

# **Safe Replication Service**

#### How to create safe replications in a flexible way

Peter Wittenburg, Morris Riedel

MPIPL, Nijmegen, FZJülich, Germany Mark van de Sanden SARA, The Netherlands







## **Community Service Wishes**

#### In Progress as Services (Task Forces set up)

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

### In Specification/Discussion as Services

Aggregated EUDAT Metadata Domain
Researcher Data Store (Simple Upload, Share and Access)
Common Authentication/Authorization Infrastructure

### In Progress as Research Issues (WP7)

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



### **Principal Task**



**11010010** 



## Data, Metadata, PIDs

110100101



Replication to be policy rule based, i.e. explicit and therefore auditable.



EUDA



## **Requirements I**

- Community managers(CM) can specify M replicas to be stored for N years at specific centers.
- CM can specify which objects/collections need to be copied.
- CMs want to know whether the replicas are **identical** with the source.
- CMs want to execute regular checks across all object instances and give indications in case of differences.
- CMs want the replicas to be accessible by users while maintaining the access permissions as defined by the originating community center.
- The work of all centers participating is being audited (RAC or DSA) to establish trust and therefore all activities need to be explicitly be described by policy rules.
- The solutions should require as little changes as possible on the community data organization side, although general principles must be in place: PIDs for objects, separate metadata and data, PID reference in the metadata.





## **Requirements II**

- This service should be offered as a highly available and robust service with high throughput network will be limiting factor.
- The transmission needs to be secure in so far as the traffic needs to be carried out between certificated services/servers. At first instance no encryption needs to be done.
- It is expected that an access logging is being done and that aggregated statistics are being provided.
- It is expected that proven and widely used software is being used and that the approach is modular involving different protocols and solutions were possible.
- Multiple URLs must be supported by the chosen PID framework and it must cope with the many PIDs required by the communities for all their DOs.





# Your input

- comments?
- additional requirements?

### Thanks for your attention.

