth-limited data environments



Signature Quality Metrics

- Going beyond ROC curves
- Framework to combine signature fidelity, risk, cost, and utility



Current research areas and additional expertise needs

- ▶ Identifying new application areas for SDI methodology
- ▶ Augmenting SDI methodology with new computational and mathematical approaches
- ▶ Training and internship opportunities in signature discovery
- ▶ Joint proposal activities

