



Universiteit Utrecht

iRODS and Yoda - introduction

Ton Smeele - ITS/ResearchIT, Utrecht University



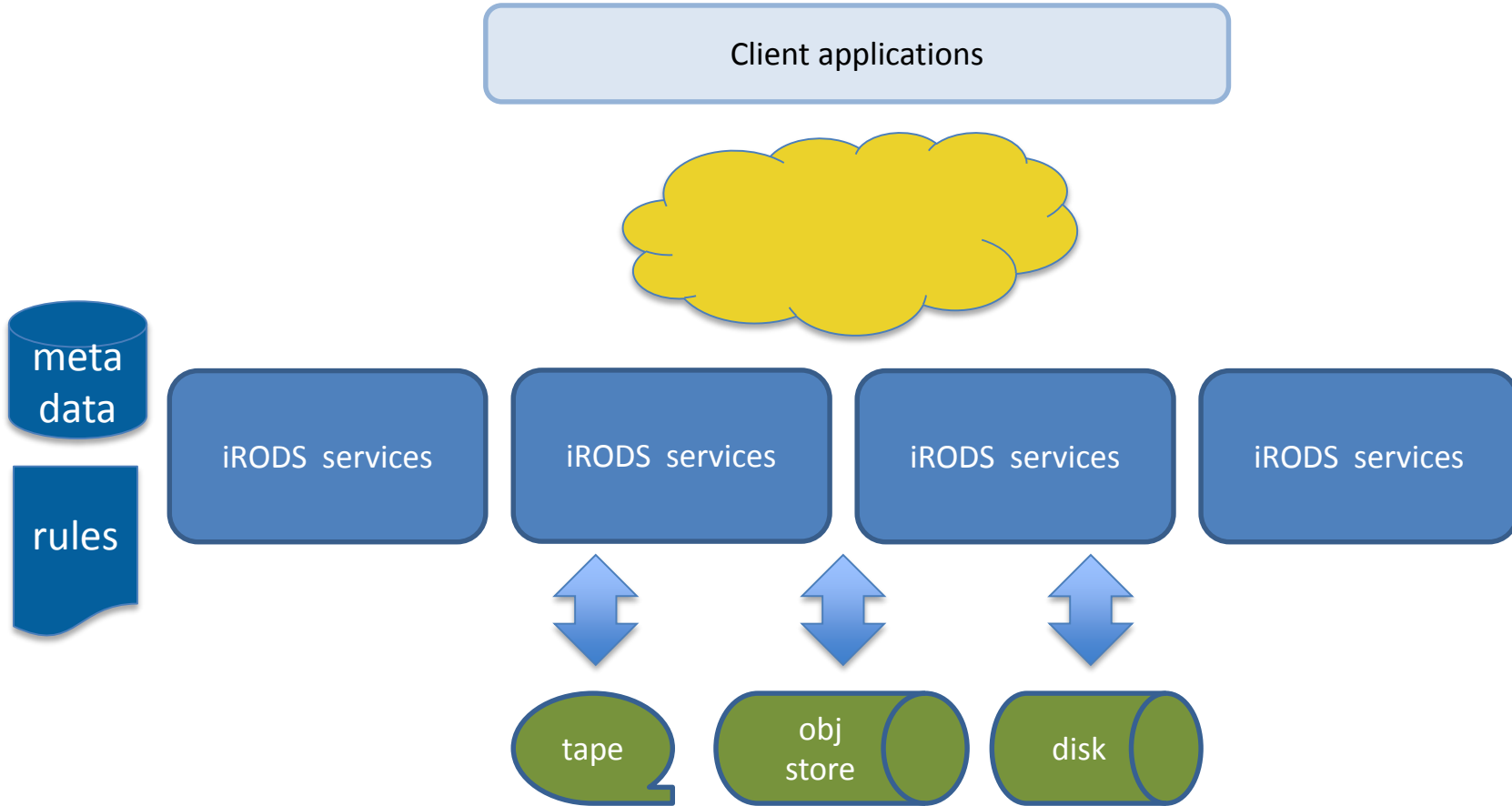
What is iRods

- **data grid** software, to logically order and use large amounts of *data* along with their *metadata* from sources spread across multiple locations, computers and storage infrastructures
- An **integrating infrastructure** situated in the layer between user applications and operating system/storage facilities
- Includes services that allows users to access, manipulate and manage data in a **uniform way**, annotate the data with metadata and share data and metadata with other users.

Support for metadata is crucial when you need to manage your research data



iRODS unifies and manages access to persistent data





iRODS Rule Engine: operationalize policies

Policies:

- Ensure that *primary* research data is kept safe in at least two geographically spread locations managed by the university.
- Ensure data is properly described and accounted for.

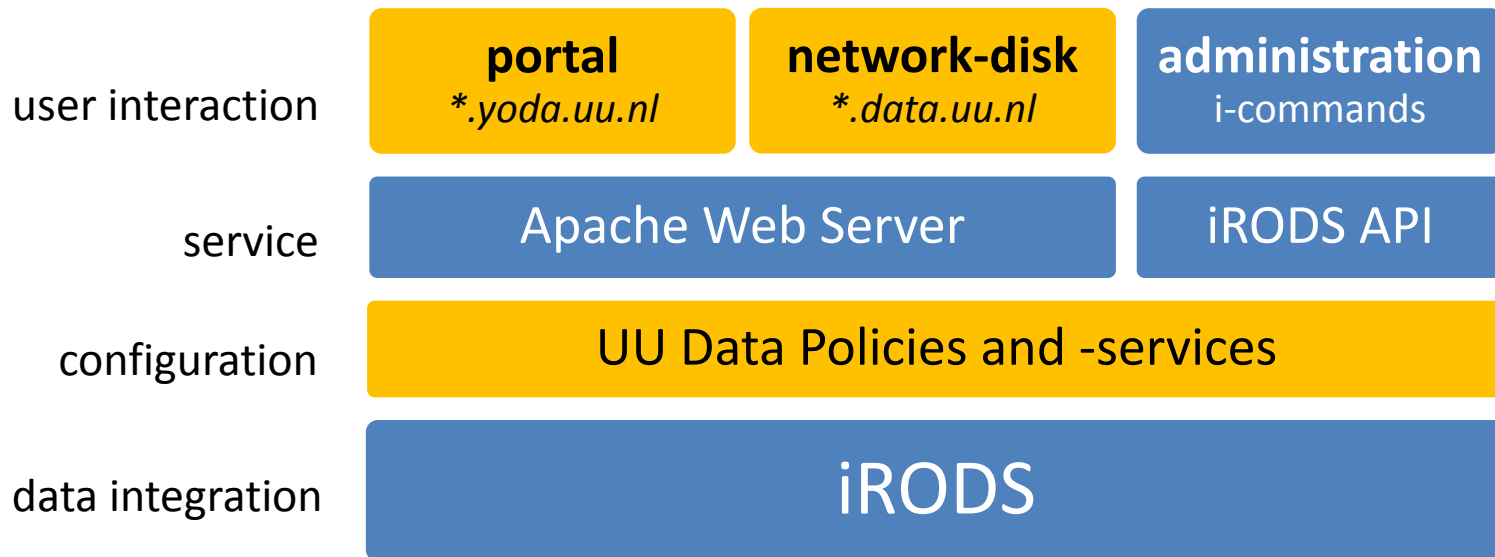
Partial rule:

GIVEN (a new data file is being added)
WHEN (file has metadata "primary research data") , (file is of type image)
THEN
(**auto-extract** the geo-location, date using the file's EXIF properties) **AND**
(**store the file** on our own Utrecht located bulk storage system) **AND**
(**store a replica** of the file on our Almere located storage system) **AND**
(**maintain** geo-location, actor and date as **metadata with the file**)



Our iRODS implementation is called "Yoda":

preconfigured iRODS based system, delivered and supported as a service
– enhanced with (graphical) user interfaces, policies and rules





Manage data during (and after) research



Collaborate safely as a group ("Research" folder)
-> membership self-managed by researchers



Maintain integrity, deposit a folder in the vault
-> metadata can vary per community,
-> datamanager approves deposit



Allow FAIR reuse, publish a data package
-> datamanager approves publication, DOI citable data



demonstration



Utrecht University examples of iRODS deployment

- **primary research data management: YOUTH**
 - centrally manage storage and integrity of lab experiment data
- **manage, publish and find interdisciplinary research data sets: I-Lab**
 - interdisciplinary collaboration on "Institutions" related data
- **research study collaboration: Equine research**
 - international collaboration with concurrent access to data
- **archive: Cell Biology**
 - safely store many terrabytes of research data along with metadata
- **backup: NMR**
 - daily synchronization research data of local server to Yoda



What we have learned so far

iRODS is a very versatile and useful toolbox for research data management
... and its successful initial deployments generates much traction in our country

Out-of-the-box it lacks a user interface, hence you need to supplement it with other components

It requires skilled staff to configure iRODS (scalability, rules)
..... and locally those skills are not yet widely available

We can benefit from each other's experience and results -> U2CONNECT.ORG

So how do we all scale up and benefit from iRODS



Announced at iRODS UGM June 2017:

SURF builds an iRODS Expertise Center

- **Team of iRODS experts already recruited/developed**
- **Initiative to provide national level iRODS services**
 - iRODS based scale out/archival storage
 - assistance on iRODS implementation and development of rules
 - hosted iRODS services
 - European Data infrastructure (EUDAT) integration



Curious? Let's get you started!

Thank you

More info:

Ton Smeele - a.p.m.smeele@uu.nl

Frank Heere - frank.heere@surfsara.nl