



# “EUDAT 1<sup>st</sup> Training Days”

## 25 – 26 June 2012

- Discuss EUDAT views and share knowledge
- Get input on requirements for CDI
- Get input on training requirements

# Community Service Wishes

## Services to become available in autumn

- Safe Data Replication  
(for Bit-stream Preservation & Access Optimization)
- Dynamic Data Staging into HPC Workspace

**iRODS  
is used**

## Launch of these Services later

- Aggregated EUDAT Metadata Domain
- Researcher SimpleStore (Simple Upload, Share and Access)

## In Progress as Research Issues (WP7)

- more elaborate policy rules and federation scalability
- generic workflow execution framework  
(automatic annotation, data mining, etc.)

# Enabling Technologies

- **Building robust and available persistent identifier service (is in place based on Handles)**

- EPIC: millions of objects, DataCite: published collections
- EPIC offers registration/resolution service for all data centers in Europe
- EUDAT: all objects need to be registered, all policy operations will use PIDs

**Handles**  
**API v1**  
**API v2**

- **Federated AAI service**

- Shib/SAML based world - still a mess due to fragmentation
- can we rely on harmonized EU wide Identity Federation?
- will individual identity providers offer needed attributes?

**complex**  
**problem,**  
**many**  
**facets**

- **Shared Workspaces**

- obviously for different purposes (storing data, automatic annotations, etc)

**in**  
**progress**

- **Monitoring and accounting**

- all participating servers/services need to show stability, availability

**in**  
**progress**

- **Network Services (of course)**



# Today and tomorrow

## **Have chosen the most evident topics for these Training Days**

- iRODS as a basis for repositories and federations of them
- Handles as basis for registered data objects and collections
- AAI as one pillar for establishing trust relations

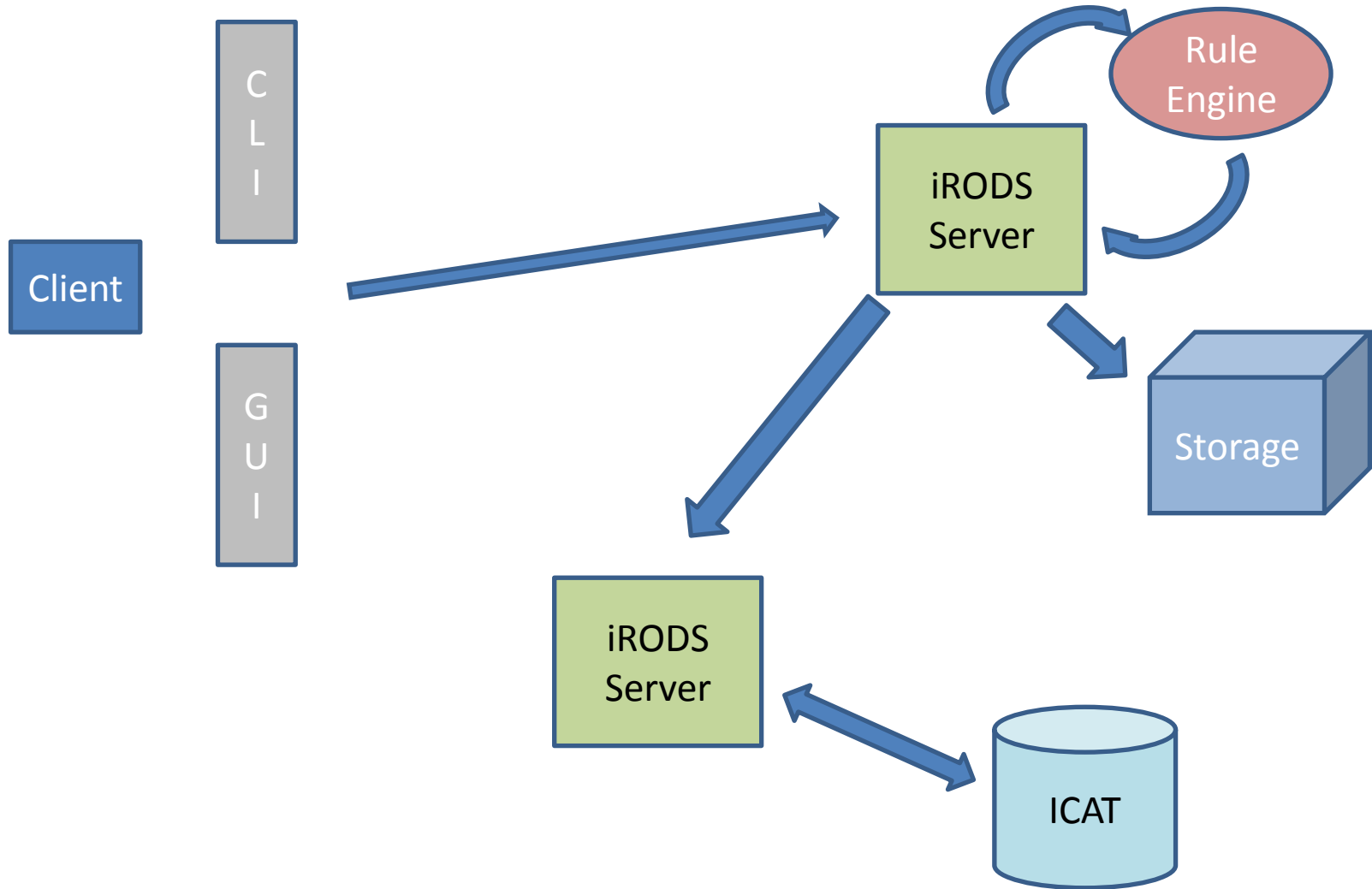
Presentation slides available at:

<http://www.eudat.eu/1st-training-days-presentations>

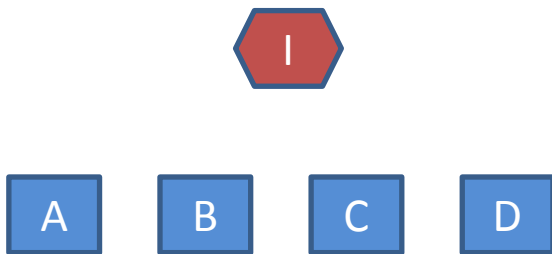
# iRODS Training Session

2:00 – 12:20	Introduction	Peter Wittenburg (MPI Nijmegen)
12:20 – 13:00	Policy rules principles and attributes	Adil Hasan (U Liverpool)
13:00 – 13:45	Lunch (provided)	
13:45 – 14:15	Concrete implementation I	Jean Yves Nief (IN2P3 Lyon)
14:15 – 14:45	Concrete implementation II	Willem Elbers (MPI Nijmegen)
14:45 – 15:30	Coffee and discussion	
15:30 – 18:00	Hands-on Session	Willem Elbers,
20:00	Dinner	

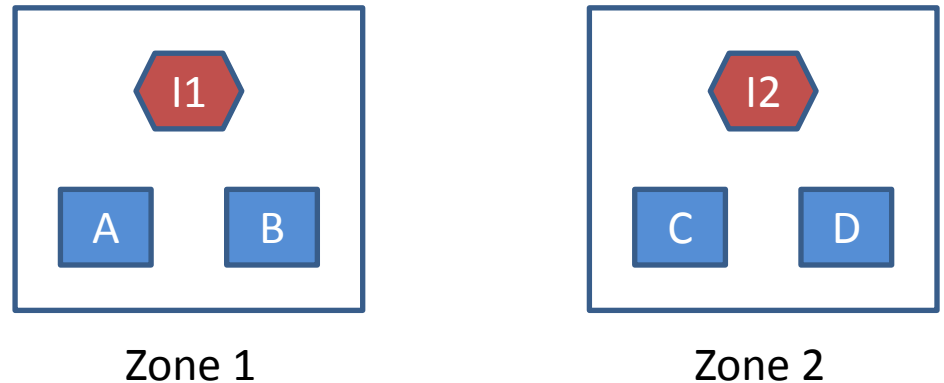
# iRODS Setup



# Linking iRODS Together



Servers in same administrative domain. All registered as resources with same iCAT (I)



Servers in different administrative domains. A and B exist in Zone 1 and C and D in Zone 2

# Additional investment in...

- Policy and rule management
  - Meta-policies and life cycles for managing changes
  - Deriving processes/rules from policies
  - Rule management



# Additional investment in...

- Federated iRODS systems
  - Rules can not be run across zones
  - Cannot replicate across zones
  - Software licensing/data access rights across zones

# CC-IN2P3 activities

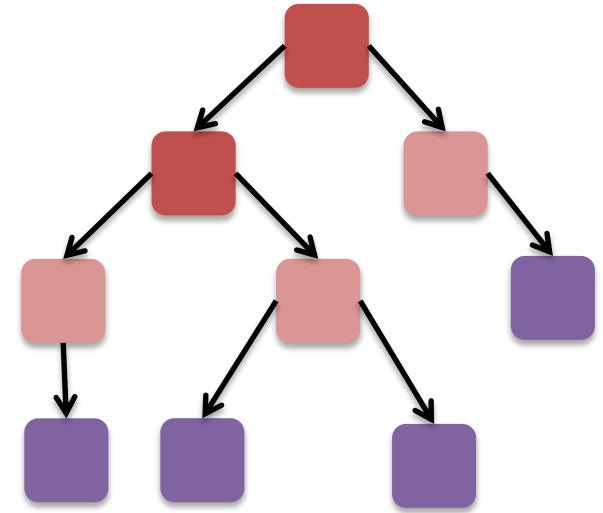


- Federate computing needs of the french scientific community in:
  - Nuclear and particle physics.
  - Astrophysics and astroparticles.
- Computing services to international collaborations:
  - CERN (LHC), Fermilab, SLAC, ....
- Opened now to biology, Arts & Humanities.



# Data Organization

- Data objects
  - Structural metadata
  - Descriptive metadata
  - Resources
    - Audio / Video
    - Text
    - Annotations
  - Each DO stored in a file
  - Complex Metadata hierarchies
- Resources are grouped in session nodes (descriptive metadata)
- Each DO is assigned a PID (we are using the handle system)
- All files are owned by a single user on the file system level





# Logical Replication Layer

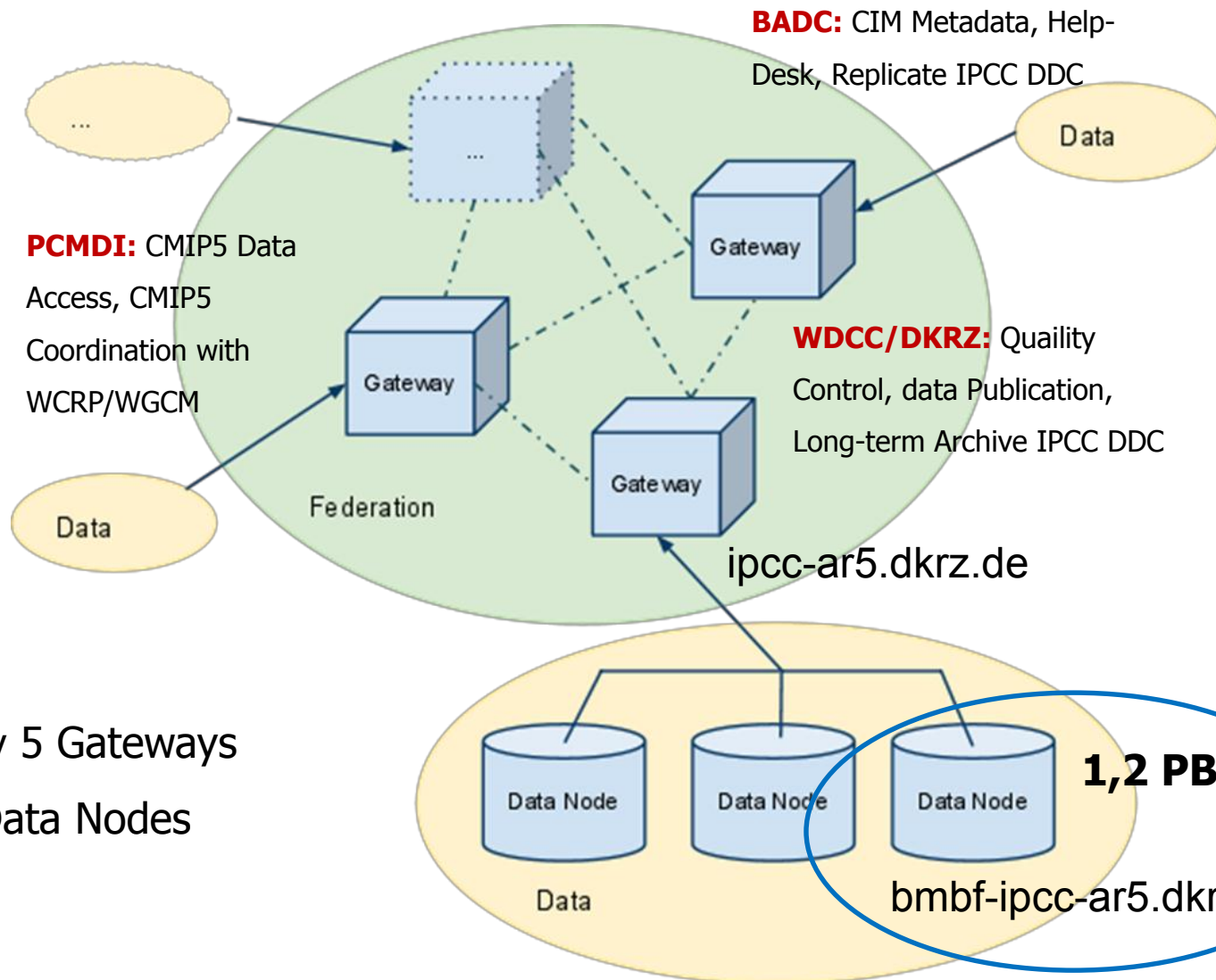
- REPLIX
  - Explorative project to implement logical replication layer and policy
  - Based on iRODS (version 3.0)
  - Demonstrated with a federation between MPI-TLA and RENCI (USA) at DAITF<sup>1</sup> meeting in Copenhagen
  - Currently moving into pre-production
    - Federation between MPI-TLA and RZG
- Why iRODS?
  - Federation support
  - Rule engine – policy based
  - Rsync based transfers
  - Java API



# Handles Training Session

08:30 – 09:00	Arrival, Coffee	
09:00 - 09:15	Introduction	Michael Lautenschlager (DKRZ Hamburg)
09:15 - 09:45	Handle System and its usage	Larry Lannom (CNRI, US)
09:45 - 10:00	EPIC Service	Pieter van Beek (SARA, A'dam)
10:00 - 10:15	DataCite Service	Sebastian Peters (TIB Hannover)
10:15 - 11:00	Use of EPIC API (Hands On)	Willem Elbers (MPI Nijmegen)
11:00 – 11:30	Coffee	

# CMIP5 Data Federation



Presently 5 Gateways  
and 22 Data Nodes

# Finalisation of Quality Assurance

- after completion of quality control of data and metadata (CIM und CF)
  - Quality status: „approved by author“
  - Data are classified as **irrevocable**
  - Long-term archiving in **WDC Climate** of DKRZ
- Execution of the DataCite data publication and integration into library catalogues
  - Each data entity (here one modelling experiment) receives a **citation reference** for direct usage in scientific publications and for transparent data access via **DOI** (Digital Object Identifier)
  - Citation reference contains data author(s) and title as well as WDC Climate as data publisher and the DOI itself
  - Resolution of the DOI leads to the „**Landing Page**“, its address is stored in the central repository of the DOI Handle Servers



# Handles Resolve to Typed Data

---

Handle	Data Type	Handle Data
10.1525/bio.2009.59.5.9	URL	<a href="http://caliber.ucpress.net/doi/abs/10.1525/bio.2009.59.5.9">http://caliber.ucpress.net/doi/abs/10.1525/bio.2009.59.5.9</a>
	HS_ADMIN	handle=0.na/10.1525; index=200; [delete hdl,add val,read val,modify val,del admin,add admin,list]
	10320/loc	<locations chooseby="locatt, country, weighted"> <location id="1" cr_type="MR-LIST" href="http://mr.crossref.org/ iPage?doi=10.1525%2Fbio.2009.59.5.9" weight="1" /> <location id="2" cr_src="unca" label="SECONDARY_BIOONE" cr_type="MR-LIST" href="http://www.bioone.org/doi/full/10.1525/ bio.2009.59.5.9" weight="0" /> </locations>



~~SAGA~~

- SIGNALLING  
- CRAWLERS

Master  
MPI-PL

12345/abc	
LOC	25% http://mp.pl/abc 75% http://10916/123
SHA1	DEADC0FFEE
REPL	SARA CINECA

Mirror1 (SARA)

10916/123	
LOC	url: 11858/thing http://www.sara.pl/thing
SHA1	DEADC0FFEE
ORI	12345/abc
REPL	GWDC

MIRROR2  
CINECA?

MIRROR 1.1  
GWDC

11858/thing	
LOC	http://gwdc.pl/for
ORI	10916/123
SHA1	BADC0FFEE5
TORI	12345/abc ???

m.2.1

m.2.2

# AAI Training Session

11:30 – 12:00	Introduction - AAI and research infrastructures	Peter Gietz (DAASI Tübingen)
12:00 - 12.30	An introduction into certificates in all their variants and aspects	Jens Jensen (STFC Oxford)
12:30 - 13:00	FIM (concentrating on <a href="#">SAML2</a> and the national IDFs)	Willem Elbers (MPI Nijmegen)
13:00 – 14:00	Lunch (provided)	
14:00 – 14:30	Collaborations in research infrastructures	Niels van Dijk (SARA A'dam)
14:30 – 15:00	Two use cases from the communities	Twan Goosen (MPI Nijmegen), Hannes Thieman (DKRZ Hamburg)
15:00 – 15:30	EUDAT synthesis vision	Johannes Reetz (RZG Munich)
15:30	Close	



# EUDAT's Mission in AAI

- Provide federated access to EUDAT core services.
- Making use of existing AAI infrastructures and federations.
- Make EUDAT services AA interoperable with Community users and services.
- By this way meet communities requirements in the course of such an integration approach
- Secure the EUDAT service infrastructure and make it trustworthy

# Community Perspective

Community	Other projects	# users	technology	status	IGTF
photon/neutron facilities	EUROFEL, PanData, CRISP	>30,000 visiting researchers per year	Shibboleth/ SAML	Umbrella prototype	no
Social Sciences and Humanities	DARIAH, CLARIN, CESSDA, DASISH, Bamboo	Hundreds now, potential for 10000+ across SSH	Shibboleth/ SAML	Prototype CLARIN Service Provider federation deployed in 3 countries	yes
High Energy Physics	WLCG	10,000+ globally	X509	Production	yes
Earth Sciences	Earth System Grid Federation, GENESI-DEC, CMIP5, Metafor, IS-ENES, CORDEX, Exarch, Climate Data Exchange	5000+ for CIMP5	OpenID, X.509 and SAML	Production - earth system grid	Not yet, but foresee for EGI integration
Life sciences	ELIXIR, BioMedBridges, BBMRI, NCoEDG & potentially 10 BMI ESFRI Projects	3 million researchers access data via EBI website each Year	not chosen yet	Security is included in BioMedBridges project workplan, pilot is being planned with EGI	no

from **Federated Identity Management for Research Collaborations**  
 Research paper: 23rd April 2012



# Pragmatism: Approach in 3 Phases

1. Single sign-on for EUDAT members: reuse of the EUDAT (staff) login with other EUDAT services.  
Done using Atlassian CROWD SSO, managed by CSC
2. Stepwise integration of EUDAT core services with AuthN and AuthZ tools used by communities, starting with CLARIN (Shib) and ENES (OpenID/SAML).  
Started to investigate options to “shibbolethize” iRODS services
3. Long term perspective, supporting as much requirements as possible, including “citizen scientists”, federated identities within the EUDAT administrative domain, credential delegation