Data Replication

EUDAT Task Force

Morris Riedel, Jedrzej Rybicki
{m.riedel, j.rybicki}@fz-juelich.de
Juelich Supercomputing Center (GER)

EUDAT

Date: 8.03.2012

Content

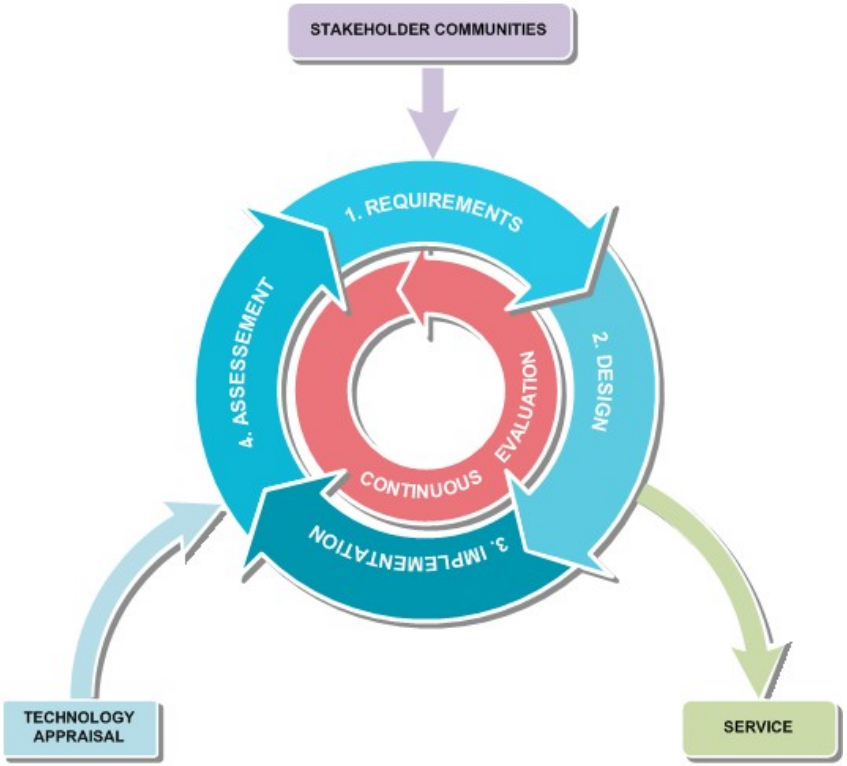
1. The idea of Task Forces
2. Goal: What is replication?
3. EUDAT Architecture
4. Technical details of replication
5. Time line
6. Summary

Requests for the presentation

Temporary Slide

- Where we are and how we intend to implement the services
  - But don't be too technical!
- Safe Replication:
  - Timing
  - Functionality
  - Main components

EUDAT Working Principle



Task Force Islands

EPOS
EUROPEANPLATEOBSERVINGSYSTEM

enes European Network for Earth System Modelling

CLARIN
Common Language Resources and Technology Infrastructure

EUDAT

Task Force Islands

EPOS
EUROPEANPLATEOBSERVINGSYSTEM

enes European Network for Earth System Modelling

CLARIN
Common Language Resources and Technology Infrastructure

EUDAT

Task Force Islands

CINECA

EPOS
EUROPEANPLATEOBSERVINGSYSTEM

enes European Network for Earth System Modelling

JÜLICH FORSCHUNGSZENTRUM

sara

rzg RECHENZENTRUM GARCHING

CLARIN
Common Language Resources and Technology Infrastructure

EUDAT

Data Replication

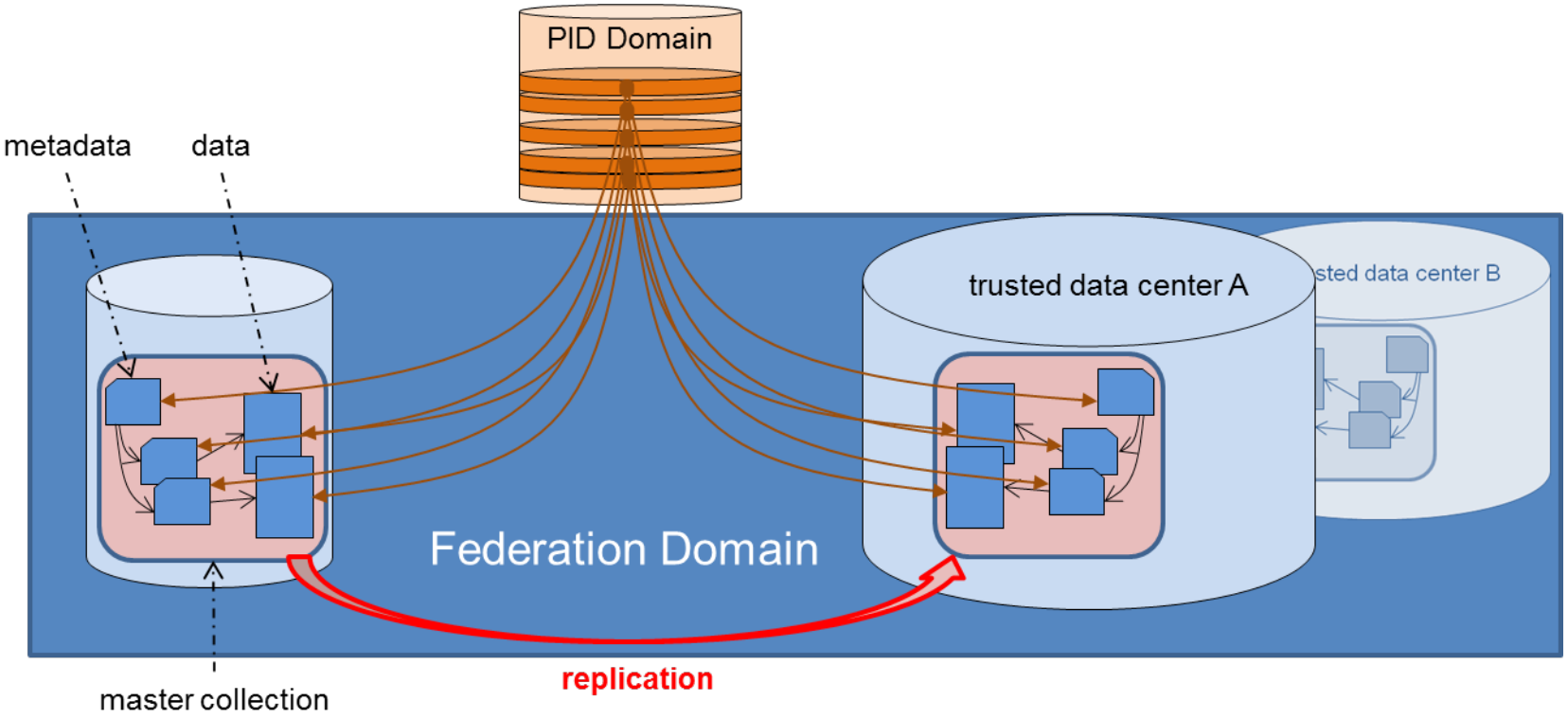
Motivation:

- Ensure bitsream preservation
- Enable data curation functionality
- Improve data accessibility

Common functionality:

Create M replicas (identified by a PID record) at different data centers for N years, exclude centers X, maintaining the given access permissions.

High-level Idea



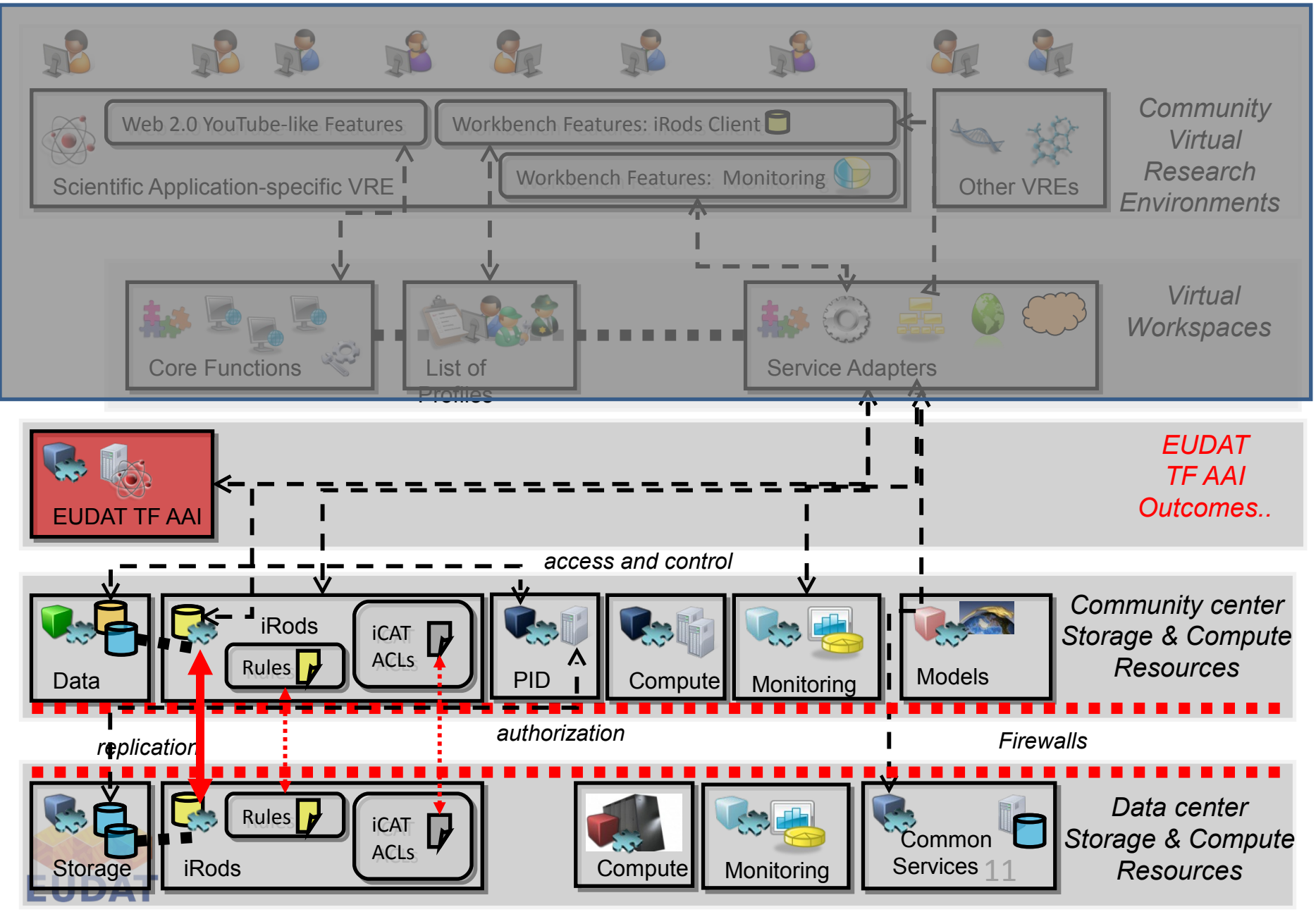
Components

Technologies:

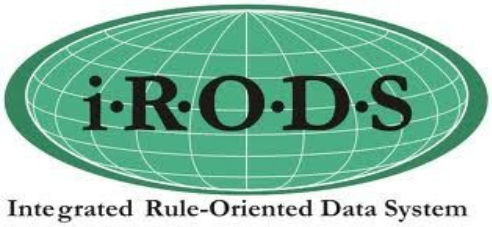
- Long Term Archives → Comunity specific
- Policy-based Replication → iRODS
- Persistent Identifiers → EPIC/Handle

Orthogonal aspects:

- AAI
- Monitoring
- Center Registry
- Metadata


Web 2.0 YouTube-like Features
Workbench Features: iRods Client
Workbench Features: Monitoring
Scientific Application-specific VRE
Other VREs
Community Virtual Research Environments
Core Functions
List of Profiles
Service Adapters
Virtual Workspaces
EUDAT TF AAI
EUDAT TF AAI Outcomes..
access and control
Data
iRods
Rules
iCAT ACLs
PID
Compute
Monitoring
Models
Community center Storage & Compute Resources
replication
authorization
Firewalls
Storage
iRods
Rules
iCAT ACLs
Compute
Monitoring
Common Services
Data center Storage & Compute Resources

iRODS



Integrated **R**ule **O**riented **D**ata **S**ystems

Data grid software system developed by the Data Intensive Cyber Environments (DICE) research group

Deployments: NASA, CC-IN2P3, EU SHAMAN, Australian ARCS, UK e-Science, King's College,...

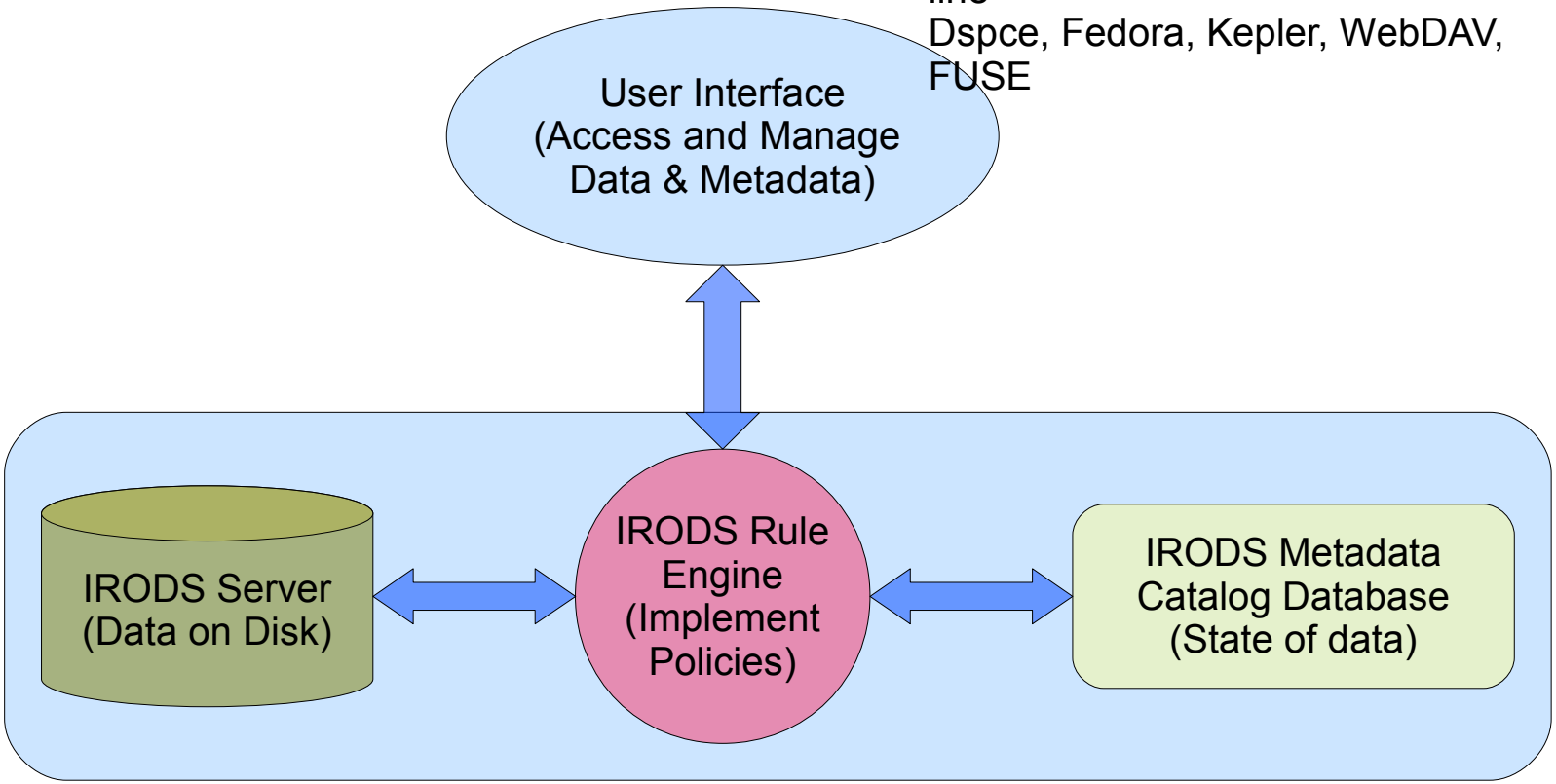
Adaptive Middleware with Rule Oriented Programming (ROP)

- One-size does not fit all
- Community specific operations can be realized by defining **rules**
- Rule: workflows composed of micro services
- Execution: triggered by middelware or user

EUDAT

IRODS Components

Web GUI, iRODS GUI, Command-line
Dspce, Fedora, Kepler, WebDAV, FUSE



iRODS Rules Example

- Format:

```
#action|conditions separated by &&|call_functions separated by ##|
rollback_functions separated by ##
```

- Replication upon ingest:

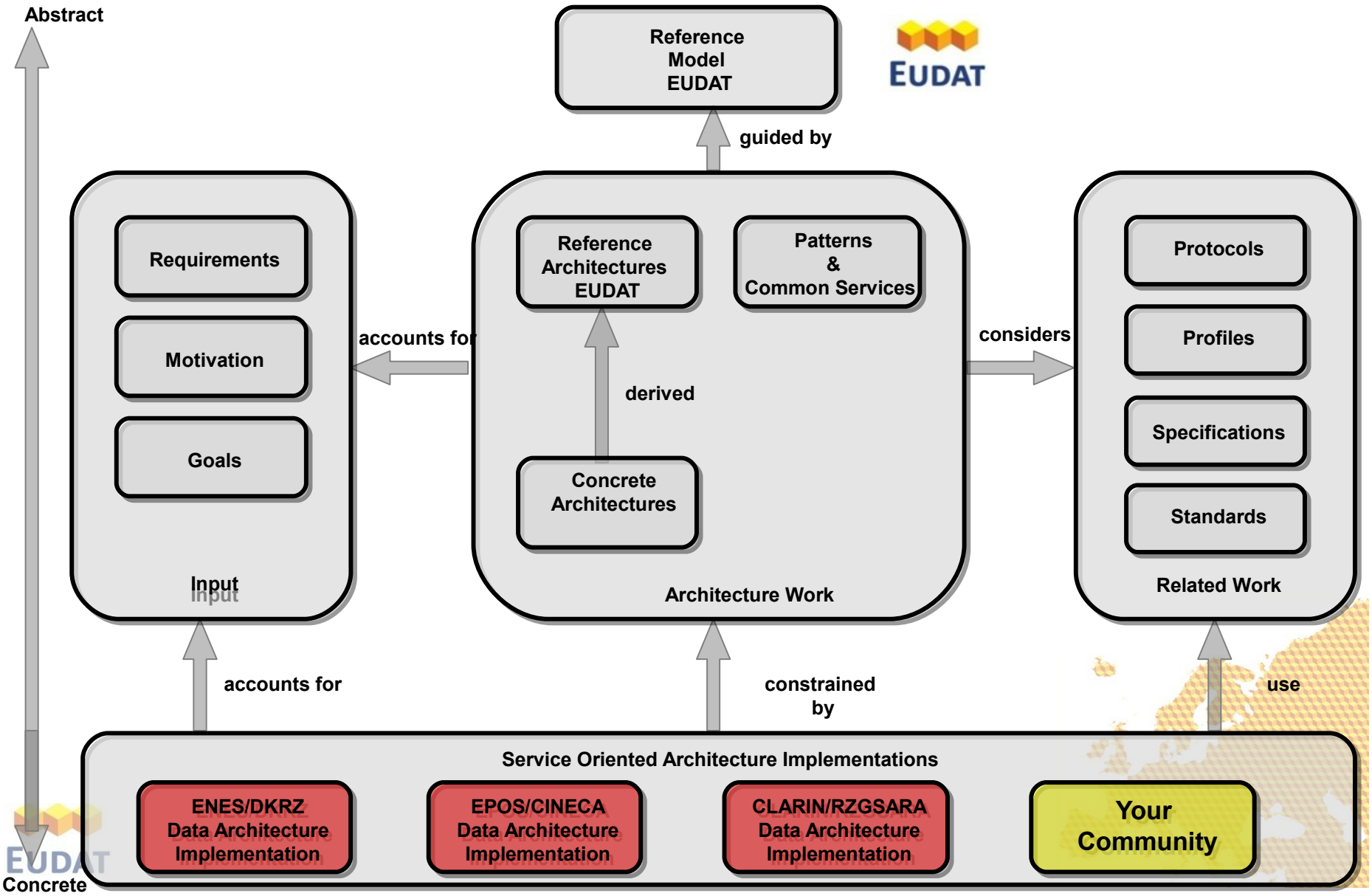
```
acPostProcForPut||msiSysReplDataObj(seq,all)| nop
```

- Example with a workflow of microservices:

```
acCreateUser||msiCreateUser##acCreateDefaultCollections##msiCommit|
msiRollback##msiRollback##nop
```

EUDAT

From Community Solution to EUDAT Architecture



Implementation Plan

Segment 1: Service Building (February 2012: done!)

- Install software (iRODS,...)
- Test replication (Test data, metrics)
- Discuss replication policies

Segment 2: Test and Evaluation (May 2012)

- Test PID registration
- Evaluate performance and accessibility

Segment 3: Production (July 2012)

- Integrate in monitoring
- Produce documentation
- Pass over to production WP5/WP6

EUDAT

Summary

Data Replication Task Force

- Implements data replication service
- No one-size-fits-all-solution is sought after...
  - 80% of the island solution will be common
  - 3*20% of the community specific functionalities
- Easy integration of the new communities:
  - Enjoy basic services
  - Use existing microservices and rules to tailor a solution that suits you

JOIN US!