Policy-Encapsulated Objects

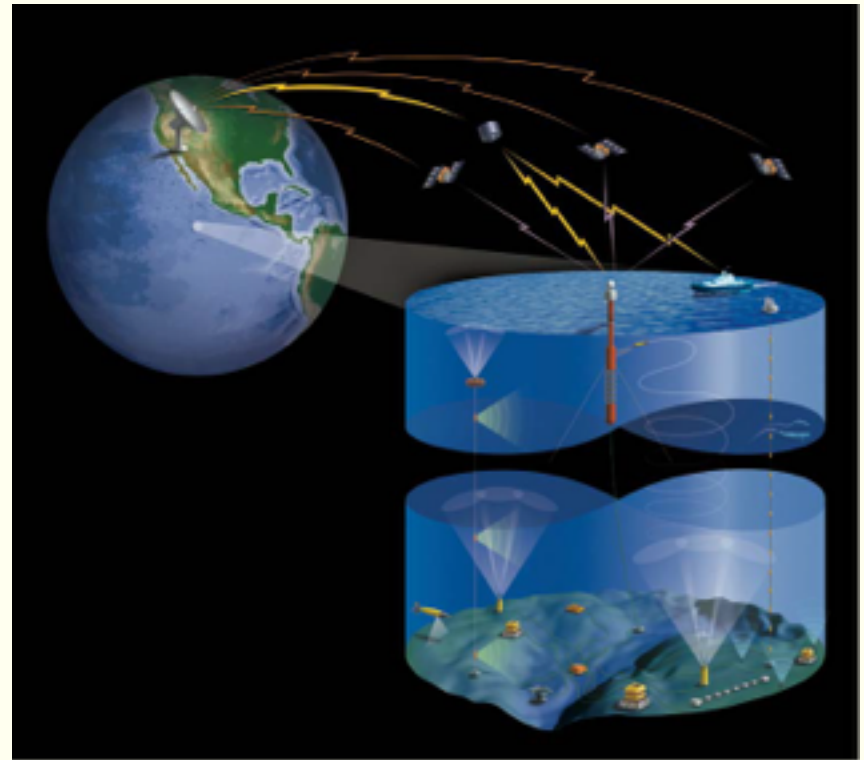**Arcot (Raja) Rajasekar**

rajasekar@unc.edu
The University of North Carolina
at Chapel Hill

Outline

- Motivations
- Anatomy of PEOs
- Architectural Issues
- Q&A

Acknowledgement:
Prof. Reagan Moore and I had discussions about PEOs around 4 or 5 years back but never got around to doing anything about them. Reagan is part of the intellectual genealogy of the PEOs.

CoreGen3

iRODS

What does iRODS bring to the table?

– Federated virtual dataspace
 (also other spaces: userspace/resources/…)
– Rich Metadata Discovery System
– Extensible System Information (ACLs, Audits,…)
– Distributed Data Pipelines
– User-tunable workflows & μ-service
– Distributed Rule Engine
– Etc., Etc., …

<u>**FAIR Data**</u>
<u>**Principles**</u>
  Findability
  Accessibility
  Interoperability
  Reusability

Extensible Big Data Life Cycle Mana

What is Missing?

- Portability beyond iRODS
- Chain of Custody beyond iRODS
  iRODS looses control when a dataset is taken out its zone
  What is lost:
  – Continuous Integrity Maintenance
  – Continuous Authorization & Authentication
  – Continuous Auditing
  – Continuous Versioning, Edit control
  – Linkages with Metadata – (user, system, …)
  – Linkages with ACLS, Workflows, Pipelines, μ-servi

  FAIR Data
  Principles
    Findability
    Accessibility
    Interoperability
    Reusability

  – Things are good as long as they are inside iRO
  – Move it out or even out of a zone to another,
        we loose control
  – Dangling data lifecycle
  – This is true not just for iRODS
    but also any storage system
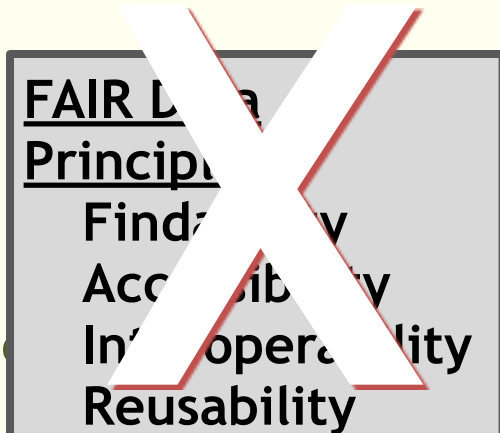
Extraterritorial **jurisdiction** (ETJ) is the legal ability of a
government to exercise authority beyond its normal boundaries.

E.T. phone home.

ET the extraterrestrial,
1982, Universal Studios

Power up!!

- Answer: Make a data object to be active
- Now data objects are passive
  - They have no control over what happens to them,
  - Where can they be stored
  - Which application can handle them,
  - Which user can view them,
- All actions on an object controlled by outside entities and processes
  - even inside iRODS

GIVE POWER TO DATA OBJECTS -> FREE THE DATA
        Give them Independence to control their destiny
   Give them Policies and
              Set them free

What is a Policy Encapsulated Object?

INTERPRETER

POLICIES & SERVICES

METADATA

DATA

PEO

Virtual Machines   **Containers**
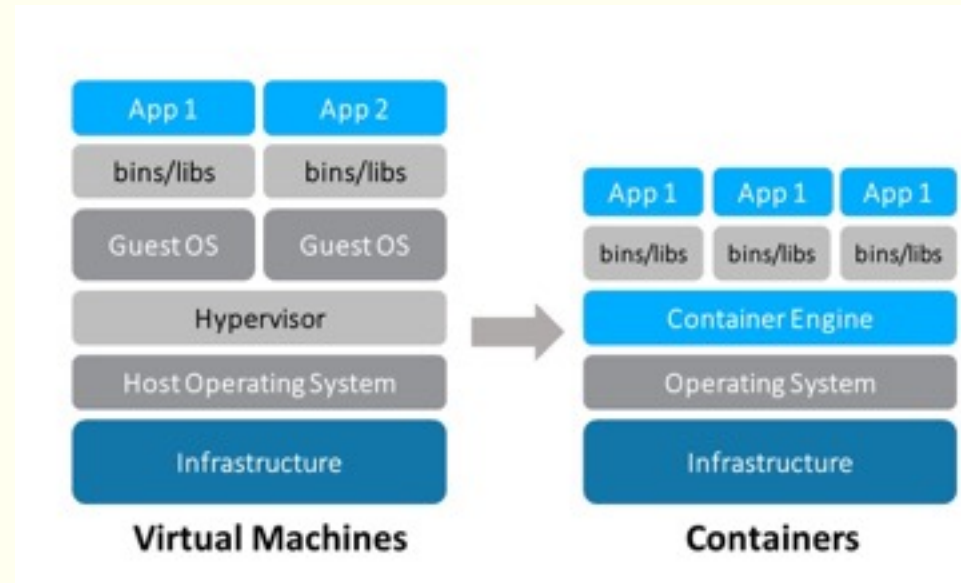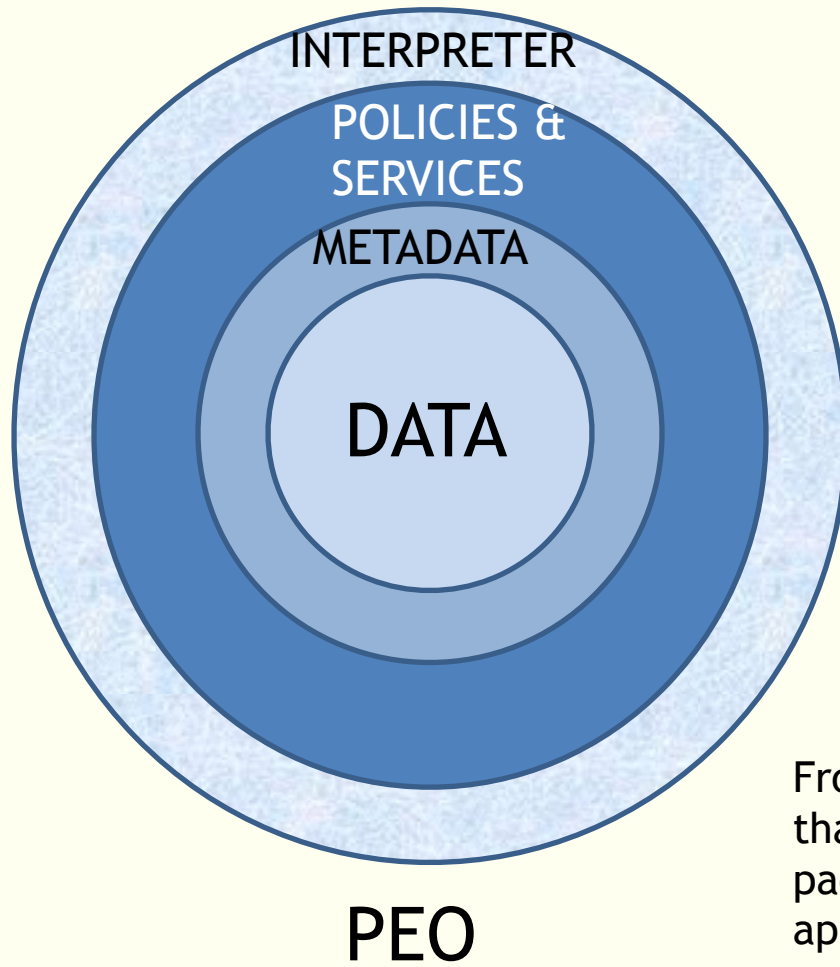
From Docker:. A container is a standard unit of software that
packages up code and all its dependencies so the application
runs quickly and reliably from one computing environment
to another. A Docker container image is a lightweight, standalone, executable package of software that

PEO = Trusted Data

- Live Data Object
- Trust & Integrity
  - Reproducibility
  - Trusted Environment
    - Trust goes both ways
- Self-containment
  - Portability
  - Independence
- Chain of Custody
  - FAIR Data Principles
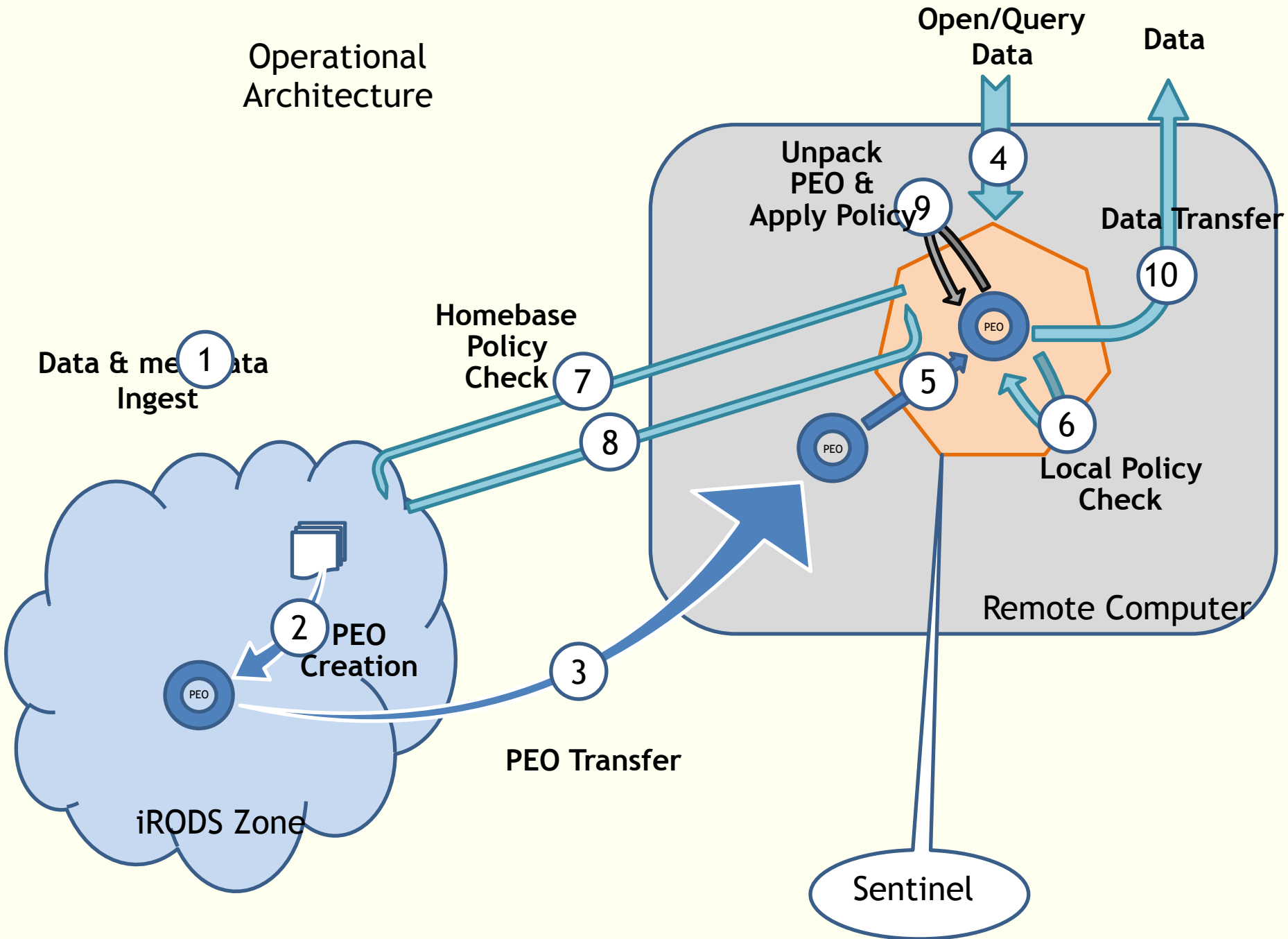 - Full Data Life-cycle Compliance
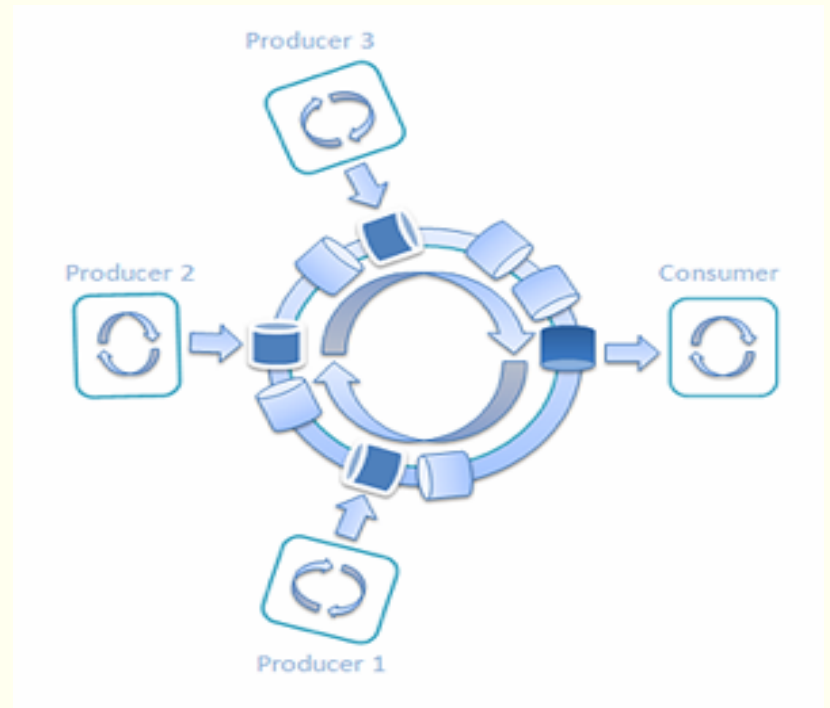
Types of POE

- Tethered POE
  - Checks back to Home Zone
  - Home Zone can update and recall!!
  - Change Policy – Yank ACLs!!
  - Audit Trail & Remote Editing can be synchronized
  - Kill from far!!
- Autonomous PEO
  - Simpler
  - Self-reliant

Architecture

- Active Data Architecture
  - Active Policies (Execute on Event)
  - Apply policies when they mature
    - Event-driven or Periodic
- Message-based Architecture
- Event-based Architecture
- Rule Interpreter Engine
  - Failure Action or Recovery

Operational Architecture

Open/Query Data

Data

Unpack PEO & Apply Policy

④

⑨

Data Transfer

⑩

Homebase Policy Check

⑦

⑧

Data & metadata Ingest

①

PEO

⑤

Local Policy Check

⑥

PEO

Remote Computer

②

PEO Creation

PEO

③

PEO Transfer

iRODS Zone

Sentinel

Use Cases

- Security
- Privacy
- Autonomy
- Automation
- Compliance
- Fidelity
- Tight beam data transfer
- Integration with Blockchain

Q & A

rajasekar@unc.edu