Policy:
Assertion or assurance that is enforced about a collection or a dataset

Problem

Computer actionable policies

- Enforce management,

- Automate administrative tasks,

- Validate assessment criteria,

- Automate scientific analyses

- etc.

A generic set of policies that can be revised and adapted by user communities and site managers does not exist.

- ***Domain scientists*** who want to build-up a collection or a repository

- ***Data centers*** for automating policies

Results

Identification of 11 important policy areas:

- Contextual metadata extraction
- Data access control
- Data backup
- Data format control
- Data retention
- Disposition
- Integrity (including replication)
- Notification
- Restricted searching
- Storage cost reports
- Use agreements

Results

Working Group: **Practical Policy**

Outcomes Policy Templates: Practical Policy Working Group, September 2014

Version: August 24, 2014

https://www.rd-alliance.org/filedepot?cid=104&fid=556

Templates

- Interactions of policies and DO attributes
- Policy descriptions
- Technology independent
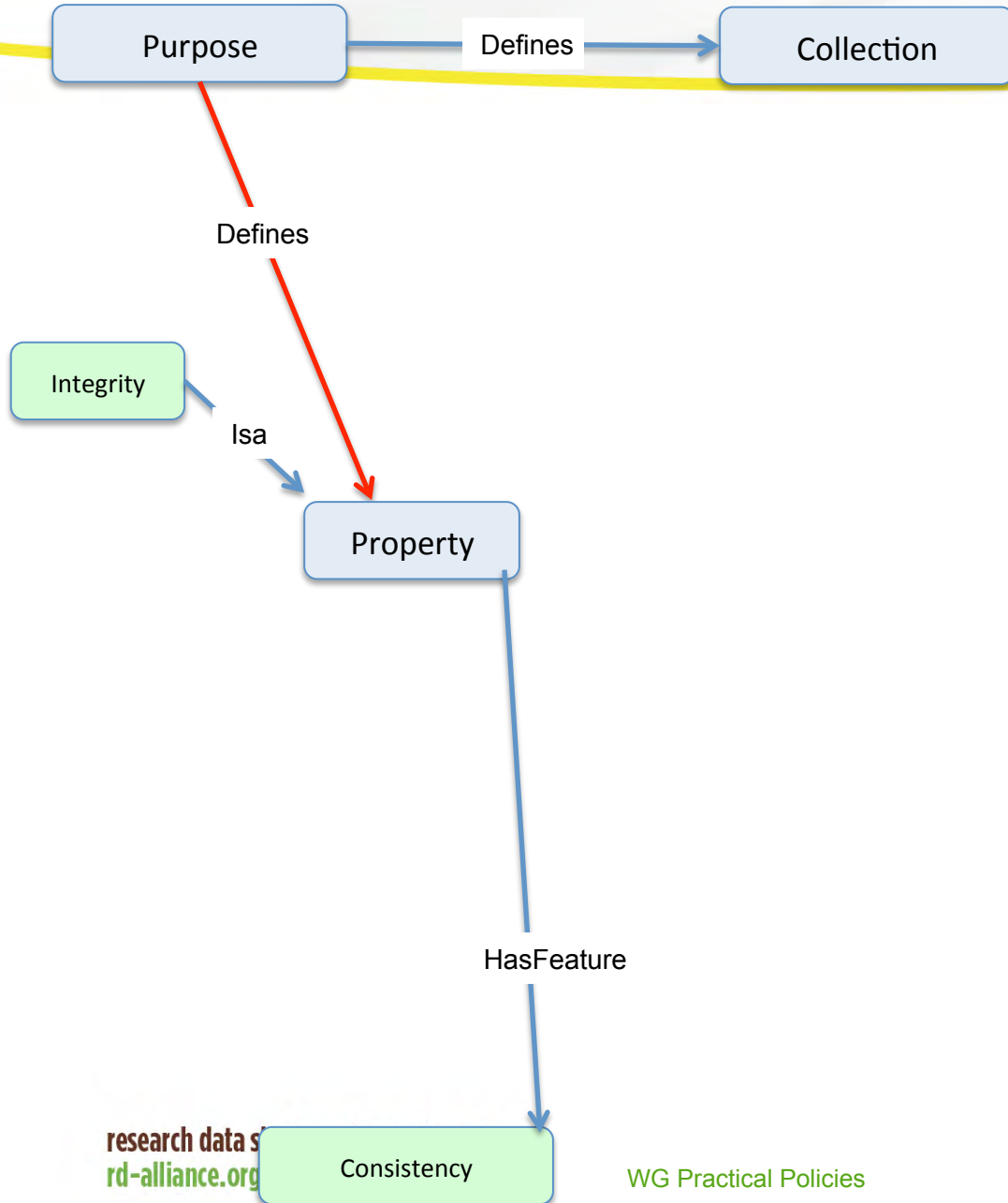- Reviews of the provided policy areas in progress

Results

Working Group: **Practical Policy**

Implementations:
Practical Policy Working Group,
September 2014

Version: August 24, 2014

https://www.rd-alliance.org/filedepot?cid=104&fid=553

- Examples for implementations:
  - English language descriptions
  - iRODS
  - GPFS
- ~50 pages

research data sharing without barriers
rd-alliance.org   WG Practical Policies   RDA
                                                        RESEARCH DATA ALLIANCE   6
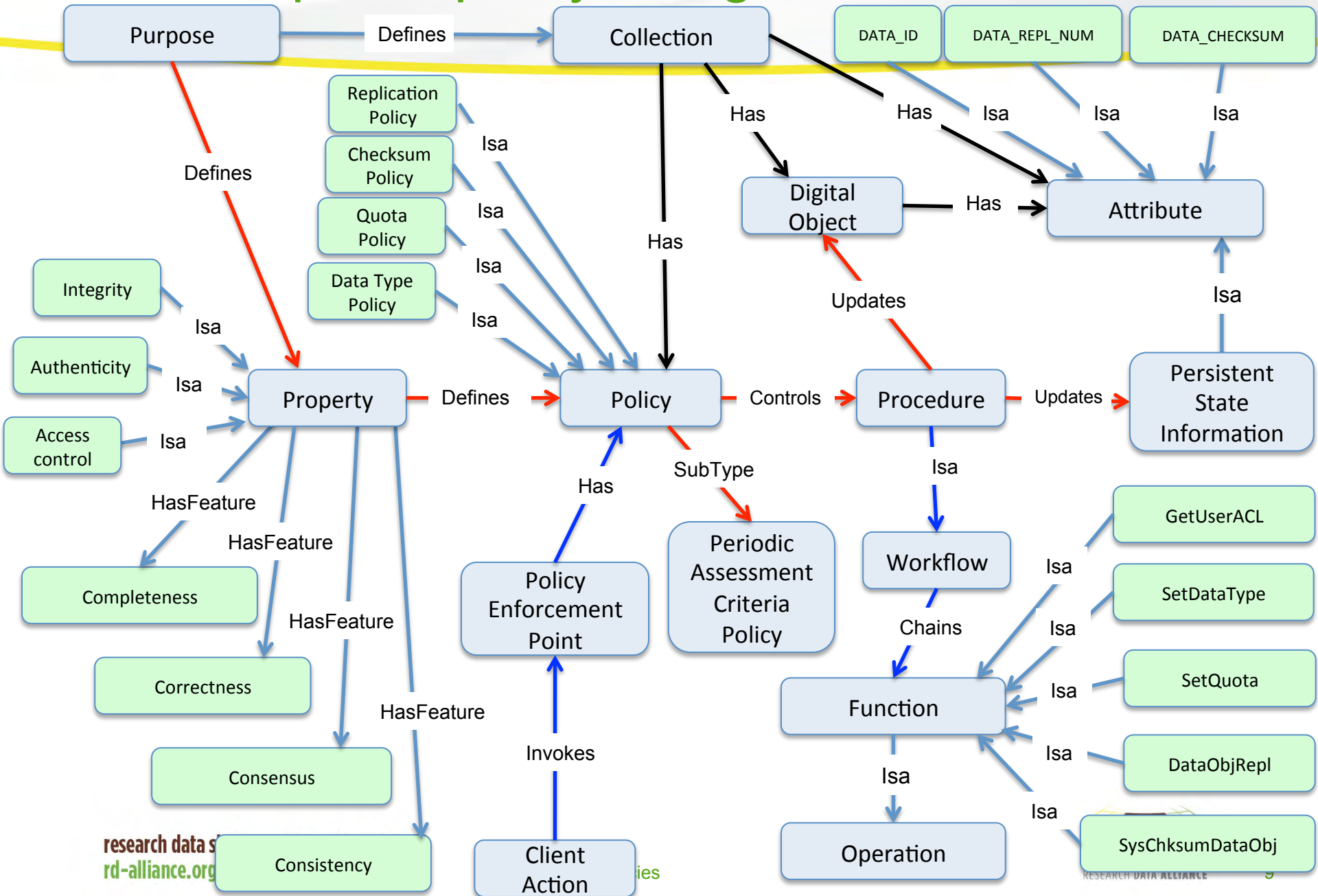
Concept Graph by Reagan Moore

research data s
rd-alliance.org

Concept Graph by Reagan Moore
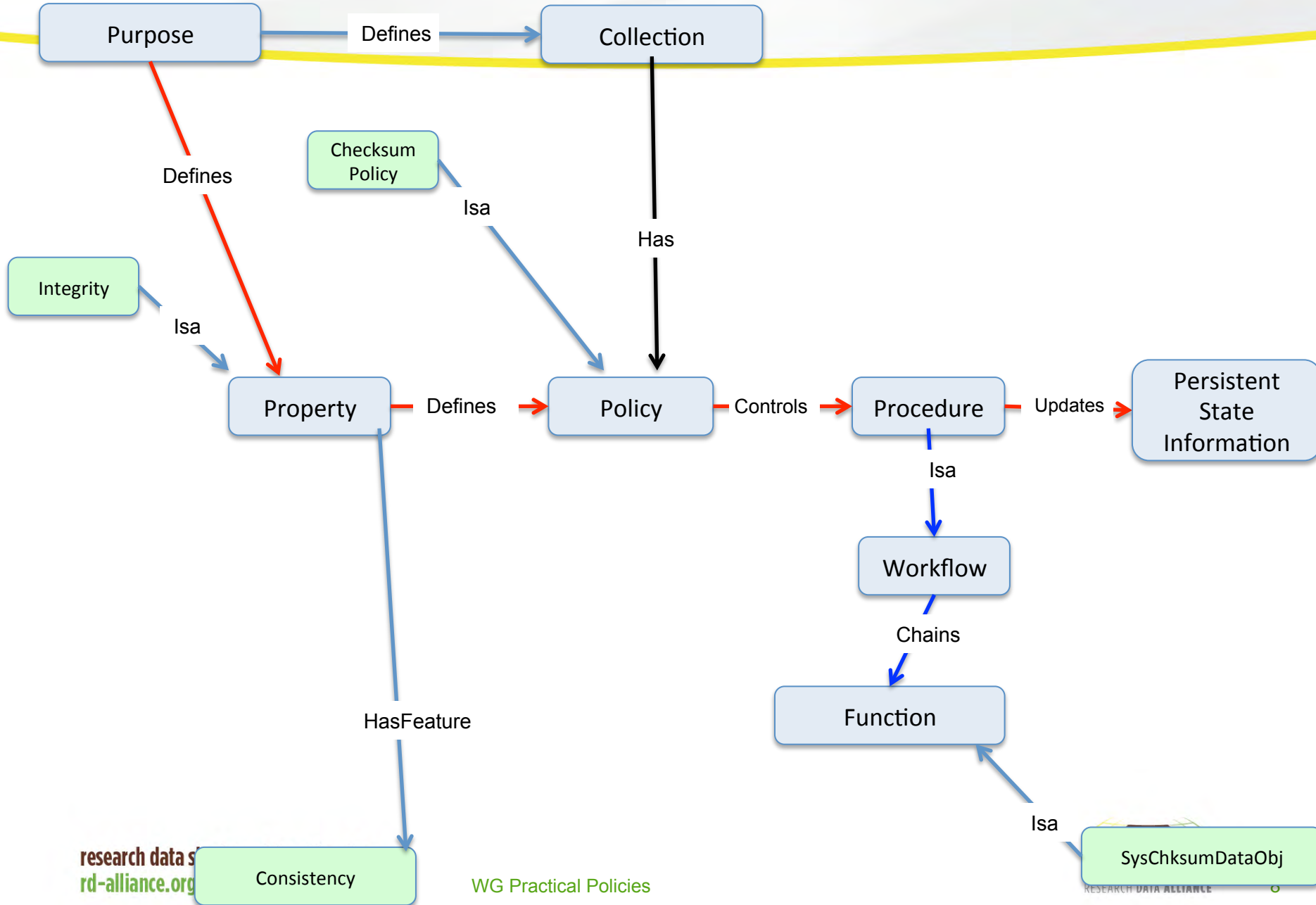
Concept Graph by Reagan Moore

Example: Integrity policy

Scenarios:

- Verification of integrity on ingestion through validation of a checksum
- Replication of the file across multiple storage locations to ensure the ability to replace a corrupted file.
- Periodic verification that the files are not corrupted, that the required number of replicas exist, and that the replicas are correctly distributed.

Integrity	Constraint	State attributes for Constraint
	Periodic check	Time_interval
	For collection	Collection_name
	For file	File_name
	For files on storage system	Storage_system_name
	Operations	**State Attributes for Operation**
	Set checksum	File_ID
		File_checksum
	Create replica	File_ID
		File_replica_number
		File_replica_location
		File_replica_creation_time
		File_physical_path_name
		File_replica_checksum
	Verify checksum	File_ID
		File_checksum
	Verify replicas	File_id
		File_replica_number
		File_replica_checksum
		File_storage_location

Implementation: iRODS

research data sharing without barriers
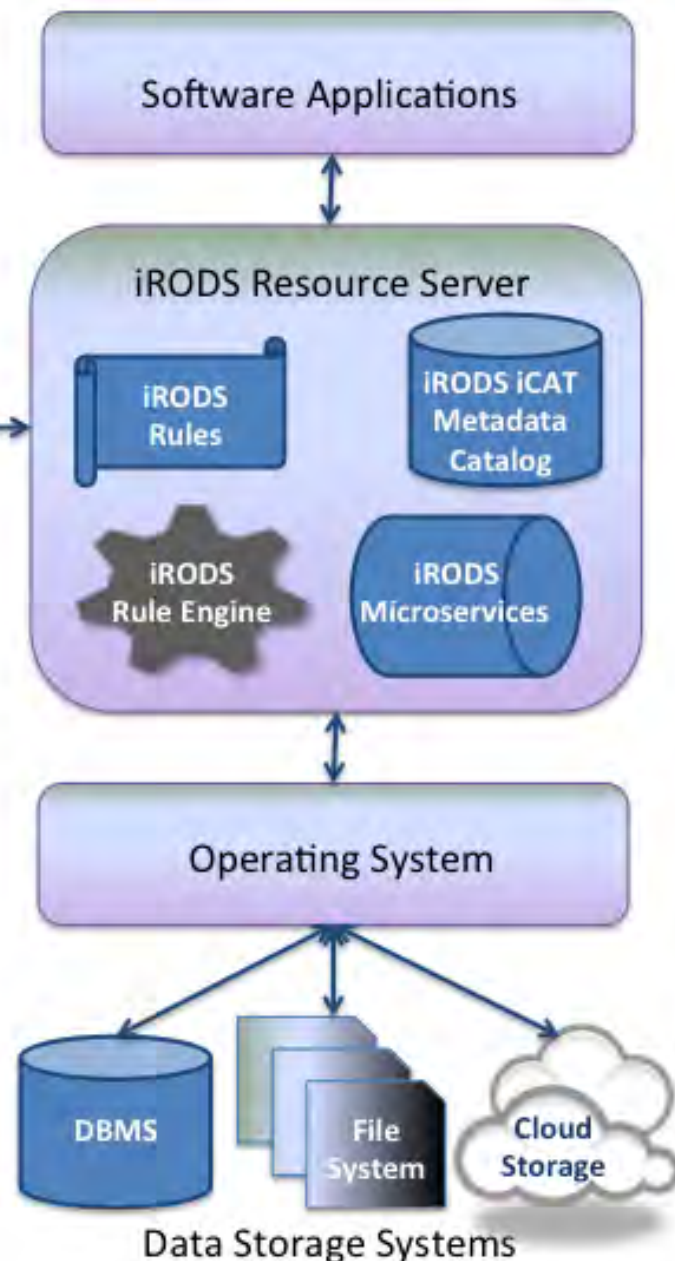rd-alliance.org
WG Practical Policies
12
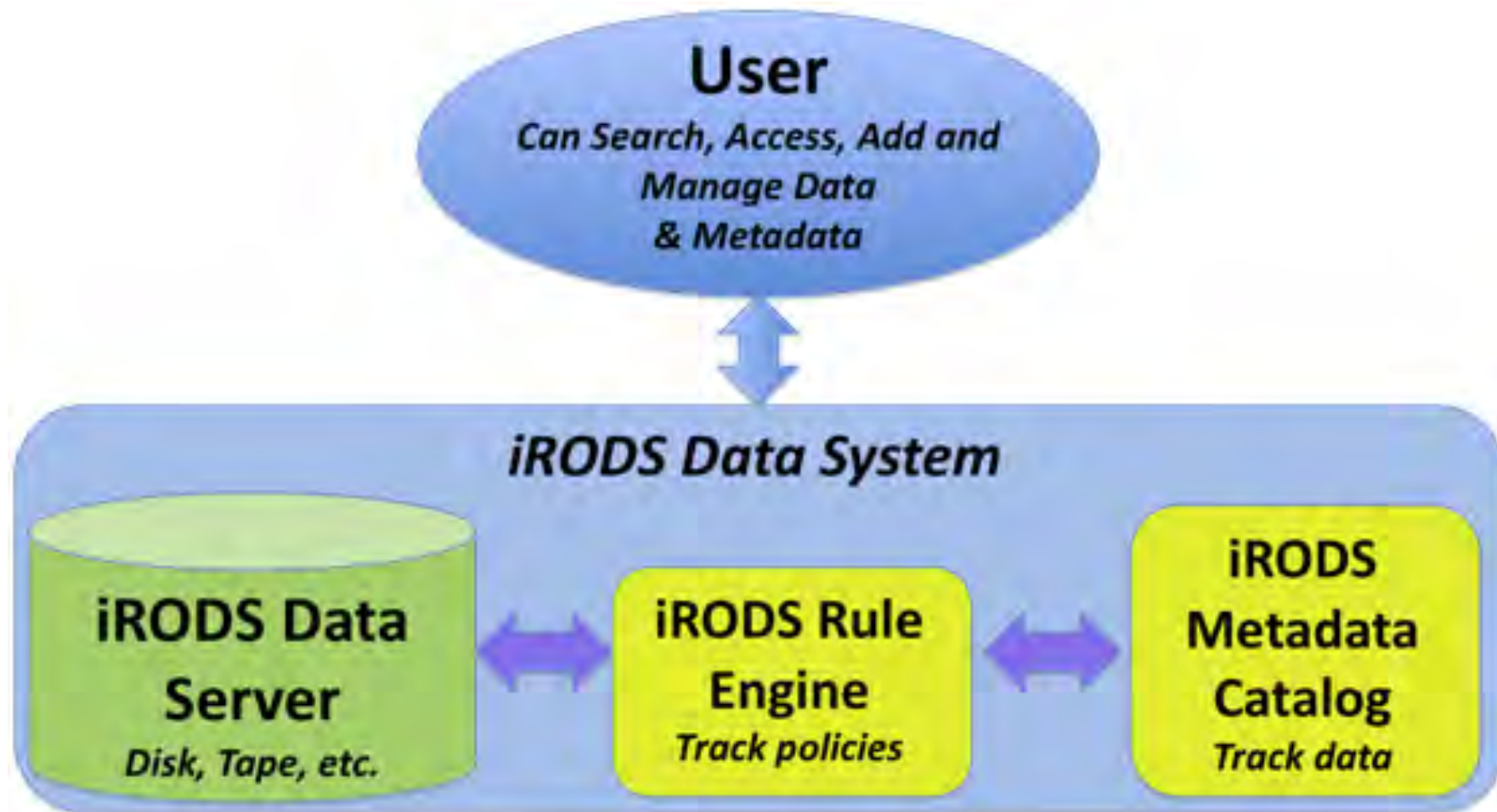
Implementation: iRODS

iCAT contains:

- **User domain related MD**

- **System related MD**
  - User namespace
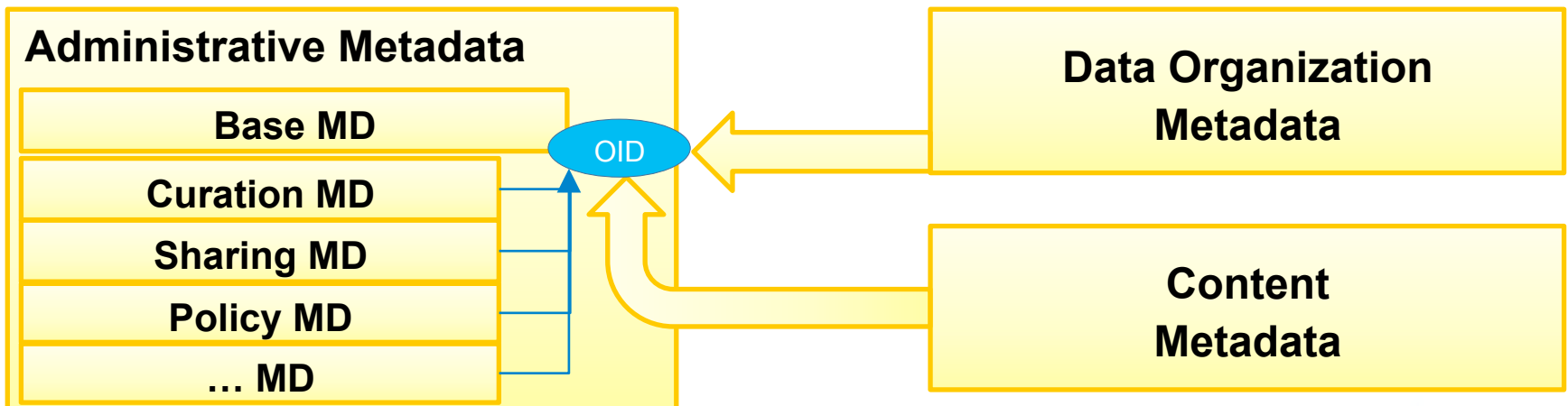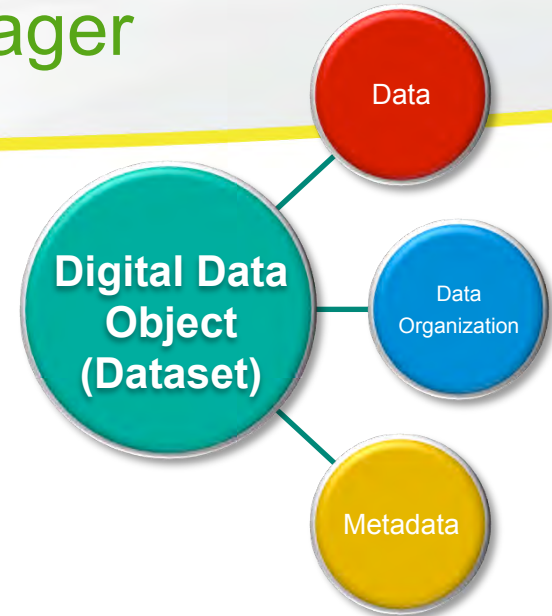 - File namespace
  - Storage namespace

Repository System: KIT Data Manager

Repository

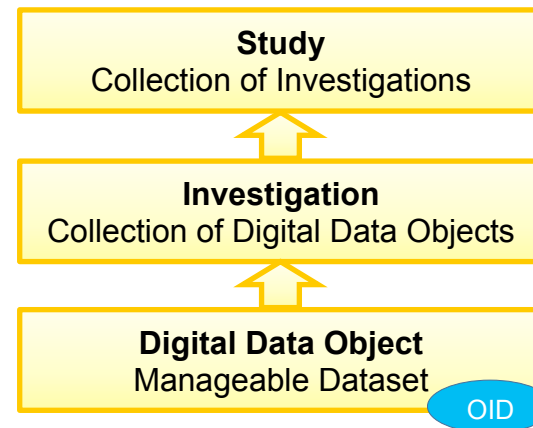Managed location/destination/directory/ bucket where **digital data objects** are

- Registered
- Permanently stored
- Made accessible and retrievable
- Curated

Policy Metadata

Policy MD

enable ***Integrity Policy***	
Periodic check	Time_interval
For digital object	OID
On user	User_ID
Set checksum	File_ID1
	File_ID1_chksum
	…

enable ***Data access control***
enable ***Data backup***
enable ***Data format control***
enable ***Data retention***
enable ***Disposition***
+++
+++

Study
Collection of Investigations

⬆

Investigation
Collection of Digital Data Objects

⬆

Digital Data Object
Manageable Dataset

OID

Conclusions

- Policy information carried by metadata
- → Integration in the Data Fabric

- Templates provide crude structure of a vocabulary
- → Creation of a human+machine accessible vocabulary

For information please contact
- Reagan Moore rwmoore@renci.org and
- Rainer Stotzka rainer.stotzka@kit.edu