ASPiS

integrating iRODS with Shibboleth and provenance engines

Eric Liao, Mark Hedges, Tobias Blanke
King's College London

Andrea Weise, Adil Hasan, Jens Jensen
University of Reading, University of Liverpool, Science and Technology Facilities Council

iRODS Workshop @ CC-IN2P3, 2009

Outline

Project Overview

- Funded by JISC e-Infrastructure programme.
- Partners:
  - Centre for e-Research, King's College London
  - University of Liverpool
  - Science and Technology Facilities Council
  - (University of Reading - very helpful PhD student)
- Project Goals:
  1. access management in iRODS - integration with Shibboleth (and authorisation systems such as PERMIS).
  2. integration of iRODS with provenance capture systems.

Project Overview

- Funded by JISC e-Infrastructure programme.
- Partners:
  - Centre for e-Research, King's College London
  - University of Liverpool
  - Science and Technology Facilities Council
  - (University of Reading - very helpful PhD student)
- Project Goals:
  1. access management in iRODS - integration with Shibboleth (and authorisation systems such as PERMIS).
  2. integration of iRODS with provenance capture systems.

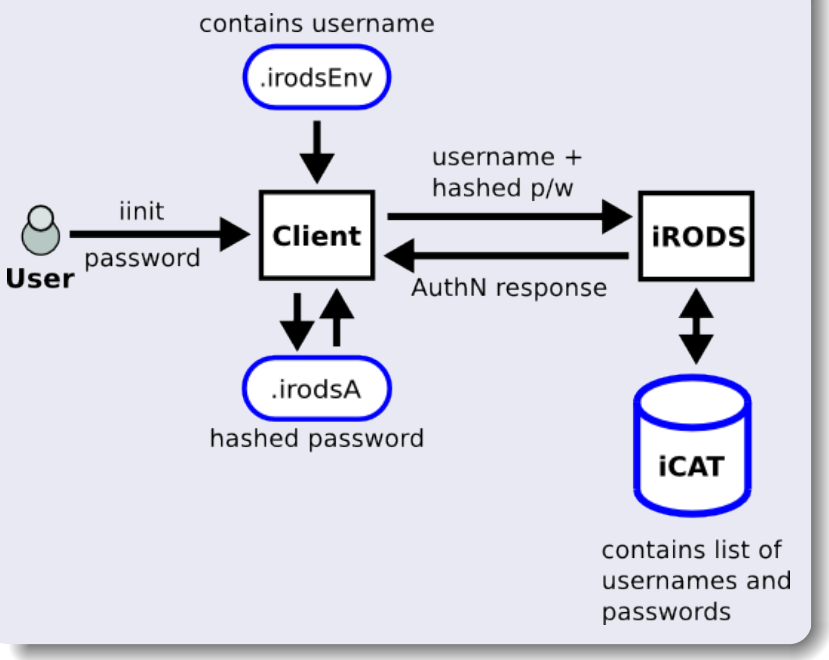
Project Overview

- Funded by JISC e-Infrastructure programme.
- Partners:
  - Centre for e-Research, King's College London
  - University of Liverpool
  - Science and Technology Facilities Council
  - (University of Reading - very helpful PhD student)
- Project Goals:
  1. access management in iRODS - integration with Shibboleth (and authorisation systems such as PERMIS).
  2. integration of iRODS with provenance capture systems.

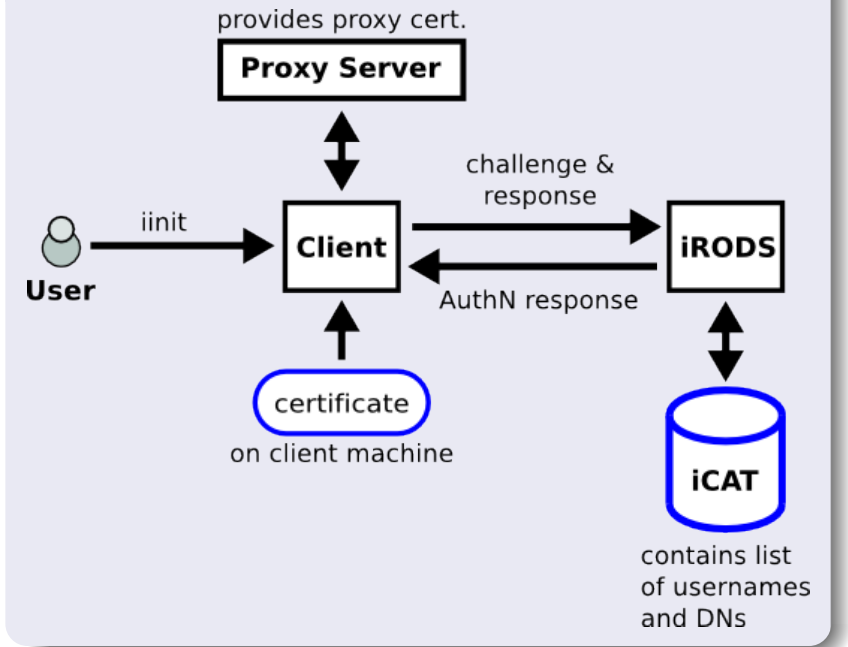
Outline

iRODS Authentication

Username + Password



GSI



iRODS Authorization

- iCAT stores information on:
  - Users
  - Domains
  - Groups
  - Access Control Lists (ACLs)
- Access managed according to:
  - Mode of access (read / write / delete / annotate)
  - By user, domain, group
- Information held centrally

iRODS Authorization

- iCAT stores information on:
  - Users
  - Domains
  - Groups
  - Access Control Lists (ACLs)
- Access managed according to:
  - Mode of access (read / write / delete / annotate)
  - By user, domain, group
- Information held centrally

iRODS Authorization

- iCAT stores information on:
  - Users
  - Domains
  - Groups
  - Access Control Lists (ACLs)
- Access managed according to:
  - Mode of access (read / write / delete / annotate)
  - By user, domain, group
- Information held centrally

Observed Issues

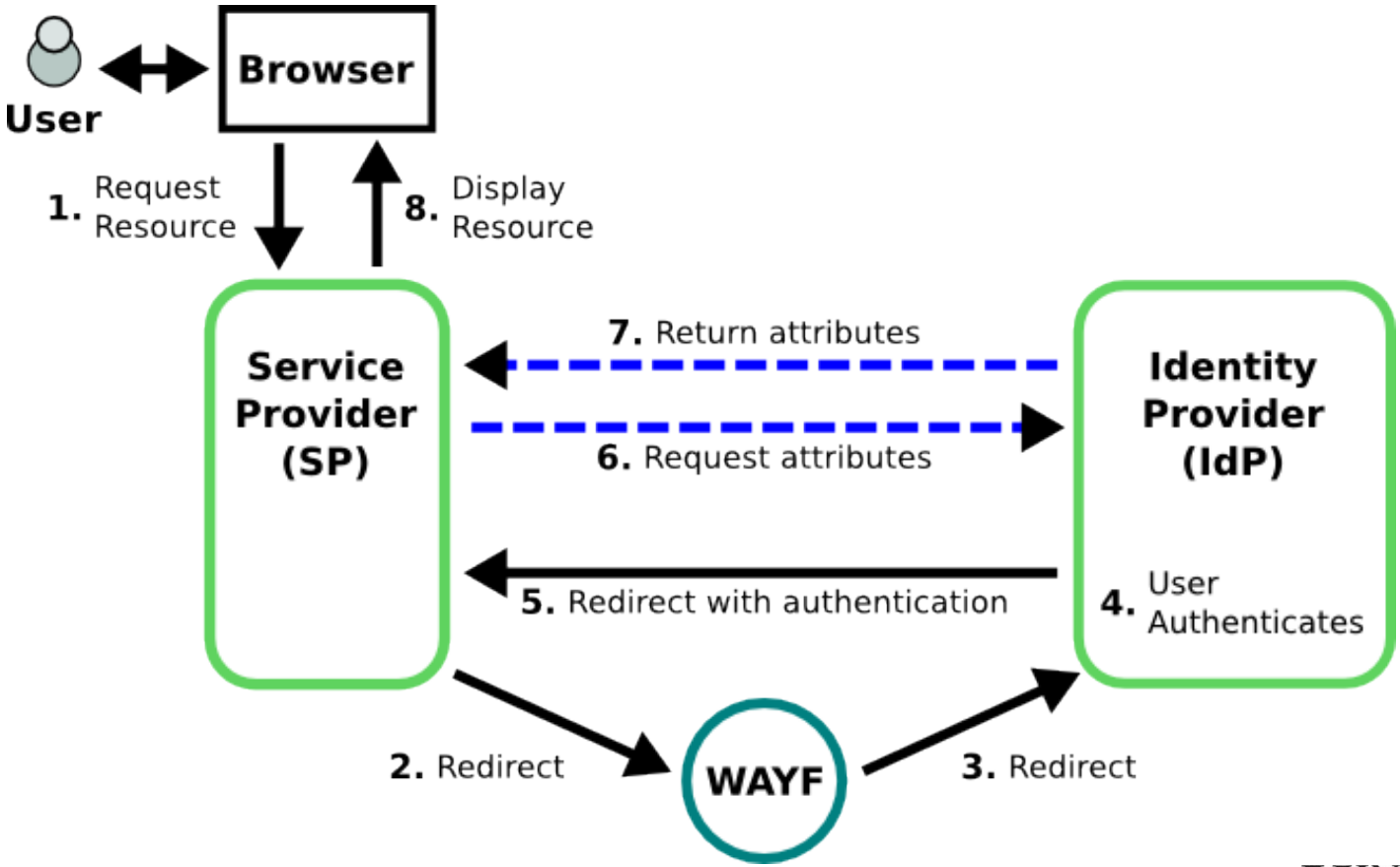
- Centralised management of user identities and access rights
- Doesn't scale well
- Different organisations cannot maintain their own lists of users in data grid - duplication, lists can get out of sync
- Inflexible authorisation system - no locally managed admin of access rights
- Certificates a barrier to uptake of grids in some communities

Outline

Shibboleth Overview

- Architecture for federated access to web based resources
- Based on circle of trust among organisations
- User identities managed locally to their institution
- Access to resources managed locally to the owning institution
- Adopted by JISC as a solution for managing access to distributed web resources

Shibboleth Information Flow



UK Federation

- **UK Access Management Federation for Education and Research**
  - Based on SAML (**S**ecurity **A**ssertion **M**arkup **L**anguage)
  - Provides a single access solution to online resources/services
  - Metadata based on the Internet2 eduPerson LDAP schema
- Core Federation eduPerson attributes
  - *ScopedAffiliation* → staff@kcl.ac.uk, visitor@stfc.ac.uk
  - *TargetedId* → idp.kcl.ac.uk!sp.stfc.uk!<opaque string>
  - *PrincipalName* → eric.liao@kcl.ac.uk
  - *Entitlement* → urn:mace:ac.uk:irods.stfc.ac.uk:visitor

UK Federation

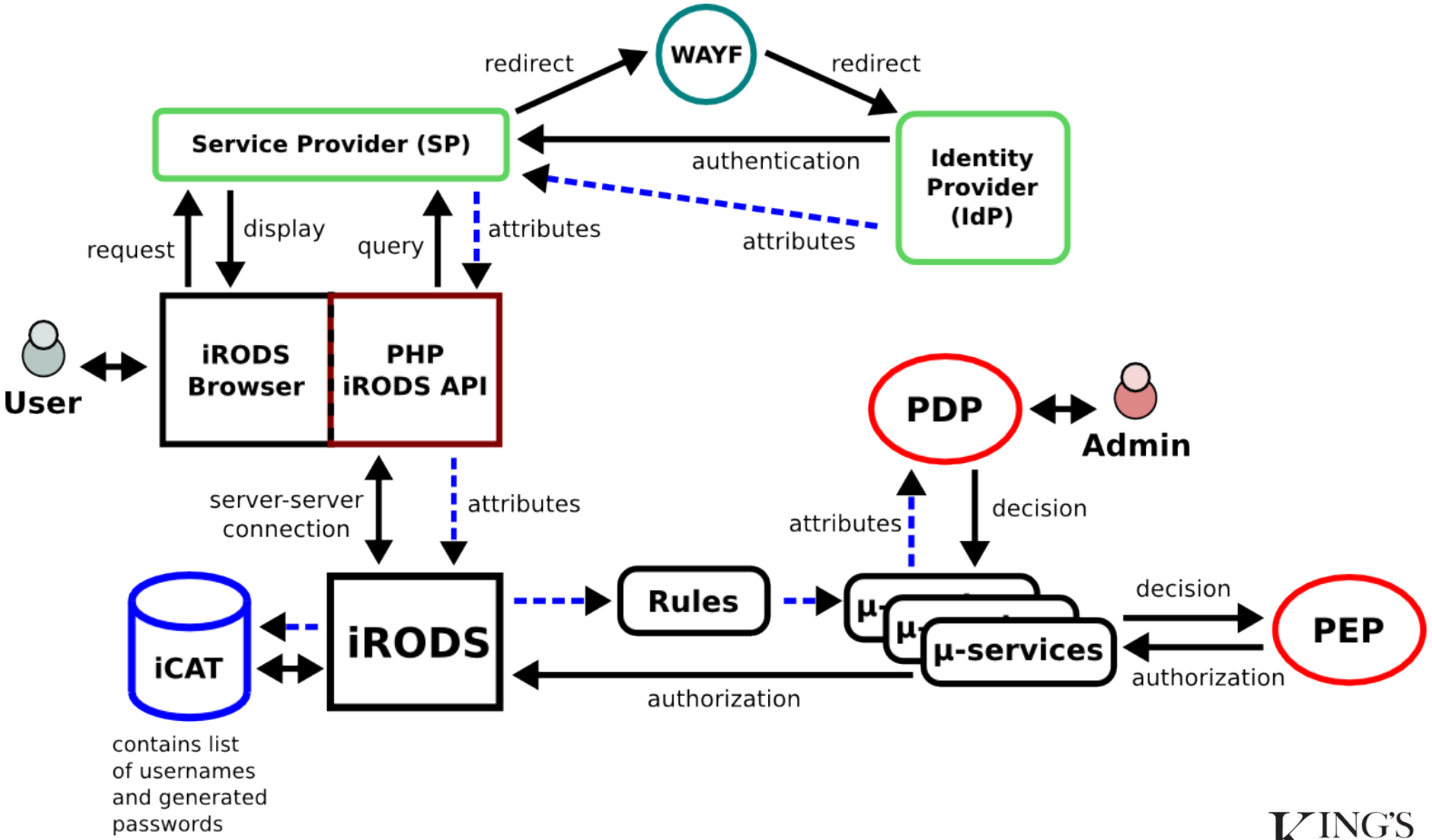
- **UK Access Management Federation for Education and Research**
  - Based on SAML (**S**ecurity **A**ssertion **M**arkup **L**anguage)
  - Provides a single access solution to online resources/services
  - Metadata based on the Internet2 eduPerson LDAP schema
- **Core Federation eduPerson attributes**
  - *ScopedAffiliation* → staff@kcl.ac.uk, visitor@stfc.ac.uk
  - *TargetedId* → idp.kcl.ac.uk!sp.stfc.uk!<opaque string>
  - *PrincipalName* → eric.liao@kcl.ac.uk
  - *Entitlement* → urn:mace:ac.uk:irods.stfc.ac.uk:visitor

Outline

Access Control Requirements

- Devolve authentication service to user's home institution
- Common interface layer to decouple authorization services
- Access control allowing fine-grained access rights to be defined for roles, not just user identities
- No interference to iRODS core system

Access Control Architecture



Outline

Overview

- Provenance $\longrightarrow$ history of operation applied to a digital object
- Provenance is an important issue
  - Gives history of events
  - Allows to verify the authenticity of data
  - Determines quality of data
  - Supports researchers in many ways (e.g. re-executing experiments)

Provenance in iRODS

- iRODS does not capture changes made to data
- iRODS's metadata is not sufficient to capture workflows

Overview

- Provenance $\longrightarrow$ history of operation applied to a digital object
- Provenance is an important issue
  - Gives history of events
  - Allows to verify the authenticity of data
  - Determines quality of data
  - Supports researchers in many ways (e.g. re-executing experiments)

Provenance in iRODS

- iRODS does not capture changes made to data
- iRODS's metadata is not sufficient to capture workflows

Overview

- Provenance $\longrightarrow$ history of operation applied to a digital object
- Provenance is an important issue
  - Gives history of events
  - Allows to verify the authenticity of data
  - Determines quality of data
  - Supports researchers in many ways (e.g. re-executing experiments)

Provenance in iRODS

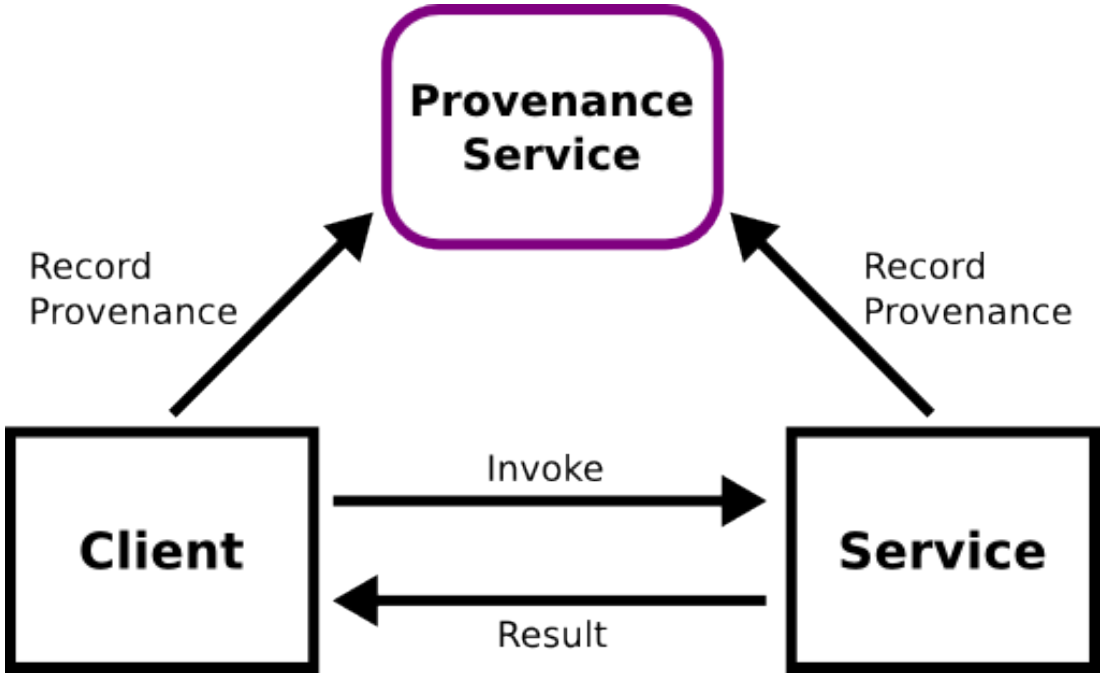
- iRODS does not capture changes made to data
- iRODS's metadata is not sufficient to capture workflows

Key Requirements

- Manage data throughout its lifecycle
- Capture and record information about the data analysis
- Enforce ownership of data thoughout its lifetime
- Ensure data access is auditable
- Ensure infrastructure is robust and scalable

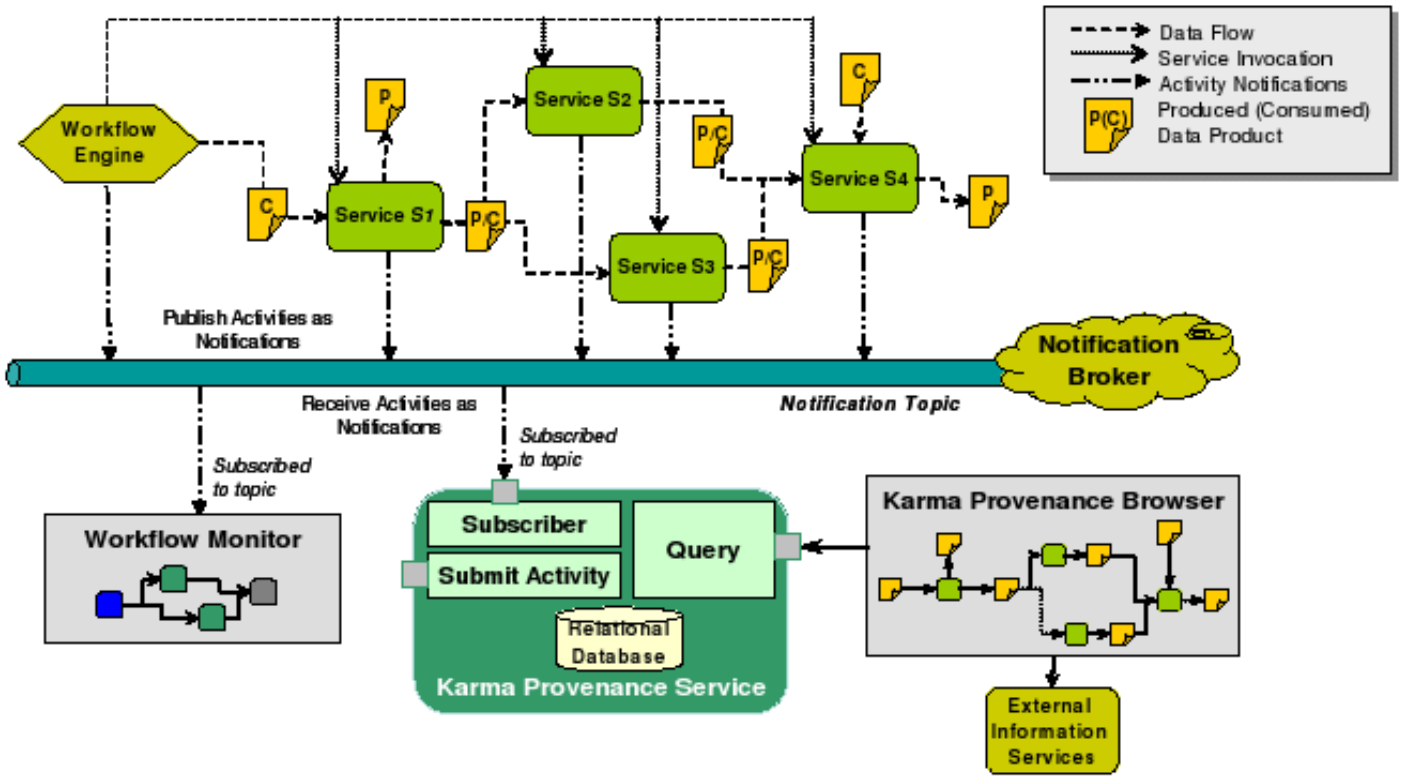
Outline

PASOA



- Independent protocols for recording and accessing provenance

Karma



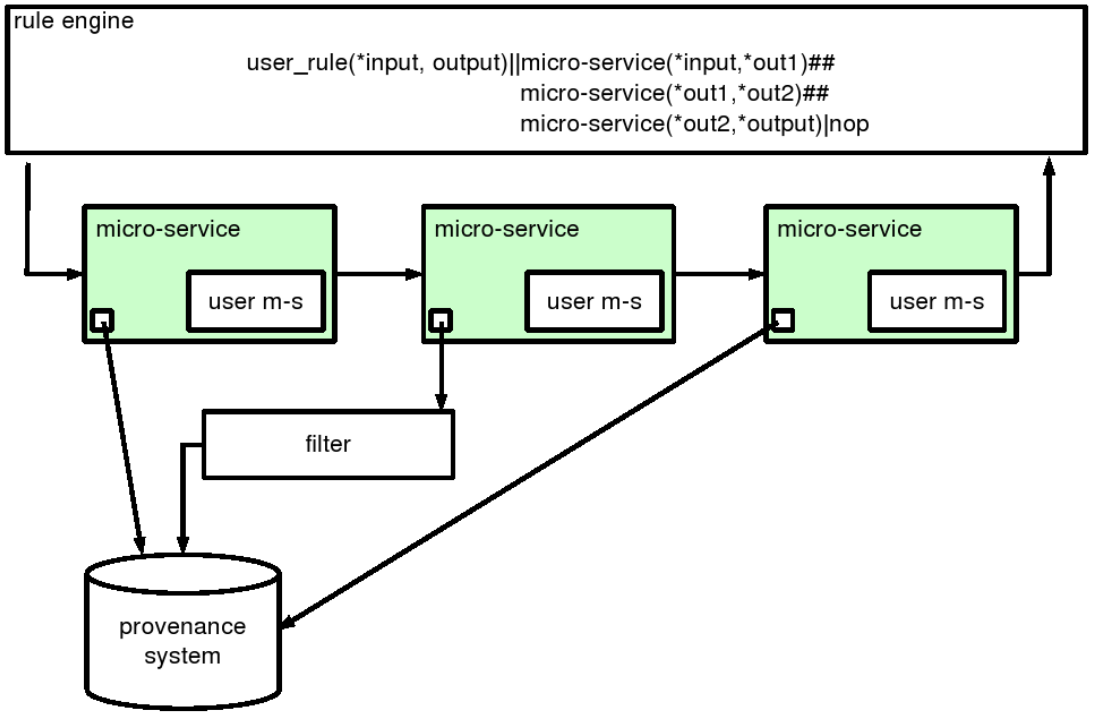
- Publish-subscribe notification protocol

Outline

Provenance System Requirements

- Meet provenance requirements
- No interference with iRODS core system
- Provenance system should be applicable for any other system
- Easy to use
- Eliminate single point of failure within PASOA and Karma

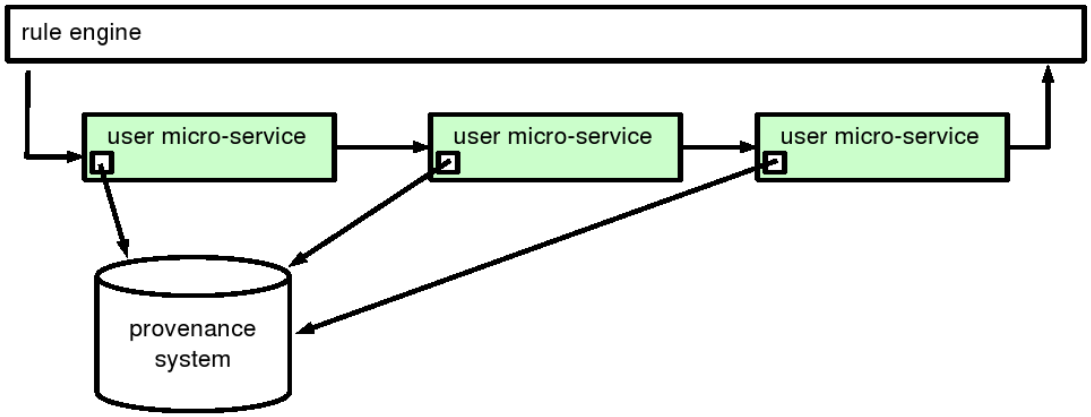
Provenance System Design Ideas



Microservice Wrapper

- Embed user microservice in provenance microservice
- Capturing all information
- User microservice has to be modified

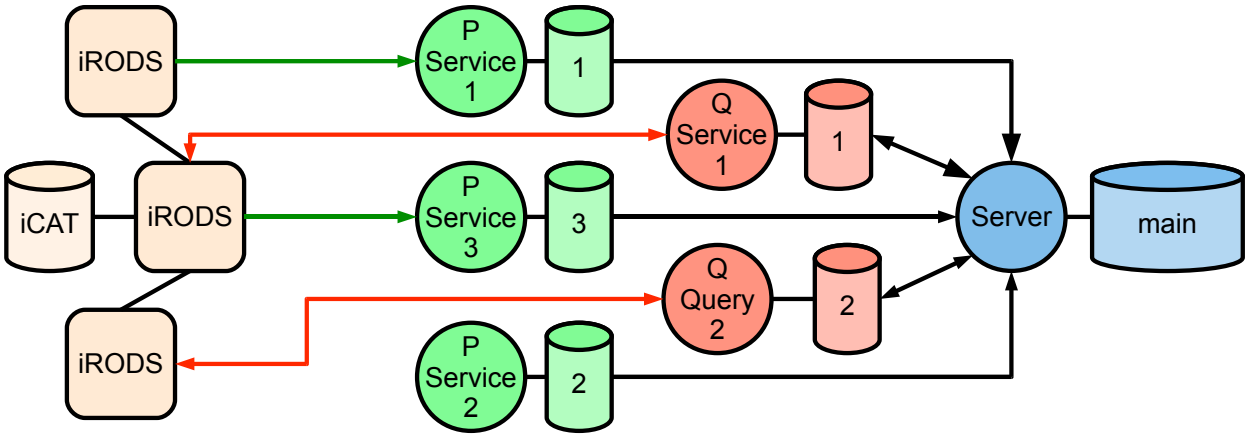
Provenance System Design Ideas



Microservice Chain

- Embed provenance microservice in user microservice
- Only specific data is captured
- User deals with capturing

A Provenance Framework



- Recording service (P-Service) + Querying service (Q-Service)
- Balanced distributed web service lookup system

Work so far & Future plans

Completed Work

- Liased with potential users and determined initial use cases
- Developed prototypes for iRODS-Shibboleth integration
- Developed prototypes for iRODS-Provenance integration

Future Work

- Refine prototypes of access control and provenance systems
- Integration of access control and provenance systems
- Testing with use cases

Work so far & Future plans

Completed Work

- Liased with potential users and determined initial use cases
- Developed prototypes for iRODS-Shibboleth integration
- Developed prototypes for iRODS-Provenance integration

Future Work

- Refine prototypes of access control and provenance systems
- Integration of access control and provenance systems
- Testing with use cases

Contacts

eric.liao at kcl.ac.uk

mark.hedges at kcl.ac.uk

tobias.blanke at kcl.ac.uk

a.hasan at rl.ac.uk

j.jensen at stfc.ac.uk