

iRODS workflows for the data management in the EUDAT pan-European infrastructure

iRODS UGM 2017

Utrecht, 14-15.06.2017

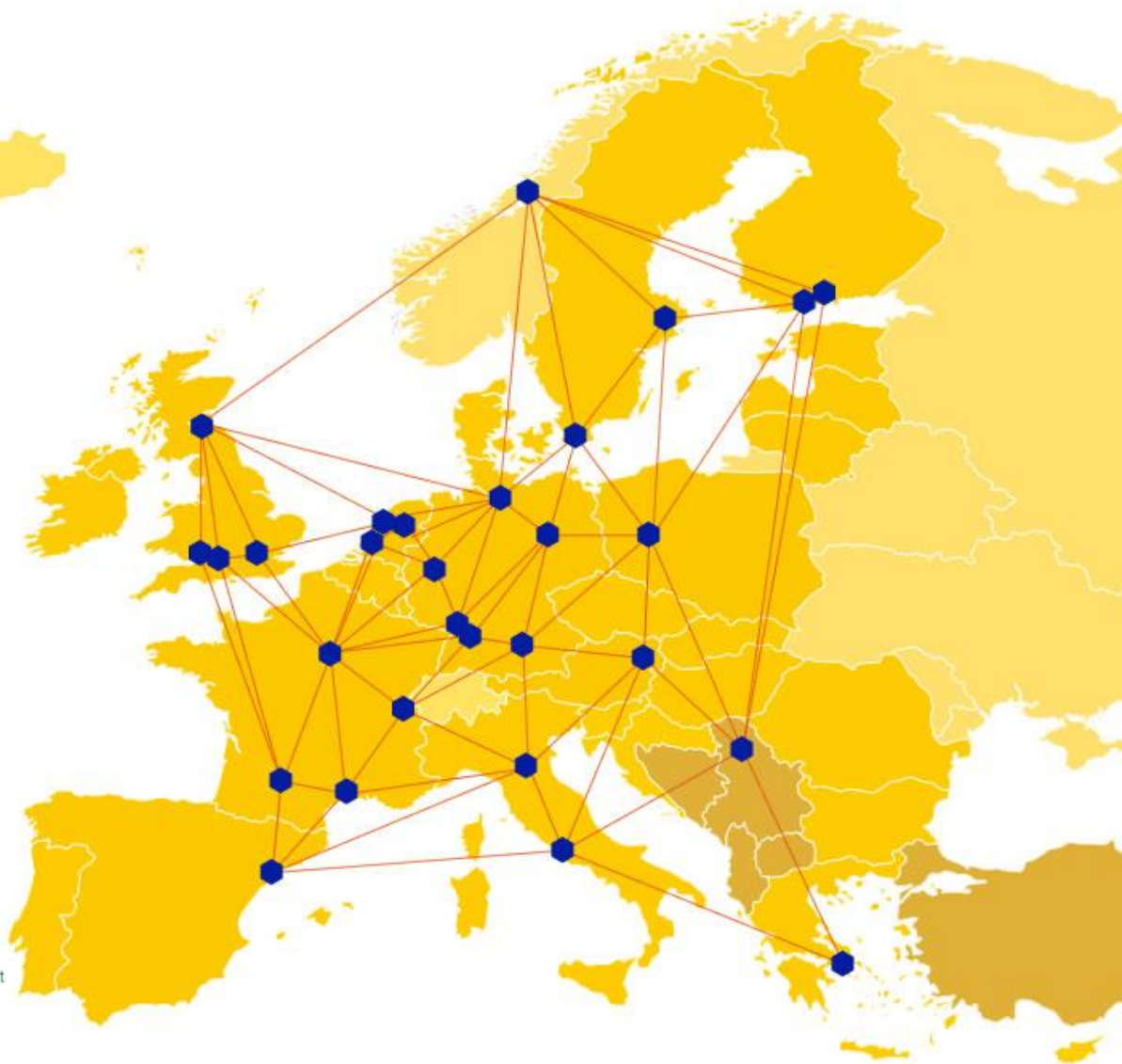
Claudio Cacciari
(c.cacciari@cineca.it)

Outline

- Introduction to EUDAT
- The challenge
- The solution
- B2SAFE service
- B2SAFE module
- Data management: replication
- Persistent Identifiers
- Implementation
- Future work
- Conclusion

Introduction

- The European project EUDAT built a data e-infrastructure, called **Collaborative Data Infrastructure** (CDI)
- connecting 16 data and computing centres
- to support over 50 research communities spanning across many different scientific disciplines.



The challenge

- One of the main challenges to implement such infrastructure was to enable the users to manage their data in the same way across the different data centres despite each centre has its own peculiarities at hardware, software and policy level

Seamless data management

IN

Heterogeneous environments

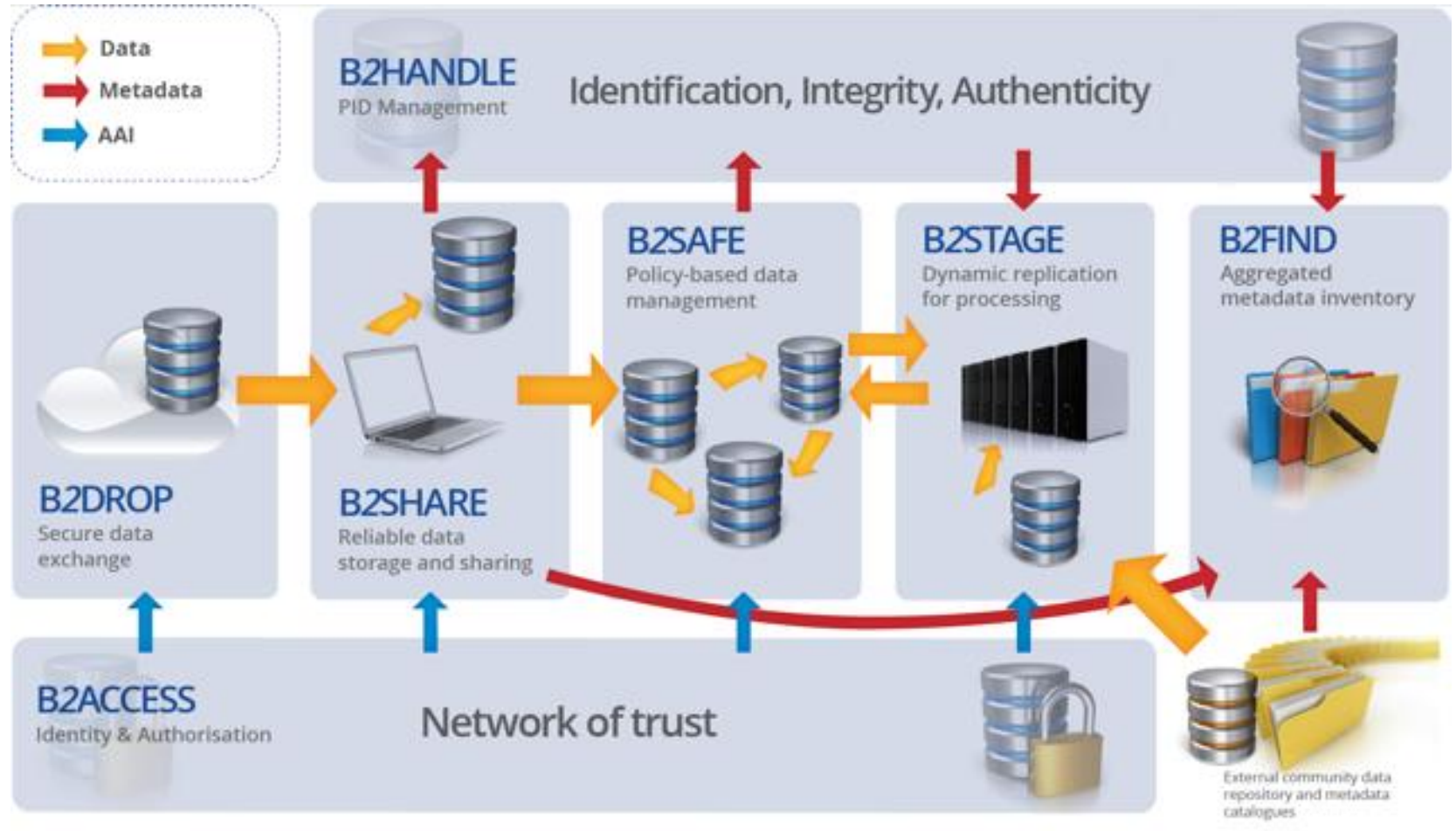
The solution

- EUDAT adopted **iRODS** to deal with this heterogeneity relying on its features:
 - To define a **common abstraction layer** on top of the different storage systems.
 - To provide **a shared set of software interfaces** and clients to perform data management operations.
 - To **federate** different administrative regions.
 - To enforce a common set of **policies** through data management workflows.

B2SAFE service

- The CDI has an architecture based on services, which form an integrated suite
- iRODS is part of the B2SAFE service, which supports the long-term data preservation.

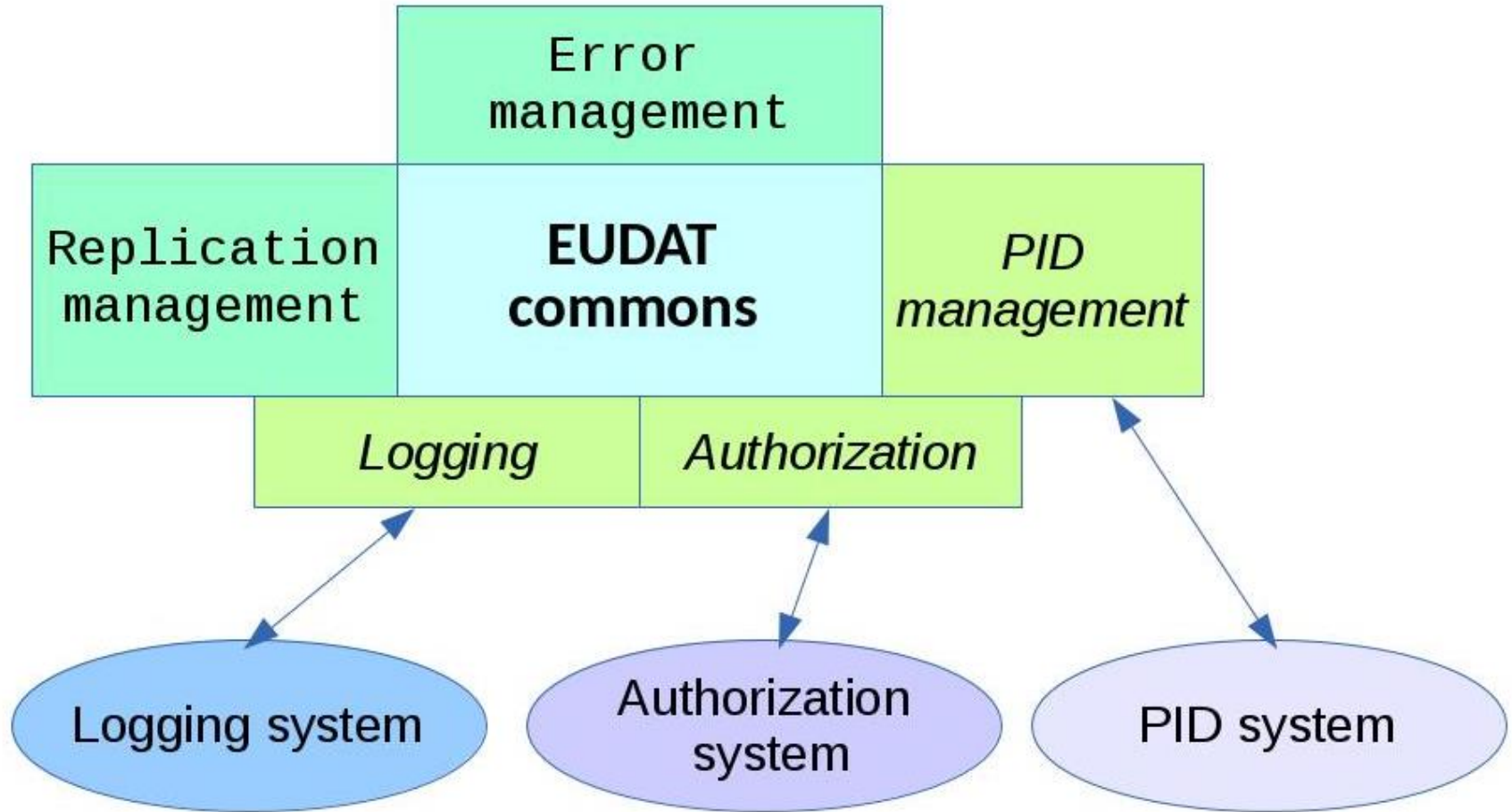
EUDAT service suite



B2SAFE additional module

- The B2SAFE service extensions to iRODS are implemented through rules and python scripts and can be grouped by functionality:
 - logging,
 - authorization,
 - persistent identifiers (PIDs) management,
 - data replication ,
 - error management,
 - utilities.

B2SAFE functions



Data management workflow: replication 1

- B2SAFE's main objective is to enforce policies for the long-term data preservation.
- In this context one of the most important strategies
 - to keep the data safe and support disaster recovery scenarios, is



- the replication of data to multiple sites, geographically distributed.

Data management workflow: replication 2

- Further benefits:
 - The data replication is a way to optimize the data exploitation. Because many of the CDI's data centers offer computing resources, therefore, the data replication allows moving the data close to those resources;
 - many scientific communities are distributed across Europe, hence having the data close to their institutions improve their accessibility.

Cross-zone replication

- iRODS offers already replication mechanisms, but within the same zone. We needed to replicate data sets across different zones, which implies to deal with a certain number of issues related to
 - the tracking of the replicas,
 - the fault tolerance,
 - the data integrity,
 - the performance.

Replication: iRODS rules 1

- we defined a rule called *EUDATReplication*, which relies on all the aforementioned extensions.
- The rule can be triggered client-side, with the “irule” command, but it is usually called within a policy enforcement point in “core.re”

```
*source="/CINECA01/home/original_path"  
*destination="/CINECA01/home/mypath";  
*recursive = "true";  
*registered = "true";  
*status = EUDATReplication(*source,  
                             *destination,  
                             *registered,  
                             *recursive,  
                             *response);
```

Replication: iRODS rules 2

- It is triggered when a new object or a new collection is uploaded to a specific path.
- The rule can receive as input the path either of an object or of a collection and replicate it to the proper destination.

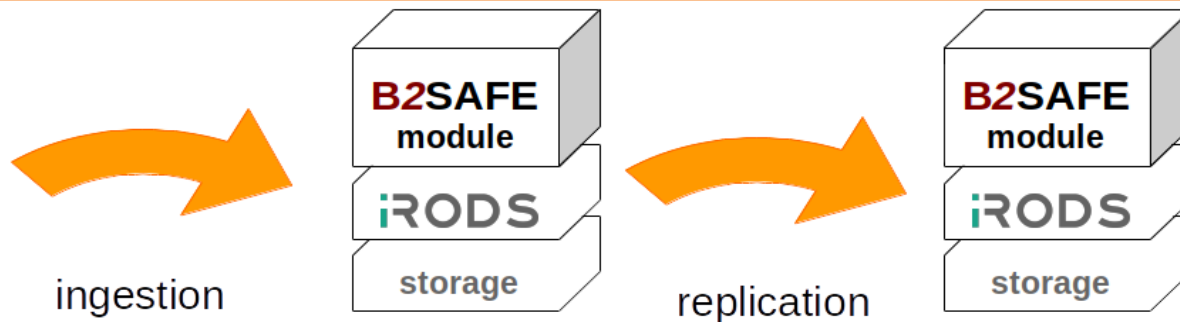
➤ **EUDATReplication**

- EUDATCatchErrorDataOwner
- EUDATRegDataRepl
 - EUDATSearchAndCreatePID
 - EUDATPIDRegistration
- EUDATCheckIntegrity

Replication process

Where are my replicas?

What happens when the collection is moved to a different location?



Persistent identifiers

- The persistent identifiers (PIDs) management consists of
 - multiple rules
 - and a python based client (epicclient2.py), which is able to connect to an instance of the EUDAT B2HANDLE service.
- A PID is a unique identifier, based on the Handle scheme, which is composed by a prefix and a suffix, for example:
842/f5188714-f8b8-11e4-a506-fa163e62896a
- The B2HANDLE service is a distributed service, which allows publishing PIDs and making them globally discoverable, relying on a software component called Handle System, supported by DONA.

EUDAT PID record profile

- By design, the handle scheme permits to extend arbitrarily the set of attributes associated to a PID, called PID record.
- EUDAT defined a PID record profile to formalize the EUDAT extended attributes

EUDAT PID record profile: single object's attributes

URL

- The http address of the object.

EUDAT/CHECKSUM

- The MD5 checksum of the object

EUDAT/CHECKSUM_TIMESTAMP

- The timestamp in ISO UTC/ZULU time of the update operation of the checksum

EUDAT/FIXED_CONTENT

- Boolean value to show if the content of the object is immutable (true) or can be updated without changing the PID (false)

EUDAT PID record profile: replica's attributes

EUDAT/FIO

- First Ingested Object: the PID pointing to the location of the first CDI node which has ingested the object.

EUDAT/PARENT

- The PID pointing to the parent element in a replication chain

EUDAT/ROR

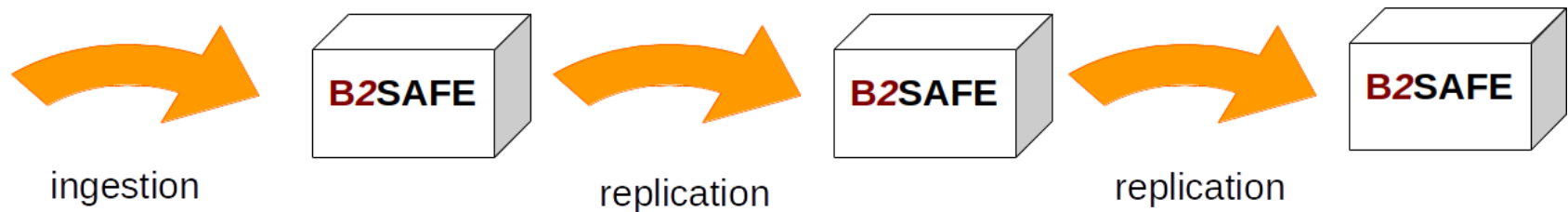
- A pointer to the community's Repository of Record element in a replication chain. It can be a PID or any other identifier chosen by the community

EUDAT/REPLICA

- A list of PIDs pointing to the replicas in a replication chain

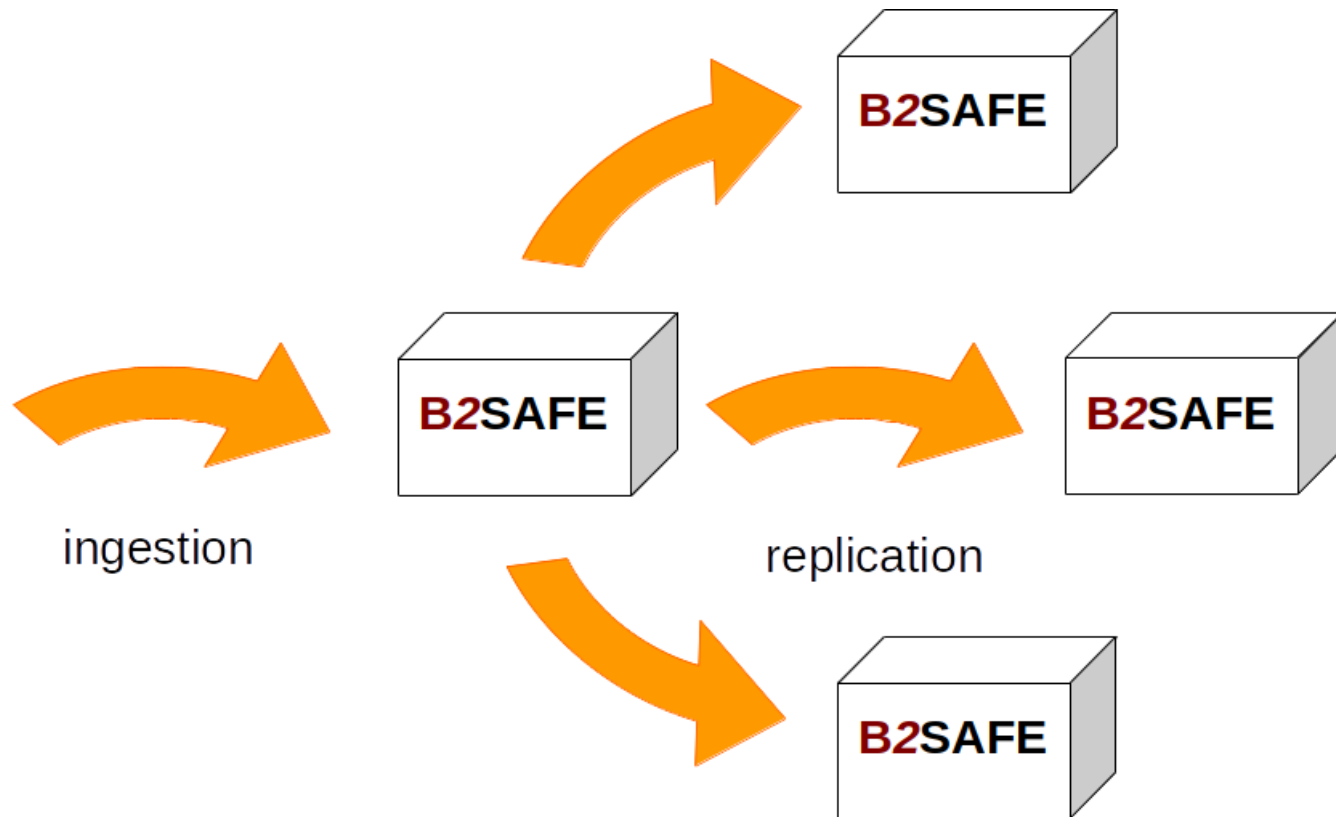
Replication: tracking replicas 1

- The replication sequence can involve multiple steps and supports different patterns. It could be a single chain of replicas and replicas of replicas



Replication: tracking replicas 2

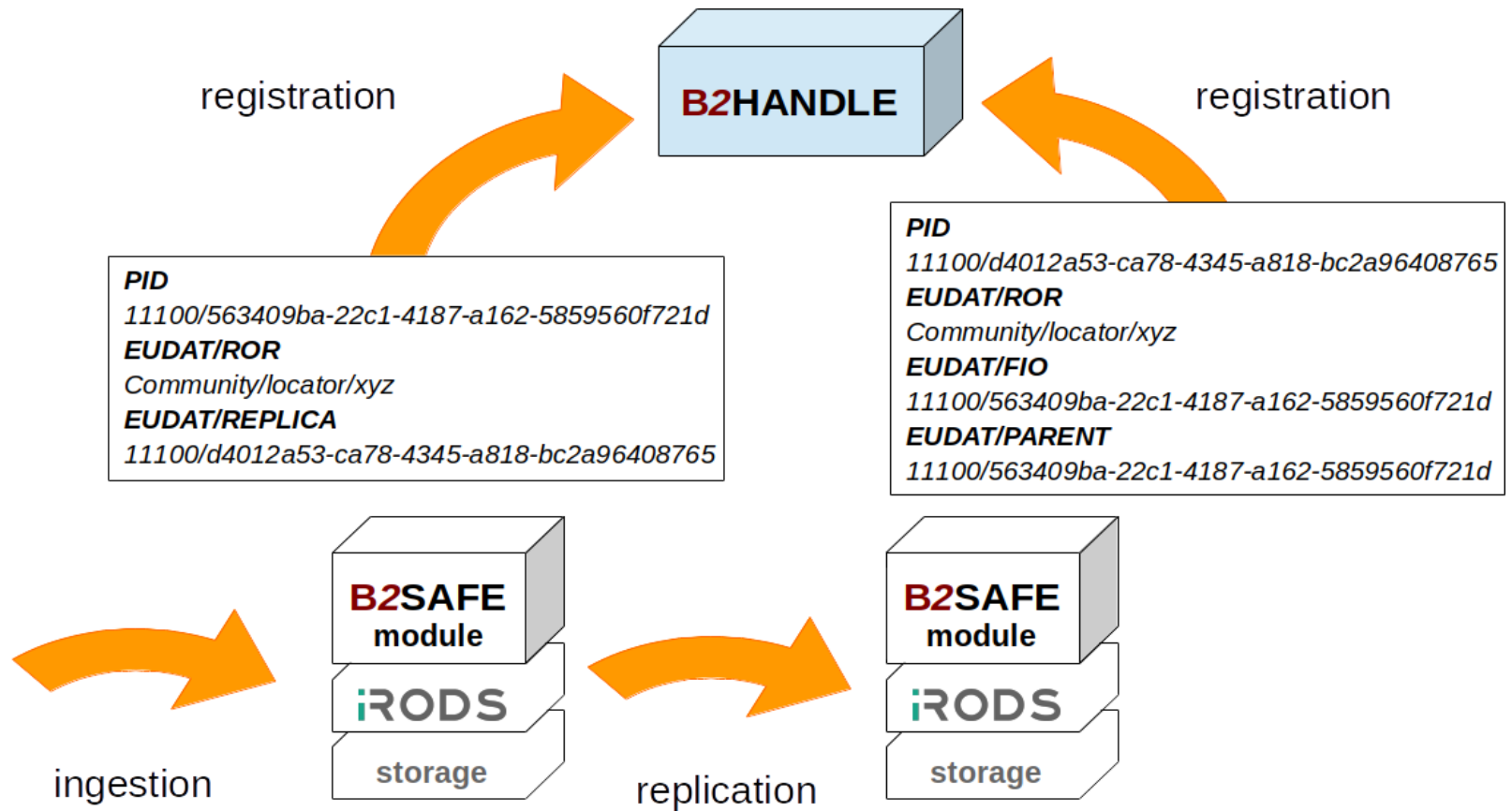
- or, for example, have a star configuration, where each replica is copied directly from the master.



Replication: double linked chain

- All the different patterns share a certain number of elements, which are tracked and form a double linked chain:
 - each parent's PID record includes pointers to its replicas
 - each replica's PID record includes a pointer to the parent.
- Each replica's PID record includes
 - the pointer to the first copy of the object ingested into the CDI (First Ingested Object, FIO)
 - If it exists, the pointer to the master copy, stored outside the CDI, in the community's domain, known also as Repository of Records (RoR).

Replication process



Replication: replica's tracking benefits

- This approach has three main benefits:
 - it permits to the B2SAFE administrators to be always aware of the location and the number of copies of every object and collection stored on the infrastructure
 - it allows the users to find the data location that best fits their needs.
 - in case of failure of one node of the CDI hosting a copy of the data, the user can always follow the pointers in the PID records to find another accessible copy.

Future work

- the architecture:
 - Some of the components of the B2SAFE service are good candidates to be implemented as iRODS plugins.
 - Other components could be, potentially, replaced by iRODS new features, like the messaging framework.
- the data management workflows:
 - Checksum comparison: currently the B2SAFE administrator has to configure this procedure separately from the replication workflow. It is possible to achieve a better integration.

Conclusions

- The B2SAFE service implements two fundamental data management workflows:
 - the data replication
 - the assignment of globally discoverable identifiers,
- which can be used as building blocks from the users to define more complex and customized data policies.

B2SAFE developers team

- ▣ Claudio Cacciari (Cineca) – c.cacciari@cineca.it
- ▣ Robert Verkerk (SURFsara) - robert.verkerk@surfsara.nl
- ▣ Adil Hasan (SIGMA2) - adilhasan2@gmail.com
- ▣ Javier Quinteros (GFZ) - javier@gfz-potsdam.de
- ▣ Julia Kaufhold (MPCDF) - julia.kaufhold@mpcdf.mpg.de

Thanks for you attention

Questions?

www.eudat.eu