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

**Claudio Cacciari**
CINECA
Via Magnanelli 6/3,
Casalecchio di Reno (BO),
Italy
c.cacciari@cineca.it

**Robert Verkerk**
SURFsara
SURF Science Park Building,
Science Park 140
1098 XG, Amsterdam,
The Netherlands
robert.verkerk@surfsara.nl

**Adil Hasan**
SIGMA2
Abels gate 5, Trondheim,
Norway
adilhasan2@gmail.com

**Javier Quinteros**
German Research Centre for Geosciences
(GFZ)
Telegrafenberg
D-14473 Potsdam, Germany
javier@gfz-potsdam.de

**Julia Kaufhold**
Max Planck Computing and Data Facility
(MPCDF)
Gießenbachstraße 2
85748 Garching, Germany
julia.kaufhold@mpcdf.mpg.de

## ABSTRACT

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. 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. Therefore, EUDAT adopted iRODS to deal with this heterogeneity relying on its features:

- To define a common abstraction layer on top of the difference storage systems.
- To provide a shared set of software interfaces and clients to perform data management operations.
- To enforce a common set of policies.
- To federate different administrative regions.

On the other hand, each community has its own characteristics and often it requires specific customizations to cope with its data life cycle. Hence, beyond this common horizontal layer, through iRODS, EUDAT can offer the flexibility of a vertical integration with the community's tools and policies. In order to implement those policies and functions, in the context of this project, we extended iRODS with a set of rules and scripts, which form, together with the underlying software stack, the B2SAFE service. It allows the replication of data collections across different iRODS zones, takes care to assign a unique identifier to each data object and collection, to log every failed transfers and to store a minimal set of metadata together with the data themselves. The unique identifiers are stored in a de-centralized registry, called B2HANDLE, which makes them globally resolvable and persistent. In this article, we introduce the B2SAFE architecture and highlight the integration between iRODS and the B2HANDLE system and the corresponding workflows.

**Keywords**

Data Management, EUDAT, replication, B2SAFE, persistent identifier, handle.

*iRODS UGM 2017,* June 13-15, 2017, Utrecht, Netherlands

[Authors retains copyright. EUDAT receives funding from the European Union's Horizon 2020 programme – DG CONNECT e-Infrastructure Contract No. 654065.]

### INTRODUCTION

In the last years, the European Commission promoted and funded initiatives to strength the data and computing infrastructures, which supported the research communities. Before those initiatives, some infrastructures were already in place, others only planned, but the panorama was fragmented with lack of interoperability and the concrete risk of increasing maintenance costs. The EUDAT project [1] was born as an answer to those concerns [2], with the objective to build a real pan-European data e-infrastructure.

EUDAT adopted iRODS as one of the main component of its infrastructure, represented in Figure 1 and called Collaborative Data Infrastructure (CDI). The CDI, which is now fully operative, has an architecture based on services, which form an integrated suite, depicted in Figure 2. iRODS is part of the B2SAFE service [3], which supports the long-term data preservation.

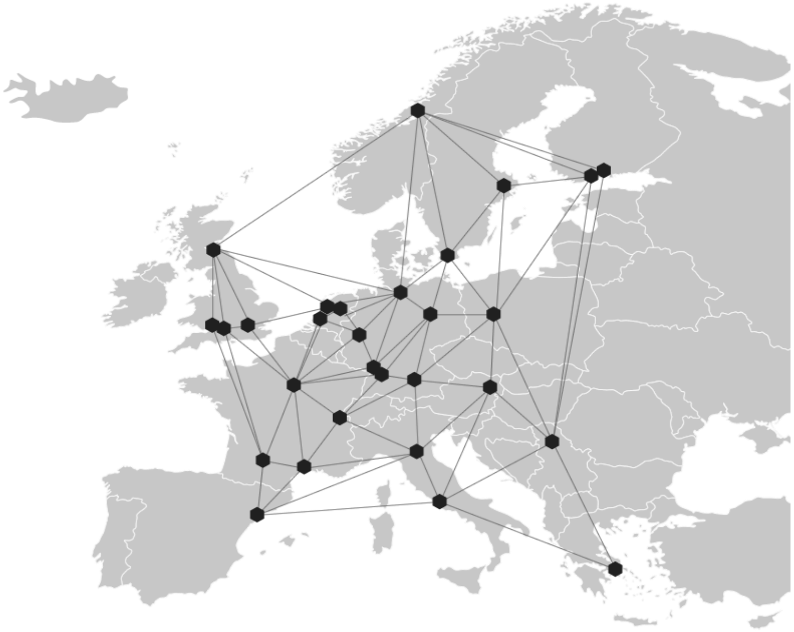

**Figure 1. EUDAT Collaborative Data Infrastructure.**

### CHALLENGE

The Collaborative Data Infrastructure wants to offer a common experience to the users, providing them tools to manage their data in the same way across the different centers, which are part of the CDI. This is a great challenge, because it connects 16 data and computing centres to support over 50 research communities spanning across many different scientific disciplines and they are heterogeneous at hardware, software and policy level. The B2SAFE service relies on iRODS to overcome these obstacles, extending it through rule sets and scripts.

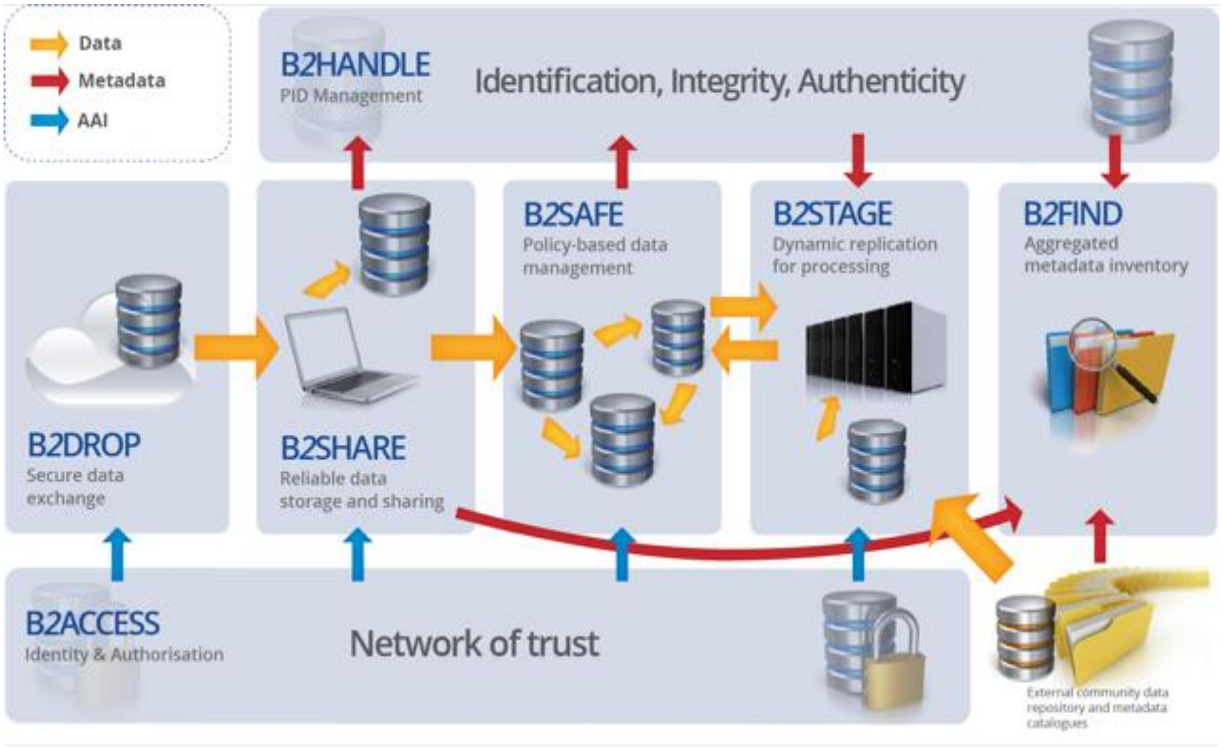
Data
Metadata
AAI
B2HANDLE
PID Management
Identification, Integrity, Authenticity
B2SAFE
Policy-based data management
B2STAGE
Dynamic replication for processing
B2FIND
Aggregated metadata inventory
B2DROP
Secure data exchange
B2SHARE
Reliable data storage and sharing
B2ACCESS
Identity & Authorisation
Network of trust
External community data repository and metadata catalogues

**Figure 2. EUDAT data infrastructure service suite.**

### SOLUTION

The B2SAFE service provides a homogeneous and user-friendly interface to the user, dealing with four aspects related to critical differences among the various data centres:

1. Each CDI's data centre has different storage systems or, in the best case, different configurations of the same systems. The iRODS back-end abstraction layer allows hiding this complexity. In many cases, they are just plain file systems, but sometimes even tape libraries, or combinations of the two in a hierarchical solution. Part of this abstraction is also the common namespace, which simplifies both the data management for the users, and the back-end system maintenance for the administrators.
2. Because of this abstraction, it is possible to use a common set of software interfaces and clients to perform data management operations. In particular, B2SAFE allows the user to connect to it via iRODS icommands [4], WebDAV [5] and GridFTP. This last interface is an integration with the Globus GridFTP server [6], implemented through a specific library developed by EUDAT [7].
3. While the first and the second point address hardware and software heterogeneities, B2SAFE relies on the federation feature [8] to overcome administrative limitations. Each data centre in the CDI is an independent administrative unit; therefore, it overlaps perfectly with the concept of iRODS zone. Those zones are federated, allowing users to get access to multiple nodes of the infrastructures in a seamless way. However, the federation does not follow a full-mesh schema, where every zone is federated with each other, but it is configured as a set of islands, where each island, composed by two or more nodes, represents the storage area of a specific research community.
4. EUDAT defined a set of policies, implemented through iRODS rules, in order to enforce the data preservation best practices within the B2SAFE service. This set is packaged as an additional module [9], which is deployed on top of iRODS. In this way, the service offers common policies across the whole infrastructure.

### IMPLEMENTATION

The B2SAFE service extensions to iRODS are implemented through rules and python scripts and can be grouped by functionality, as shown in Figure 3: logging, authorization, persistent identifiers (PIDs) [10] management, which rely also on software tools independent from iRODS; and data replication and error management, which are based completely on rules. Moreover, there is a set of utilities to support all the aforementioned groups.

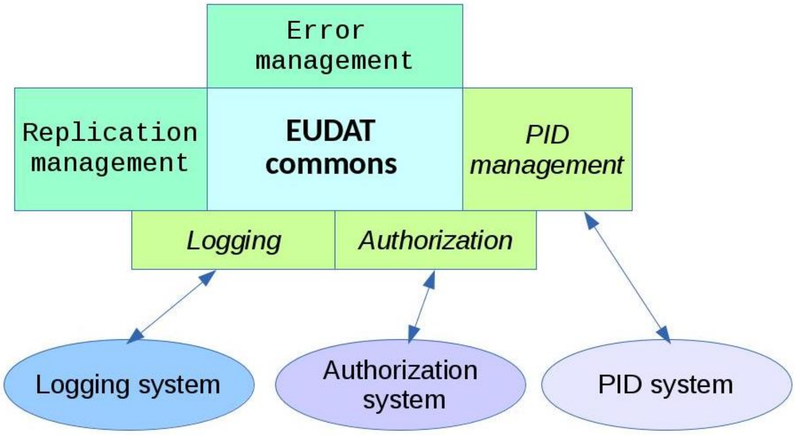

**Figure 3. B2SAFE architecture.**

### Logging

The logging mechanism is independent from the core iRODS logging system and its purpose is to log the information related to the data replication rules. Currently the information is stored in files. In the future, the log messages could be queued into a messaging system.

### Authorization

The authorization extension is a tool to control "who can execute which rule" and it is implemented through a rule ( *EUDATAuthZ(*user, *action, *target, *response)* ), which compares three attributes: the username, the "action" and the "target" against a set of assertions stored on a JSON (JavaScript Object Notation) formatted file and it returns true in case of match, false otherwise. For example:

```
acPreProcForExecCmd(*cmd, *args, *addr, *hint) {
    if (*cmd != "authZmanager.py") {
        EUDATAuthZ("$userNameClient#$rodsZoneClient", *cmd, *args, *response);
    }
}
```

The above code, placed in the rule set core.re, enforces the permission to execute external commands according to the assertions defined in the assertion file, which could be like the following one:

```
{
 "assertion 1":
      { "subject": [ "jack, james, master*" ],
        "action":  [ "getLDAPattributes.py" ],
        "target":  [ "password" ]
      }
```

}

Which means that the users “jack”, “james” and all those beginning with word “master” are allowed to execute the command “getLDAPattributes.py” with the argument “password”. Even in this case the mechanism can be further developed, in the future, to query an external authorization system.

### Persistent identifier management

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 [11]. A PID is a unique identifier, based on the Handle scheme [12], which is composed by a prefix and a suffix, for example: 842/f5188714-f8b8-11e4-a506-fa163e62896a, where 842 is the prefix. The B2HANDLE service is a distributed service, which allows publishing PIDs and making them globally discoverable, relying on a software component called Handle System [12], supported by DONA [13]. The client exploits the B2HANDLE library [14] to create, modify and search across PIDs. 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. In the Table 1, we reported just the mandatory ones, which are the only relevant for the current explanation.

| Type | Description | Example |
|---|---|---|
| URL | The http address of the object. | http://eudat.data.center/b2safe/api/object/CINECA01/home/collectionOne/myobject |
| EUDAT/CHECKSUM | The MD5 checksum of the object | f63fe6ae1540199f65d6dd3d7048c46b |
| EUDAT/CHECKSUM_TIMESTAMP | The timestamp in ISO UTC/ZULU time of the update operation of the checksum | 2008-09-05T16:30Z |
| 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) | True |
| EUDAT/FIO | First Ingested Object: the PID pointing to the location of the first CDI node which has ingested the object | 11100/785973e1-cd53-4c2f-bfe9-ed60d355725b |
| EUDAT/PARENT | The PID pointing to the parent element in a replication chain | 11100/563409ba-22c1-4187-a162-5859560f721d |
| EUDAT/ROR | A pointer to the community’s Repository of Record element in a | Community/locator/xyz |

| | | |
|---|---|---|
| | 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 | 11100/d4012a53-ca78-4345-a818-bc2a96408765, 99000/28c50c7c-65d3-436a-bd69-a333d82df192 |

**Table 1. EUDAT PID record profile: mandatory attributes.**

### Replication

As mentioned before, 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. Besides, 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; and many scientific communities are distributed across Europe, hence having the data close to their institutions improve their accessibility.

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 and the performance. Thus, 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", so that 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. This operation relies on the PIDs management rules and scripts to assign a PID to the source, the destination and link them together using the attributes of the EUDAT PID record profile, as depicted in Figure 4.

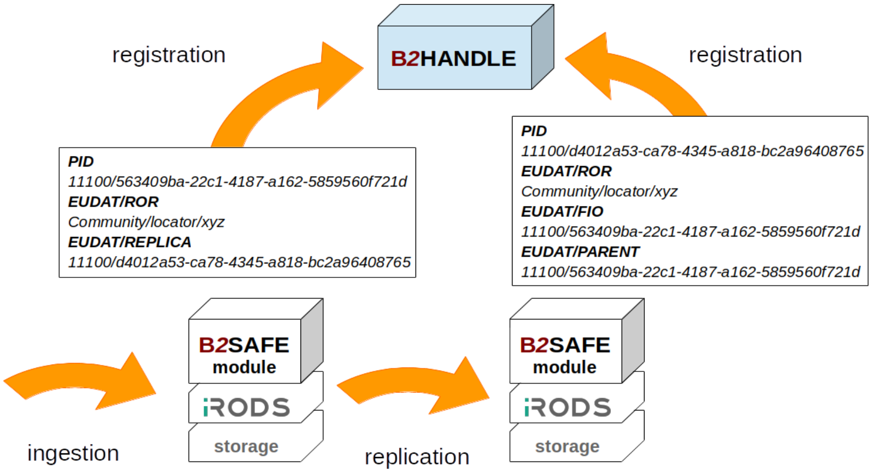
registration
B2HANDLE
registration
PID
11100/563409ba-22c1-4187-a162-5859560f721d
EUDAT/ROR
Community/locator/xyz
EUDAT/REPLICA
11100/d4012a53-ca78-4345-a818-bc2a96408765
PID
11100/d4012a53-ca78-4345-a818-bc2a96408765
EUDAT/ROR
Community/locator/xyz
EUDAT/FIO
11100/563409ba-22c1-4187-a162-5859560f721d
EUDAT/PARENT
11100/563409ba-22c1-4187-a162-5859560f721d
B2SAFE
module
iRODS
storage
B2SAFE
module
iRODS
storage
ingestion
replication

**Figure 4. Replication process with PID attributes.**

The replication sequence can involve multiple steps and supports different patterns. It could be a single chain of replicas and replicas of replicas, or, for example, have a star configuration, where each replica is copied directly from the master. Anyway, 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 and each replica's PID record includes a pointer to the parent. Moreover, each replica's PID record includes the pointer to the first copy of the object ingested into the CDI (First Ingested Object, FIO) and, 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). This approach has two 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 and it allows the users to find the data location that best fits their needs. In addition, 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.

However, this mechanism is meaningful only if the relation between the PID and the object's location is defined univocally (we are not considering here multiple versions of the same object for the sake of brevity): there must be only one PID for each object, therefore for each URL and CHECKSUM values in the PID record and vice versa. Because otherwise the meaning of persistent identifier as described in [10] is lost. The Handle server does not enforce this constraint intrinsically, therefore we have enforced it client-side in the B2SAFE PID management rules. The drawback of this further check is a decrement of the rate of the creation of new PIDs, during the PID registration. Rate that is also affected by security checks. Because, since the rules to create new PIDs are accessible to every user of a B2SAFE instance, we rely on the authorization extension to avoid that unauthorized users abuse of this feature. Usually, the performance of the whole replication procedure is not a critical aspect for the users, but sometimes, when the data set to be replicated scales to millions of objects, it is. In this case, the rule *EUDATReplication* is flexible enough to break the procedure into two sub-processes: the first one replicates the data, the second registers them. Thus the asynchronous PID registration can be scheduled later, once the data transfer is completed, allowing the B2SAFE administrators to find the most suitable period to perform it.

In the case of a failed replication, a certain number of checks are in place to identify the cause of the failure. Namely lack of permissions, failure in the PID registration and failure in the data transfer resulting in different size or checksum of the source compared to the replica. The replication rule saves the information about each failed task using the EUDAT logging extension. Then this information can be passed as input to another rule, *EUDATTransferUsingFailLog*, which is able to retry each failed replication. This asynchronous solution has been preferred to a synchronous one because often the reason of the failure lasts for a certain interval of time and it would be clearly useless to repeat the transfer within the same interval.

## FUTURE WORK

The work described in the previous chapters can be improved along two directions. One is about the architecture: some of the components of the B2SAFE service are good candidates to be implemented as iRODS plugins. In particular, the authorization extension could benefit from this implementation in term of performance and better integration with the other rules. Other components could be, potentially, replaced by iRODS new features. It is the case of the logging mechanism, which could be superseded by the messaging framework. The other direction is related to the data management workflows. One of the PID record attributes is the object's checksum that is recorded to support the fixity feature, required to grant the data integrity as intended in [15]. The checksum of the master copy and of the replicas is compared periodically to verify that they are still coherent, however the B2SAFE administrator has to configure this procedure separately from the replication workflow. It is possible to achieve a better integration, abstracting some local system dependent details and automatizing part of the steps and creating in this way another building block to ease the definition of complex data policies.

## CONCLUSIONS

Meeting the requirements of so many different scientific communities, it is really challenging for a pan-European data infrastructure like EUDAT. In this context, iRODS is important to support two types of integration. One aims to hide the complexity of the technology, providing a common layer on top of which is possible to build a shared set of policies for the data management. We call it horizontal integration, because it spans across data centers and communities forming a common base to integrate them at the same level, closing the gap, which often divides the first ones, more focused on the low-level technology layer, from the latter, more interested into the high level policies. Indeed, the heterogeneity is not only at technical level, but often also the maturity of the different communities differs in terms of awareness and knowledge of data management best practices and policies. To impose top-down solutions to the researches is usually counterproductive, hence it is necessary to help them customizing the services to make them closer to their daily practice and we call it vertical integration. The B2SAFE service is an example of this approach, because it implements some fundamental data management workflows, like the data replication and the assignment of globally discoverable identifiers, which can be used as building blocks from the users to define more complex and customized data policies.

## REFERENCES

[1] EUDAT, https://www.eudat.eu, Visited last on 24.05.2017

[2] Lecarpentier, D., De Sanden, M.v., Wittenburg, P.: Towards A European Collaborative Data Infrastructure. Multi Science Publishing, doi:10.1260/2047-4970.1.0.233 (2012)

[3] B2SAFE service, https://www.eudat.eu/services/b2safe, Visited last on 24.05.2017

[4] iRODS icommands, https://docs.irods.org/4.2.0/icommands/user, Visited last on 25.05.2017

[5] WebDAV protocol, http://www.webdav.org, Visited last on 25.05.2017

[6] Globus GridFTP server, http://toolkit.globus.org/toolkit/docs/latest-stable/gridftp, Visited last on 25.05.2017

[7] B2STAGE DSI library, https://github.com/EUDAT-B2STAGE/B2STAGE-GridFTP, Visited last on 25.05.2017

[8] iRODS federation, https://docs.irods.org/4.2.0/system_overview/federation, Visited last on 25.05.2017

[9] B2SAFE module, https://github.com/EUDAT-B2SAFE/B2SAFE-core, Visited last on 25.05.2017

[10] Digital Preservation Handbook, second Edition, http://handbook.dpconline.org/, Digital Preservation Coalition (2015), Visited last on 25.05.2017

[11] B2HANDLE service, https://www.eudat.eu/services/b2handle, Visited last on 29.05.2017

[12] Handle system, https://www.handle.net, Visited last on 29.05.2017

[13] DONA, https://dona.net, Visited last on 29.05.2017

[14] B2HANDLE library, http://eudat-b2safe.github.io/B2HANDLE/index.html, Visited last on 29.05.2017

[15] Principles and Good Practice for Preserving Data, http://www.ihsn.org/sites/default/files/resources/IHSN-WP003.pdf, Interuniversity Consortium for Political and Social Research (ICPSR) (2009), Visited last on 29.05.2017