

# Policy Towards Sharing Smart Grid Data, Workflows and Services

Focus Group

NSF WORKSHOP  
Smart Grids Big Data

Moderators:

Arcot (Raja) Rajasekar

University of North Carolina, Chapel Hill

Yingchen (YC) Zhang

National Renewable Energy Laboratory

# Challenges in Sharing Digital Artifacts

- **Findability**
  - Identifications – DOI
  - Distributed Data collections
  - Dark Data
  - Autonomous Administrations
  - Protocol differences
- **Availability**
  - Loss, Corruption, Broken Links
  - Security, Privacy, Business Confidentiality
- **Interoperability**
  - Context capture (Metadata)
  - Evolving technology
  - Protocol/API/Format changes
- **Reusability**
  - Reproducible results
  - Re-purposable collection
  - Integrity & Fidelity
- **Massive scale – Big Data**
  - Size, Numbers, Speed, Types, Formats
  - Users, Resources, Applications, Data
- **Increasing lifespan**
  - Long-term curation
  - Context capture (Metadata)
  - Technology obsolescence

## Five V's of Big Data

**Volume**

**Velocity**

**Variety**

**Veracity**

**Value**

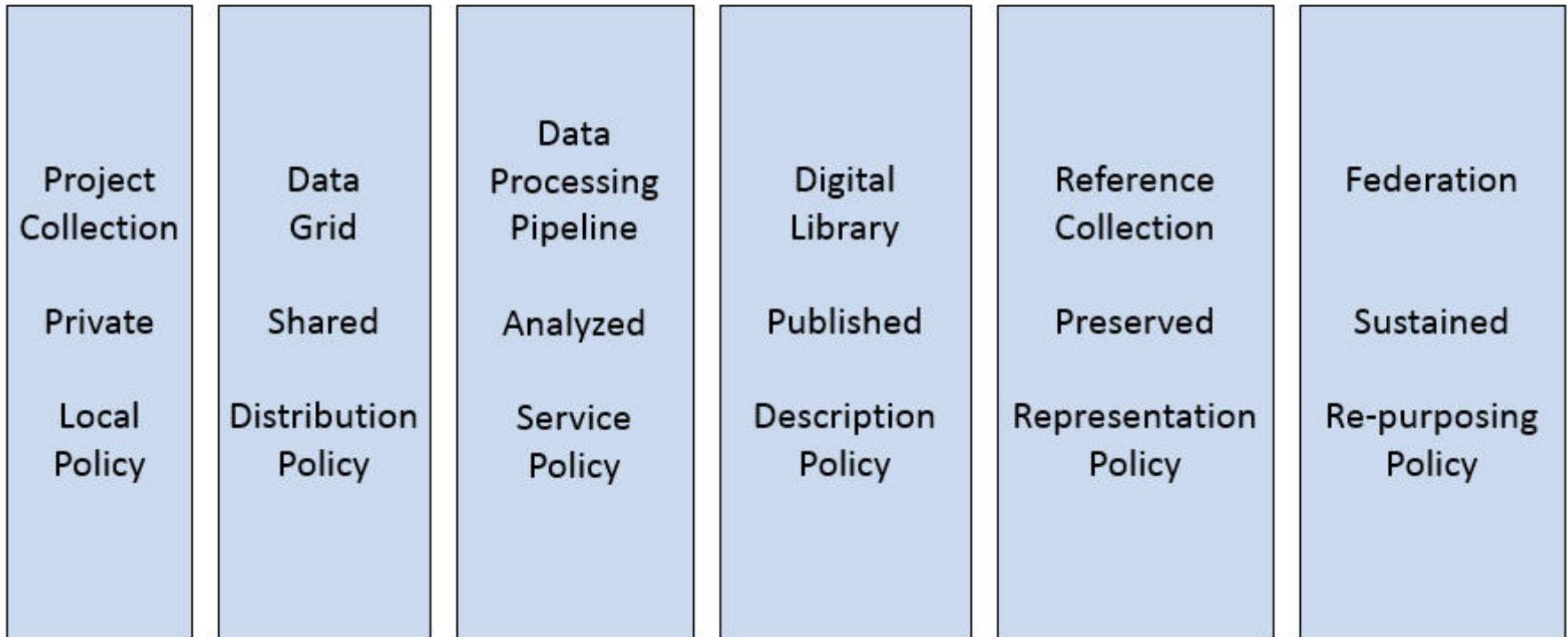
# Why do we need Policies?

- Manage interactions with users
  - Control each user action
- Manage interactions with technology
  - Control interfaces to versions of technology
- Manage properties of the DA collection
  - Enforce management decisions
  - Verify collection properties
- Regulate the data sharing data applications
  - Enforce security, privacy, and confidentiality
  - Ensure accuracy, interoperability, and scalability

# Policy Changes During A Full Life Cycle

## Collection Life Cycle

Each collection life cycle stage re-purposes the original collection



Stages correspond to addition of new policies for a broader community  
Virtualize the collection life cycle through policy evolution

# Some Essential Data Center Policies

- 1. Ingestion Policies**  
ETL, Authentication, Integrity
- 2. Contextual Metadata**  
Automate extraction of metadata from files
- 3. Privacy**  
Automate application of appropriate access controls
- 4. Availability**  
Automate creation of replicas; Create multi-resolution/format replicas
- 5. Conformance/Fidelity Check**  
Automate identification of “format”
- 6. Publication**  
Increase findability and connectedness;  
Federation
- 7. Data Retention & Disposition**  
Apply a retention period; Automate disposal
- 8. Continuous Integrity**  
Automate verification and replace bad copies
- 9. Application accuracy**  
Provide means to validate the accuracy, interoperability, and scalability of the data applications
- 10. Auditing/Notification**  
Automate auditing and manage events about changes to the collection
- 11. Security**  
Automate Virus checking
- 12. Technology Obsolescence**  
Automate data migration;  
Containers/virtualization

# Questions & Discussions

- What are the top three data needs in your research/business?
- What are the top three data challenge you are facing?
- What are the top policies that are needed for
  - Data identification & discovery
  - Data processing – Ingestion to Analytics
  - Data sharing - FAIR Principles
  - Data management - across time, space and domains
- How can policies help in bridging OT and IT?
- How can policies help in federating data from multiple resources?
- How can policies span the multi-resolution data being analyzed?
- How can automation play a role?
  - Volume and velocity are not suited for human-in-the-loop
  - Long-term data/process availability need automation
  - Institutional knowledge capture

# Focus Group Report

# Policy Towards Sharing Smart Grid Data, Workflows and Services

Focus Group

NSF WORKSHOP  
Smart Grids Big Data

Moderators:

Arcot (Raja) Rajasekar

University of North Carolina, Chapel Hill

Yingchen (YC) Zhang

National Renewable Energy Laboratory



# “Policy for Sharing” Focus Group Report

- We first looked at:
  - FAIR Principles for Sharing
  - Why we need Policies
  - Change of Policies in Time
  - Policies for a Data Center
- Then we looked at how problems in “data sharing” can be done in the Smart Grid arena
  - Legal Issues are a bottleneck for sharing Real Data
    - Multiple NDAs need to be worked on between parties
  - Synthetic Data is a solution
  - Synthetic Data need to incorporate all dimensions and resolution
  - Synthetic Data do not Capture Anomalies
  - Can we promote a Generic NDA to help sharing real data
    - Involve NSF in promoting this
  - Different types of data need different types of anonymization
  - Problem-specific Standards for sharing data is needed
    - Eg. AMI, PMU, Health Monitors might need different data scale

# “Policy for Sharing” Focus Group Report

- Three Main Ideas Emerged
  - **Clearing House for Real Data**
    - Similar to NIH Clinical Data Sharing (TraCS)
    - Utilities can place data in a trusted third-party sites
    - Can have a common NDA
    - Follow Standards for different types of Data/Problems
    - NDA repository
    - GURR can enable this interaction
    - FERC and NSF can be involved
  - **Hybrid Synthetic-Real Data**
    - Deals with Security & Privacy concerns
    - Incorporate Anomalies
    - Hard Research Problem to integrate real data with synthetic data
  - **Ask Utilities to Share Data Anonymously**
    - No NDA signed
    - No Acknowledgement made
    - Utilities get to have their data worked on by many data analysts
    - Open Data towards open solutions
    - Use in competition, Hackathons, etc.