



European Data

Grant agreement number: RI-283304

## D6.4: Final Report on Operations

Authors	Johannes Reetz (RZG), Giovanni Morelli (CINECA), Cristina Manzano (JUELICH), Kostas Kavoussanakis (UEDIN), Pascal Dugenie (CINES), David Vicente (BSC), Urpo Kaila (CSC)
Contributors	Claudio Cacciari (CINECA), Heinrich Widmann (DKRZ), Florian Kaiser (RZG), Robert Verkerk (SURFsara), Dejan Vitlacil (SNIC), Carl Johan Hakansson (SNIC), Pietari Hyvärinen (CSC)
Status	Final
Version	v1.0
Date	04/05/2015

**Abstract:** WP6 “Operations” makes the EUDAT Collaborative Data Infrastructure available for scientists and researchers by coordinating the provisioning, integration, operation and support of distributed services and resources in a lean way. This final report on the status and progress of operations is presenting an update of the 2<sup>nd</sup> year report (D6.3). In the last project year, progress has been made by enabling and providing B2DROP, B2SHARE, B2SAFE, B2STAGE and B2FIND, particularly in the context of the collaboration project from the first call in Feb 2014. The *Helpdesk System* has been further developed, new *User Documents* have been released, a *Storage Accounting system* is in place, the *SLA template for B2SAFE* has been updated and a new one for B2SHARE has been released, and the provisioning of further EPIC PID services has been completed.

Document identifier: EUDAT-DEL-WP6-D6.4	
Deliverable lead	RZG
Related work package	WP6
Authors	Johannes Reetz (RZG), Giovanni Morelli (CINECA), Cristina Manzano (JUELICH), Kostas Kavoussanakis (UEDIN), Pascal Dugenie (CINES), David Vicente (BSC), Urpo Kaila (CSC)
Contributors	Claudio Cacciari (CINECA), Heinrich Widmann (DKRZ), Florian Kaiser (RZG), Robert Verkerk (SURFsara), Dejan Vitlacil (SNIC), Carl Johan Hakansson (SNIC), Dejan Vitlacil (SNIC), Pietari Hyvärinen (CSC), Ari Lukkarinen (CSC)
Due date of deliverable	31/03/2015
Actual submission date	04/05/2015
Reviewed by	Erwin Laure (SNIC), Ioannis Liabotis (GRNET)
Approved by	PMO
Dissemination level	PUBLIC
Website	www.eudat.eu
Call	FP7-INFRA-2011-1.2.2
Project number	283304
Instrument	CP-CSA
Start date of project	01/10/2011
Duration	42 months

*Disclaimer:* The content of the document herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the document is believed to be accurate, the author(s) or any other participant in the EUDAT Consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the EUDAT Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the EUDAT Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

## CONTENTS

EXECUTIVE SUMMARY .....	5
1. INTRODUCTION .....	6
2. GENERAL OPERATIONS.....	8
2.1. Operations Coordination.....	8
2.2. Central Sites and Services Registry .....	9
2.3. User Community Support (Helpdesk) and the Trouble Ticketing System.....	10
2.3.1. Trouble Ticketing System and Support Channels.....	10
2.3.2. Contact and Request Form .....	11
2.3.3. Use of the EUDAT Contact Form and Helpdesk Service .....	11
2.4. User Documentation.....	12
3. SERVICE AND RESOURCE PROVISIONING .....	14
3.1. Principles .....	14
3.2. Service Inventory of the CDI.....	14
3.3. Storage Accounting.....	16
3.4. Operational Tools .....	17
3.4.1. Site- and Service-Registry .....	17
3.4.2. Resource Coordination Tool (RCT) .....	18
3.4.3. Monitoring.....	18
3.5. EUDAT AAI Services.....	18
3.6. Coordinating the Provisioning of EPIC/Handle PID services within EUDAT .....	19
4. CORE SERVICE ENABLING .....	21
4.1. What is a Data Project and where do Data Projects come from? .....	21
4.2. What is a Service Enabling Team? .....	22
4.3. B2SAFE and B2STAGE.....	23
4.4. B2SHARE.....	25
4.5. B2FIND .....	26
4.5.1. Collaboration Projects and Related Communities .....	26
4.5.2. 2nd Level Support .....	27
5. OPERATIONAL SECURITY .....	28
5.1. Security Policy.....	28
5.2. Acceptable Use Policy, Terms of Use.....	28
5.3. Joint Security Framework with other e-infrastructures.....	28
5.4. Security Training .....	28
5.5. Security Assessments.....	28
5.6. Risk Management .....	29
6. QUALITY ASSURANCE.....	30
ANNEX A. GLOSSARY .....	31
ANNEX B. DATA PROJECTS .....	37
ANNEX C. CONFIGURATION MANAGEMENT USING THE CENTRAL REGISTRY .....	39
ANNEX D. USER DOCUMENTATION TEAM AND DOCUMENTS.....	47

ANNEX E. SERVICE QUALITY ASSURANCE PLAN .....	50
ANNEX F. SERVICE LEVEL AGREEMENT FOR B2SAFE AND B2SHARE.....	60

## LIST OF FIGURES

Figure 1: EUDAT Infrastructure (March 2015) .....	6
Figure 2 Operations Coordination Team in 2014/2015.....	8
Figure 3: The Central Sites and Services Registry for CDI Configuration Management .....	9
Figure 4: B2-Services Helpdesk Queues.....	10
Figure 5: a) External Requests per Service Area b) Internal and External Requests per Service Area ..	11
Figure 6: Internal and External Requests (Tickets) per Queue .....	12
Figure 7: Used Storage Space.....	15
Figure 8: EUDAT Storage Accounting .....	16
Figure 9: Default Storage Accounting Panel from the RCT .....	17
Figure 10: EUDAT AAI Components.....	18
Figure 11: a) # of Projects in Production/Service/Resource b) Total Number of Data Projects per Site ...	23
Figure 12: B2SHARE Upload Panel with Customizations .....	25
Figure 14: The CDI as Network of Generic and Thematic Centres (according to WP2) .....	39
Figure 15: EUDAT Instance of the GOCDB showing the List of Registered Sites.....	40
Figure 16: Site Panel (upper part) with the Site Description .....	41
Figure 17: Site Panel (lower part) with Local Service Endpoints (Service Type, Production Level) .....	42
Figure 18: Service Endpoint Panel with Extended Properties.....	43
Figure 19: Service Groups Panel showing the available Service Groups in EUDAT.....	44
Figure 20: Particular Service Group Panel for CLARIN with the Related Service Endpoints.....	45
Figure 21: Monitoring of the Service Group (CLARIN) showing the Related Service Endpoints.....	46
Figure 22: EUDAT User Documentation Entry Point (Mar 2015) .....	47

## LIST OF TABLES

Table 1: Data projects regarding B2SAFE and B2STAGE (Status March 2015) .....	24
Table 2: Data Projects indicating Start Time (“pre sales”), Call for Collaboration Proposal and Status ....	37
Table 3: GOCDB RESTful API used in EUDAT .....	46
Table 4: User Documentation published on the EUDAT website <a href="http://eudat.eu/userdoc">http://eudat.eu/userdoc</a> .....	48

## EXECUTIVE SUMMARY

WP6 “Operations” makes the EUDAT Collaborative Data Infrastructure (CDI) available for scientists and researchers by coordinating the provisioning, integration, operation and support of distributed services and resources in a lean way. Since Oct 2011 WP6 has established a functioning distributed operational infrastructure with 13 generic data centres in partnership with a growing number of collaborating research communities, thematic centres and research infrastructures that are using the core services. This document presents mainly the achievements of WP6 since Oct 2013.

As intermediate decision preparing forum the Operations Coordination team, a network of site deputies, WP6 task and team leaders, was useful in establishing the CDI. This forum allowed taking up new demands and requirements on a short time scale.

The User Documentation has been consolidated and enhanced in collaboration with WP5.

The Helpdesk system has been refactored and new queues have been introduced. A user support web-form for external requests is available that allows external users to address their requests to specific Helpdesk channels.

The number of EPIC/Handle PID service providers in EUDAT has been increased from two to four, and a coordinator for the provisioning of PIDs services in EUDAT is acting as point of contact for PID prefix managers as well as for EUDAT PID service providers.

The operational tools have been upgraded, enhanced and further developed:

- the site and service registry has been upgraded to version GOCDB 5.2 which supports property extensions for sites, services and service groups, and an enhanced API – features that have been requested by EUDAT. The naming of the service groups and service types has been revisited and streamlined.
- the Resource Coordination Tool has been further developed (new views) and the content has been consolidated.
- the first version of a storage usage accounting has been developed and accounting information is collected for B2SAFE, B2SHARE and B2DROP. The storage accounting records are interoperable.

The CDI is now providing a full set of core services (B2-Services) and service enabling activities for these B2-Services in context of community data projects have been intensified. Service enabling teams have realized several projects in collaboration with communities and shared their knowledge. The implementation of proposals that were received from the Call for Collaboration in January 2014 has been supported. Existing data projects have been maintained and upgraded by the service enabling teams and the staff members at the service and resource provider sites. The 2nd level support has been provided by the service enabling teams and issues have been communicated back to the service developers.

Site security assessments, security audits and security scans have been regularly conducted and security incidents as well as known security vulnerabilities have been continuously managed.

A service quality management plan has been provided and the SLA template for B2SAFE and been extended to make it applicable also for B2SHARE.

## 1. INTRODUCTION

The purpose of WP6 (Operations) is to make the EUDAT collaborative data infrastructure (CDI) available to users by integrating and operating EUDAT's services on top of the storage, compute and network resources, that are provided by a number of geographically distributed European data centres, in a lean coordinated way.

The *lean coordination* of a range of tasks – namely, the provisioning of services and resources by the centres, the *research community-specific integration*<sup>1</sup>, the *production of documentation and providing support for the services handed over from WP5*, and *monitoring operational security and quality assurance* – have been developed under the supervision of the *Operations Coordination Team*. The team has worked in close cooperation with the other EUDAT work packages and been inspired by the pilot projects that were conducted in collaboration with various research communities over the last year.

During the 42 months since the EUDAT project started, WP6 has established a functioning distributed operational infrastructure from scratch. This infrastructure incorporates 13 generic data centres in partnership with a growing number of committed research communities, thematic centres and research infrastructures (see Figure 1).

As we have previously reported, during the second year of the project, WP6 primarily focussed on shaping the CDI by including additional generic data centres and further community partners (that is, data repositories). However, over the last 18 months the focus shifted so that WP6 was concentrating more on the enabling and operation of *data projects*<sup>2</sup> using one or more EUDAT core services in the context of specific scientific use cases or usage scenarios. These core services, branded as B2-Services, are described in the recent deliverables of WP5 (D5.3) and WP2 (D2.1.3).

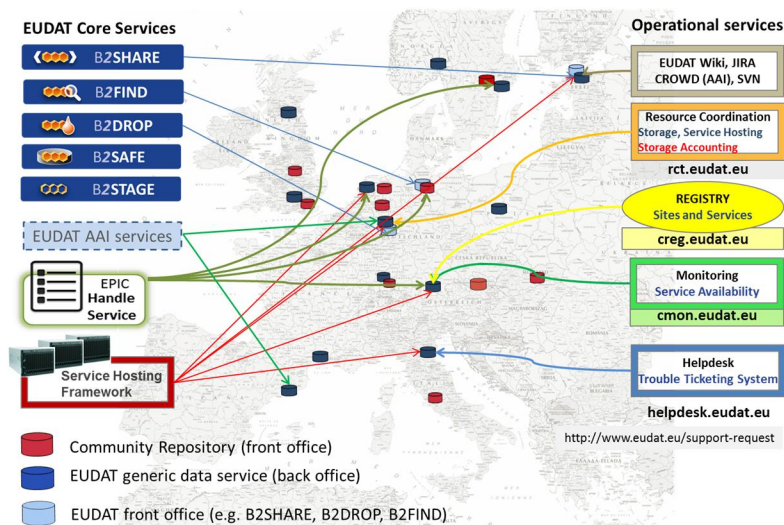


Figure 1: EUDAT Infrastructure (March 2015)

The work on establishing a CDI has been very successful. In January 2014, EUDAT put out its first call for collaboration<sup>3</sup> and received 40 proposals for collaboration projects from research communities and

<sup>1</sup> *Service enabling*

<sup>2</sup> The term "Data Project" is explained in section 4.1, p.22

<sup>3</sup> <http://www.eudat.eu/news/eudat-calls-collaboration-projects>

organisations. Thanks to the work on the CDI, it was possible to do the planning and start the implementation of 14 of these projects, and some of them are already in operation within WP6<sup>4</sup>.

Figure 1 shows the 13 generic data centres (in dark blue) that are currently participating in the EUDAT project, together with some of the research community centres (which are essentially data repositories) that are collaborating on the CDI. Sites that are offering EUDAT front office services (via a web interface) – such as B2DROP, B2SHARE and B2FIND – are indicated by the light blue colour. The generic data centres and their service components are registered by the central registry (shown in yellow); only community centres which decided to “join” the CDI are registered<sup>5</sup>. The generic data centres are running basic service components for the B2SAFE service and they can create and manage persistent identifiers (PIDs) under their own Handle prefix. Many of these centres also offer the gridFTP-based *Data Staging Interface* (DSI) on top of iRODS which makes it possible to access and stage data in and out. All the EUDAT service components are centrally monitored. The resources for the storage and service hosting capacity – which are pledged by the partners but not necessarily allocated – are coordinated and registered project-wise via the *Resource Provisioning Coordination Tool* (RCT). Figure 1 also indicates the sites that are running the operational services, the helpdesk system, the EPIC handle services and the AAI services.

Section 2 provides an update on the current status and the progress to date in coordinating and developing operations with particular regard to the advancements in operational tools and services needed for a lean operation. This section includes a report on the benefits of the upgrade of the *central site and service registry*, and provides statistical information about the usage of the helpdesk system and the organisational progress made by the *support team*. We also present the final achievements of the *documentation team*, which maintained the existing documentation and also released new documents.

Section 3 describes the achievements in *service provisioning* and *resource provisioning* with particular regard to the *utilized* resources provided by the partners. In addition, this section describes the recently developed storage usage accounting approach and details the first steps towards a production AAI service and the coordinated provisioning of the current four distributed EPIC/Handle services.

This document contains an additional section in comparison with the previous deliverable D6.3. This new Section 4 discusses the core service enabling that happened in the course of various collaboration projects (that is, joint ventures with research communities that were advantageous for the communities and EUDAT). The great thing with this type of collaboration is that the research communities and institutions can establish good practises for handling their research data while EUDAT receives valuable feedback and validation information on the appropriateness and usability of its B2-Services.

Section 5 presents the work and progress undertaken in the field of operational *security* and Section 6 describes the work that was done in the domain of *quality assurance and service quality management*.

The annex comprises a selection of documents which have been created during the reporting period: an illustrative description of the usage of the site and service registry used by EUDAT, a more detailed overview over the User Documentation team and the User Documentation, a service quality assurance plan, and the *SLA template for B2SHARE* with an update of the *SLA template for B2SAFE*.

---

<sup>4</sup> Some of the 40 proposals are handled by other work packages.

<sup>5</sup> The concept of “joining” and “using” is further explained in deliverable D2.1.3 and the service rolling plan.

## 2. GENERAL OPERATIONS

### 2.1. Operations Coordination

The purpose and scope of Operations Coordination have been described in the previous deliverable<sup>6</sup> (D6.3, pp.8). Also the *Guidelines for Integrating New Centres and Services*<sup>7</sup> which defines the necessary roles<sup>8</sup> and responsibilities are still relevant and have been taken into account.

The goal to establish a *lean operation* of the CDI requires efficient communication via operational tools, a trouble ticketing system and the avoidance of any unnecessary overhead that could overburden the staff members at the CDI constituting sites. The Operations Coordination Team<sup>9</sup> ensured the close collaboration between Service and Resources providers and the Service enabling teams in an efficient way.

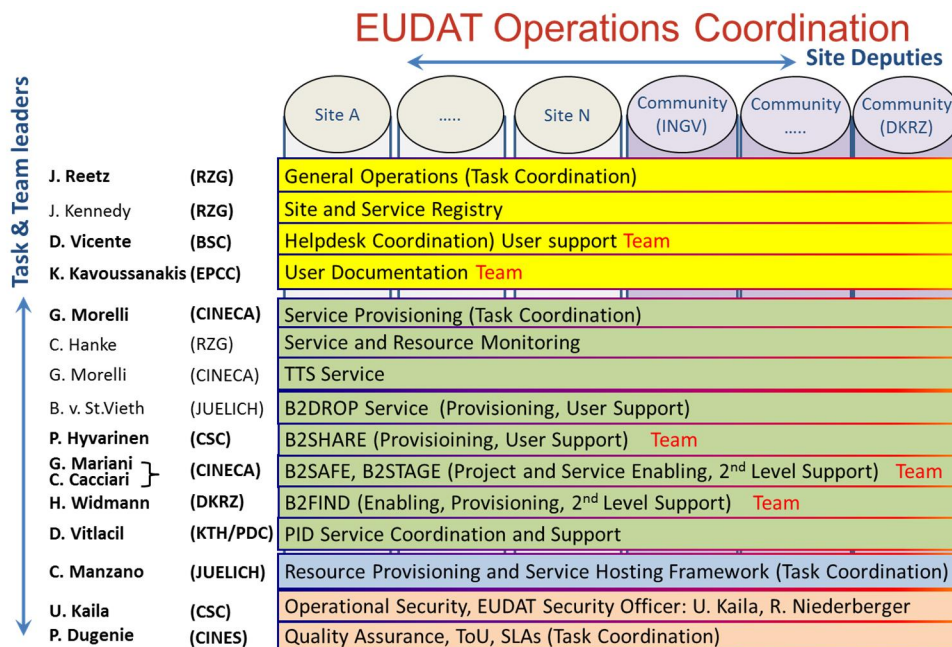


Figure 2: Operations Coordination Team in 2014/2015

Beside the Trouble Ticketing system for the helpdesk and issue management, EUDAT Operations is using a few operational tools (see Section 3.4) which are basically operated<sup>10</sup>, administrated<sup>11</sup> and managed<sup>12</sup> collaboratively by the sites. The sites are represented by site deputies who ensure that the necessary roles (site administrator, local service and resource manager, site security officer) are properly implemented. The role of a central operations coordinator has been taken by the task leaders of General

<sup>6</sup> D6.3, <http://www.eudat.eu/d63-2nd-year-report-status-and-progress-operations>, 02/12/2013

<sup>7</sup> D6.2, pp.53, <http://www.eudat.eu/d62-1st-year-report-status-and-progress-operations>, 09/11/2012

<sup>8</sup> D6.2, pp.52

<sup>9</sup> See also D6.3, section 2, for the detailed presentation of the (virtual) Operations Coordination Team.

<sup>10</sup> "operation" in terms of making the service available

<sup>11</sup> "administration" in terms of role and access management.

<sup>12</sup> "management" in terms of information management, i.e. making sure that the content of the operational tool is up-to-date.

Operations and Service Provisioning which was necessary particularly during the ramp-up phase of the EUDAT CDI.

Figure 2 shows the setup of the Operations Coordination Team during the year of the EUDAT project. Operations Coordination Team meetings were convened monthly. The figure indicates the responsibilities which are distributed across nine collaborating centres which altogether ensured an active knowledge exchange.

## 2.2. Central Sites and Services Registry

The EUDAT Registry stores and provides information about the sites and services which comprise the EUDAT CDI. It describes the configuration of the CDI with its frontend and backend services and service providers. The registry is based on the GOCDB<sup>13</sup> instance which is in production since July 2012. Throughout the EUDAT project we have worked closely with the GOCDB developers from STFC to aid with issues of customization and to define the functionality required by EUDAT.

The Registry service provides functionality which covers the initial EUDAT use-cases including:

- Web based user interface allowing users to add/edit/query information about EUDAT sites and services
- REST API for programmatic queries (read only)
- Fine grained Authorization model
- Flexible data model which allows sites and services to be grouped into logical domains(units)

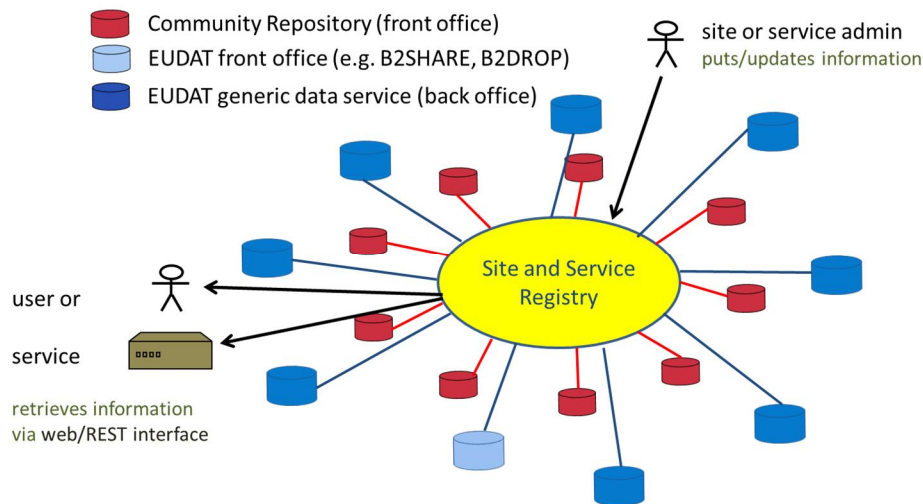


Figure 3: The Central Sites and Services Registry for CDI Configuration Management

Within EUDAT the Registry service is a core operational component of the CDI which is needed for the configuration management relevant for users (admins of generic data centers or community partners' data centers) and services (agents) alike (see also Annex B). One major example of this usage is that of the EUDAT monitoring services which use the Registry as an authoritative source of information to determine which CDI services need to be monitored. The monitoring service makes use of the REST API

<sup>13</sup> See D6.2, pp.9, section 2.3

to gather information about sites/services and construct monitoring test in real time. Thus once new services are added to the Registry they are automatically incorporated into the monitoring system.

In January 2014 the EUDAT registry was upgraded to the latest GOCDDB release (5.2). With this release came several new features which met the evolving requirements of the EUDAT.

- Allow arbitrary key/value pairs to be assigned to a site/service/group (property extensions)
- Improved Administration functionality within the GOCDDB
- The possibility to use Additional Authentication/Authorization mechanisms
- Support for Multiple projects within a single GOCDDB instance
- Enhanced API

Each of these features provides functionality which was requested by EUDAT and which enhances the usability of the GOCDDB for EUDAT and for the communities which EUDAT supports.

The addition of the key/value pairs allows and the ability to support multiple projects are of particular importance for the configuration management and community support. E.g. these allow communities which are collaborating with EUDAT to define their specific project and also to tag services with community specific meta information.

In addition to working with the GOCDDB developers to steer the developments in a collaborative way, EUDAT has also developed a suite of python scripts which allow users and services to interface with the GOCDDB for both, read/write purposes. This suite of scripts extends the functionality of the GOCDDB REST API to allow write as well as read operations. Currently the EUDAT Registry holds information about 20 sites (apart from generic data centers also some repositories), 69 services and it has 60 registered users.

## 2.3. User Community Support (Helpdesk) and the Trouble Ticketing System

Beside the ongoing support activities<sup>14</sup>, the “contact and request form” and the communication channels for the 2nd level support have been improved.

### 2.3.1. Trouble Ticketing System and Support Channels

CINECA is hosting the *BestPractical Request Tracker*, as Trouble Ticket System (TTS)<sup>15</sup> which is customised with advanced features for the EUDAT helpdesk system. The TTS is used for tracking and logging contact requests, project requests, resource requests, service requests, service enabling and issue reports, mainly addressed to the network of generic centres. The 1<sup>st</sup> level support is responding on a short timescale, normally in less than few office hours) to contact requests and it forwards the requests and issue reports to the 2<sup>nd</sup> level support. The 2<sup>nd</sup> level support is provided by the enabling teams for the B2-services making

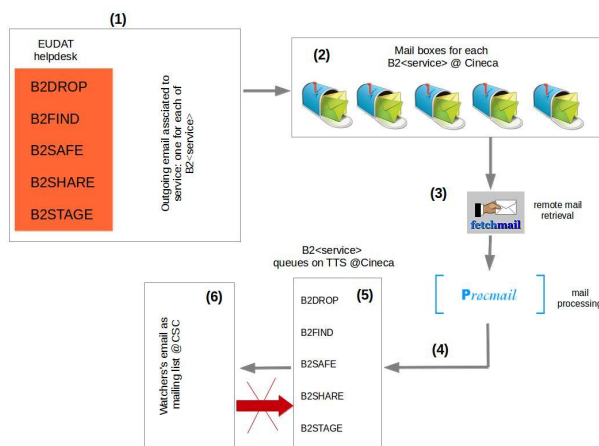


Figure 4: B2-Services Helpdesk Queues

<sup>14</sup> See EUDAT-DEL-D6.3-2<sup>nd</sup> Year Report on Operations, section 2.5 (pp.12)

<sup>15</sup> CINECA is hosting the BestPractical Request Tracker, customized with advanced features for the EUDAT helpdesk system.

use of dedicated TTS queues particularly for the communication with external parties (Figure 4).

The TTS has been further developed during the last year. Especially the EUDAT’s Contact and Support Request form on the EUDAT website has been connected.

The configuration of the TTS (queues, reminder and escalation procedures) follows the requirements of the different support levels, the users and the project management which is tracking the service enabling activities.

### 2.3.2. Contact and Request Form

The contact and request form<sup>16</sup> is accessible via the EUDAT webpage<sup>17</sup> and is linked also at the service portals for B2DROP, B2SHARE and B2FIND. The form is based on a PHP script using the Drupal module “Triggered Rules” which redirects to specific queues which are defined on the TTS. This script makes sure that the corresponding helpdesk mail addresses in the backend are not exposed to the public users. This helps to avoid SPAM as well as the fact that the form is also protected by a CAPTCHA mechanism.

Users are submitting their support request via the request form indicating the related services<sup>18</sup> (1). Each request creates an email message which is sent to a dedicated service support address (2). A *fetchmail/procmail* mechanism (3,4) retrieves the messages from the queue-specific mail boxes and dispatches them by creating new *tickets* in the corresponding queues (5). Each queue has a supervisor assigned (6) who gets notified by email. The queue supervisor makes sure that the queue has adequate watchers, i.e. experts for the specific services who can respond efficiently.

### 2.3.3. Use of the EUDAT Contact Form and Helpdesk Service

More than 60 different users contacted the EUDAT support service via the updated web form. These requests are automatically directed to the corresponding queues, depending on the given request details, or they are manually dispatched by the user support team. This is normally the case if the requester did not specify the related B2-service.

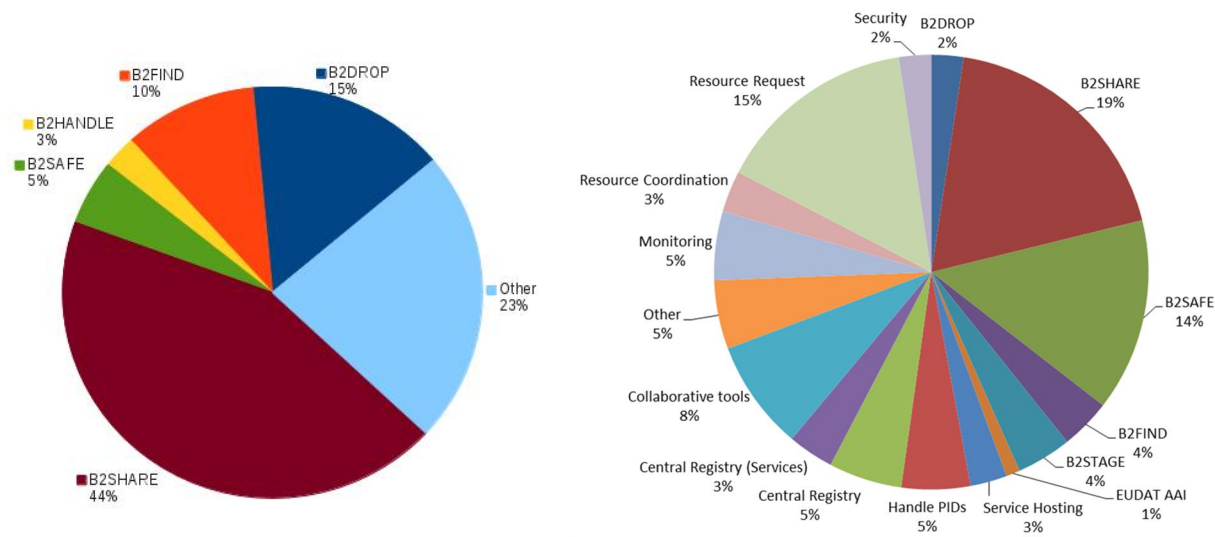


Figure 5: a) External Requests per Service Area b) Internal and External Requests per Service Area

<sup>16</sup> <http://www.eudat.eu/support-request>

<sup>17</sup> <http://www.eudat.eu>

<sup>18</sup> Unspecified requests are directed to the general support queue and the 1<sup>st</sup> level support will dispatch it appropriately.

Figure 5a shows the distribution of external requests via the web form during the last 12 months. Most external requests from this time period are related to B2SHARE, B2DROP or other aspects (contact requests). Figure 5b depicts the distribution of all the EUDAT internal and external requests over the TTS queues during the last three years. 70% of these tickets are internal ones. The large fraction of requests for B2SAFE is due to the fact that request tickets are used and kept open for tracking the service enabling activities. The large ratio of resource requests indicates that the TTS is used also as platform to coordinate the provisioning of storage and compute resources.

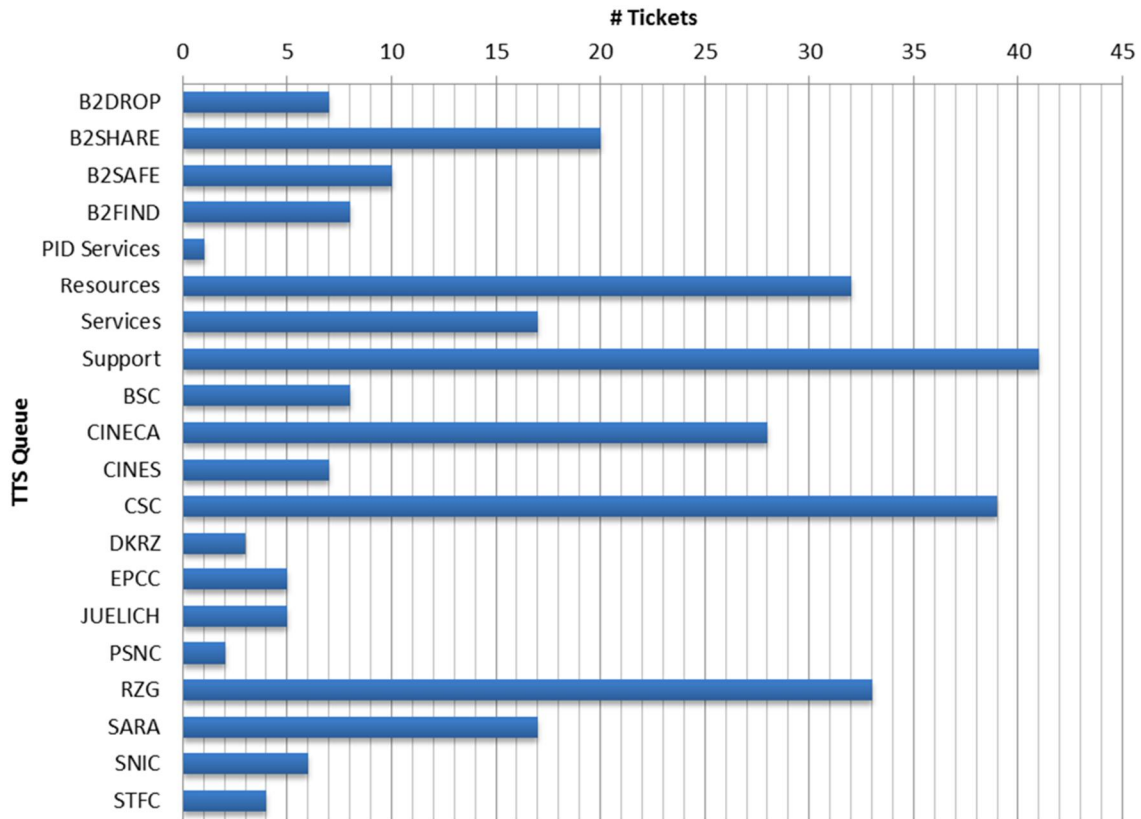


Figure 6: Internal and External Requests (Tickets) per Queue

Figure 6 displays the histogram of tickets per TTS queue (internal, external requests) from the last three years<sup>19</sup>. Recently introduced queues on B2-services received not yet many tickets. Beside larger queues for “resources”, “services” and the general support queue, the queues for CSC, CINECA, RZG and SURFsara are populated also because these sites are providing collaborative tools and central services. Queues for community centres have not been introduced yet.

## 2.4. User Documentation

The User Documentation activity of EUDAT plays an important role in engaging and retaining user communities and service providers. The *Documentation Team* compiled and created documentation for

<sup>19</sup> The queues are cleaned by removing inadequate tickets (SPAM messages, test tickets, and bulk notifications).

high-level and technical audiences by creating different documents covering the EUDAT services. The documents are published on the EUDAT website<sup>20</sup> under the categories of: Engage, Deploy and Use.

## Documents

At the end of 2013, the User Documentation had sufficient material available to cover the EUDAT service stack. Since then nine new documents have been added and seven existing documents have been updated in response to service changes. One of the changes was the update of the names of the services in line with the new branding. The team managed successfully the conflicting needs to complete documents while keeping abreast of new developments.

## Current Material

The User Documentation team has published 16 documents in collaboration with the service developers in WP5<sup>21</sup>. The full list is available from Annex D. Following documents are new in the present reporting period:

- B2SHARE User Documentation<sup>22</sup>: the B2SHARE is for researchers who do not have adequate facilities for depositing and describing their data, cannot guarantee for the persistence of their locally stored data and lack means to easily share with colleagues worldwide collections of small files and scientific data sets.
- B2FIND User Documentation<sup>23</sup>: B2FIND allows researchers to perform faceted searches on a central catalogue that is based on metadata that is harvested from various repositories.
- B2STAGE overview<sup>24</sup>, administrator<sup>25</sup> and Data Staging Script<sup>26</sup> documents: B2STAGE tools and interface services allow data to be transferred efficiently between EUDAT data nodes and external services and storage systems. These three documents provide the general introduction, configuration and User Documentation for the tool to transfer data between a EUDAT data node and an external GridFTP server.
- Monitoring user<sup>27</sup> and administrator<sup>28</sup> documents: community-user overview and configuration documents for the Nagios installation.
- How to deploy a EPIC/Handle service<sup>29</sup>: This document is encouraging and describing how sites can run and manage their own PID service based on the EPIC/Handle software stack.
- Site and Service Registry Administration<sup>30</sup>: The Site and Services Registry is an operational service that registers and provides information about the sites in the infrastructure. This document discusses the administration of the service for administrators of EUDAT sites, services and those responsible for community repositories that intend to be more integrated in the EUDAT CDI.

---

<sup>20</sup> <http://www.eudat.eu/userdoc/>

<sup>21</sup> See also the remarks on the documentation activities in the deliverable D5.3.

<sup>22</sup> <http://eudat.eu/User%20Documentation%20-%20B2SHARE.html>

<sup>23</sup> <http://eudat.eu/User%20Documentation%20-%20B2FIND.html>

<sup>24</sup> <http://eudat.eu/User%20Documentation%20-%20B2STAGE.html>

<sup>25</sup> <http://eudat.eu/User%20Documentation%20-%20B2STAGE%20for%20Site%20Administrator.html>

<sup>26</sup> <http://eudat.eu/User%20Documentation%20-%20Data%20Staging%20Script.html>

<sup>27</sup> <http://eudat.eu/User%20Documentation%20-%20The%20EUDAT%20Monitoring%20system.html>

<sup>28</sup> <http://eudat.eu/User%20Documentation%20-%20Monitoring%20System%20Operator%20View.html>

<sup>29</sup> <http://eudat.eu/User%20Documentation%20-%20How%20to%20deploy%20an%20EPIC%20service.html>

<sup>30</sup> <http://eudat.eu/User%20Documentation%20-%20Site%20and%20Service%20Registry%20Administration.html>

### 3. SERVICE AND RESOURCE PROVISIONING

This chapter provides an update of Section 3 of the report D6.3 excluding the description of the B2-Service enabling work which is now described in Section 4. It gives a brief overview over the provision of the B2-services, operational tools and how the provisioning of resources has been coordinated.

#### 3.1. Principles

EUDATs service provisioning is aiding the acceleration of *time to science* by making demanded resources and distributed services efficiently available in a reliable trustworthy manner. Trustworthiness does not just have a meaning for data depositors but also for the potential consumers of the data that, particularly in the science area, need to remain authentic<sup>31</sup> and untampered making sure that compromised data can at least always be identified<sup>32</sup> and eventually automatically healed. This requires logging and disaster recovery mechanisms which are, in the first instance, to be implemented by the sites locally (e.g. providing the adequate classes of storage featuring RAID, backup or replication capabilities). In addition, it requires a quality assessed organisation with administrative procedures under quality control (e.g. according to the FitSM<sup>33</sup> principles mentioned in Annex E and discussed in deliverable D2.3.2, Section 3.3 and Section 4). The EUDAT generic data centres and their staff members are experienced to provide trustworthy environments for their shareholders. It is the aim of these centres and their shareholders to enlarge this domain of trust also by harmonizing and improving the state-of-the-art of service provisioning by sharing and identifying the best practices of capacity and capability management, service deployment and operations, although each centre, each administrative domain, is individually responsible to comply with given standards and SLAs.

Each site is individually responsible for the quality of its Service and Resource provisioning. Some coordination effort is required to support community data projects to enable and utilize the integrated shared the B2-Services. The majority of activities regarding service provisioning is triggered and driven by the requirements from the community data projects. However, not all sites are expected to provide the same kind of services, service endpoints and resources if these are not required in the connection with concrete data projects.

#### 3.2. Service Inventory of the CDI

Important for the EUDAT CDI are the resources that are based on specific (storage) technologies and which are made available via the sites, and the suite of core services which are built and maintained by WP5 (see D5.3 about the recent developments). The core services *B2SAFE*, *B2STAGE*, *B2SHARE*, *B2DROP* and *B2FIND* have software components which are delivered in form of packages (see D5.3 Annex B) and are deployed at the different sites as indicated in Figure 1. The sites decide whether they install and provide certain service (components) and resources for a specific community projects. A decision can be dependent on the required compute capabilities, storage capacities and the network connection. The services *B2FIND*, *B2SHARE* and *B2DROP* are provided on top of resources at one site per service.

The B2-Services are complemented by additional services which could basically be outsourced but practically are provided via EUDAT. These are the EPIC/Handle PID services (section 3.6) used by *B2SAFE*, *B2SHARE* as well as by an increasing number of community repositories which are using and relying on

---

<sup>31</sup> In the context of service provisioning we are regarding the data items as they are, not considering higher levels of data description, provenance information and how to keep the links between data items authoritatively securely and safely bound.

<sup>32</sup> E.g. regular integrity checks, at least when the data is being reused or cited.

<sup>33</sup> <http://www.fedsm.eu/>

these services to register the data objects within their repositories. The EUDAT AAI service suite (section 3.5) and the service hosting framework (see D6.3, Annex G, pp.52) are also made available.

Service components for the following services are made available by the sites until end of March 2015<sup>34</sup>:

#### EUDAT B2-Service (core services)

- B2DROP: <http://b2drop.eudat.eu>  
(data sync and exchange based on ownCloud, <http://www.eudat.eu/b2drop>)
- B2SHARE: <http://b2share.eudat.eu>  
(register, describe and share data based Invenio, <http://www.eudat.eu/b2share>)
- B2SAFE: (safe data replication and handling, currently with iRODS and a customized set of rules and micro services, <http://www.eudat.eu/b2safe>)
- B2STAGE (data transfer service, <http://www.eudat.eu/b2stage>)
- B2FIND: <http://b2find.eudat.eu>  
(metadata harvesting and catalogue service, <http://www.eudat.eu/b2find>)

#### EUDAT coordinated core services

- EPIC/Handle PID services via SURFsara, DKRZ, KTH/PDC and RZG
- EUDAT AAI (UNITY via Juelich, <https://unity.eudat-aa1.fz-juelich.de:8443/home/home>)
- MyProxy services (SURFsara, CINECA)

#### EUDAT operational services

- Site- and Service-Registry: <https://creg.eudat.eu>
- Resource Provisioning Coordination Tool: <https://rct.eudat.eu>
- Trouble Ticketing Service: <https://helpdesk.eudat.eu>
- Nagios-based Monitoring: <https://cmon.eudat.eu>
- SVN repository: <https://svn.eudat.eu/EUDAT> (as an option to deposit software packages and configuration files within EUDAT)

EUDAT collaborative services: Atlassian Confluence, Jira with the Crowd SSO; Webserver with Drupal.

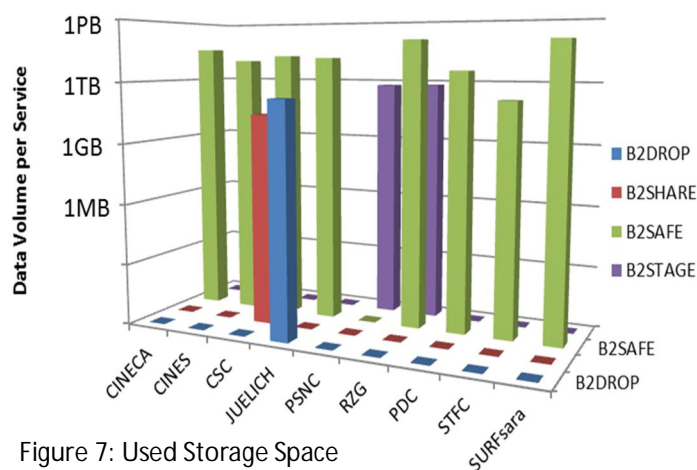


Figure 7: Used Storage Space

The provided service component and service groups are registered by the services registry (Annex B) and monitored using Nagios (D6.3, Annex H, pp.54).

A detailed list of the allocated resources per community, data project and site is accessible for entitled site and service administrators via the RCT.

Figure 7 shows the allocation of storage space for pre-production and production services. B2DROP and B2FIND are provided by CSC and Juelich.

<sup>34</sup> The site- and service registry is used to keep the detailed information up-to-date: about the service endpoints, the (required) installed software packages and the service groups to which the endpoints belong (Section 2.2, see also Annex C).

### 3.3. Storage Accounting

WP6 developed, deployed and is using a first version of a storage accounting system for EUDAT. The system allows obtaining an up-to-date overview about the distributed storage resources occupied by the registered data projects. The present implementation takes interoperability to other e-infrastructures and projects into account (e.g. principles from EMI and OGF). Developers of the EGI APEL<sup>35</sup> system were partly involved in the information exchange with EUDAT's storage accounting team.

EUDAT's accounting record (see Figure 8) follows the main principles of the EMI Storage Accounting Record StAR<sup>36</sup>. A RESTful interface is provided which allows other clients to obtain StAR-compliant records from the EUDAT accounting data base for analysis or information representation purposes.

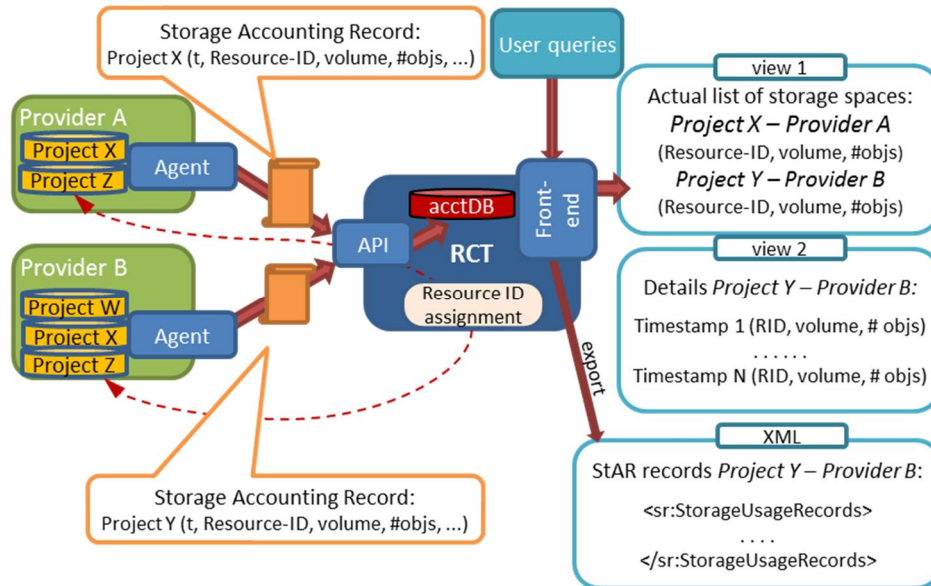


Figure 8: EUDAT Storage Accounting

Figure 8 shows the conceptual approach. The RCT<sup>37</sup>, the tool for coordinating the data project related capacities, is currently used as instance to host a central storage accounting data base. The RCT has two main functions: (a) creating unique resource identifiers which are then each linked to the dedicated storage (in terms of storage space and storage class) at a *Provider* site for a *specific Data Project* (red dashed arrows indicate the assignment of resource identifiers); (b) the occupied storage space (per provider, per project) can be then determined via agents which are connected to the providers' data services. The agents are pushing the information to the RCT which is kept in the attached accounting database. The functionality of the Resource Coordination Tool has been extended accordingly to support the provisioning of the unique resource identifiers and to link further information (storage class, provider, project) to them. Thus, RCT acts as central collector and also provides a default view of the storage accounting information to project supervisors and the entitled end users.

<sup>35</sup> See APEL's wiki: <https://wiki.egi.eu/wiki/APEL>

<sup>36</sup> See definition of the StAR accounting record: <http://cds.cern.ch/record/1452920?ln=en>

<sup>37</sup> For more details about the RCT see EUDAT-DEL-WP6-D6.3, section 3.3.1, p.23; EUDAT-DEL-WP6-D6.2, section 3.5.2, p.32 and annex C, p.56.

Each data service provider is responsible to obtain and apply the corresponding resource identifier (RID) for the storage accounting records that he regularly pushes to the accounting DB. For this purpose, a storage resource, when first assigned to a data service (e.g. an iRODS service endpoint) and to a data project, must be registered using the RCT and classified as either online, nearline or offline storage (with the specific multiplicity of copies on this resource).

For the different storage back-ends currently in use several collector agents<sup>38</sup> have been developed and are available. These agents have been adapted to the technology of the storage service. For example, a storage service via iRODS allows registering the resource identifier in the storage management database (iCAT); other kind of storage services (e.g. B2DROP, B2SHARE) need a different implementation to associate the accounting information to the corresponding RID.

The RCT accounting front-end shows the usage of the storage spaces for different B2SAFE-related projects, B2SHARE-related projects and for B2DROP.<sup>39</sup> The information about the maximum quota (if applicable in terms of volume and number of objects/inodes) is provided and listed together with the currently occupied storage space and the number of objects. This information is shown per project and provider. Another default view offers the possibility to follow the evolution of storage usage over time.

Community	Project	Storage Space	Provider	Service	Storage Class	Max Space	Max Objects	Used space	# objects
ABC	B2SAFE for ABC	ABC	BSC	bscirods BSC (B2SAFE)	nearline (2)	9.77 TiB			
BBMRI	B2SHARE	B2SHARE (BBMRI)	CSC	B2SHARE CSC	online+ (1)	10 TiB		271.05 MiB	6
CLARIN	B2SAFE for EKUT	EKUT-SR	RZG	irods0-eudat RZG (B2SAFE)	online (1)	200 GiB		66.84 GiB	1059
CLARIN	B2SAFE for LINDAT	LINDAT-SR	CINES	ariane CINES (B2SAFE)	nearline (2)	10 TiB		274.12 GiB	105
CLARIN	B2SAFE for TLA	REPLIX-SR	RZG	irods-replix RZG (B2SAFE)	nearline (2)	100 TiB		96.68 TiB	1428397
CLARIN	B2SAFE for TLA	REPLIX-SR	SURFsara	irods1 storage SURFsara (B2SAFE)	nearline (2)	80 TiB		6.48 TiB	125964
CLARIN	B2SHARE	B2SHARE (CLARIN)	CSC	B2SHARE CSC	online+ (1)	10 TiB		2.27 GiB	12
dxXa	B2SAFE for dxXa	dxXa-SR	CSC	eudat-res01 CSC (B2SAFE), eudat-res02 CSC (B2SAFE)	nearline (2)	1 TiB		158.21 GiB	15930
dxXa	B2SAFE for dxXa	dxXa-SR	STFC	eudat-node1 STFC (B2SAFE)	nearline (2)	1 TiB		158.25 GiB	15934
DRIHM	B2SHARE	B2SHARE (DRIHM)	CSC	B2SHARE CSC	online+ (1)	10 TiB		96.54 MiB	2
ENES	B2SAFE for CMIP5	CMIP5-SR	CSC	eudat-res01 CSC (B2SAFE)	nearline (2)	13 TiB		12.92 TiB	30525
ENES	B2SAFE for CMIP5	CMIP5-SR	JUELICH	eudat-s1 JUELICH (B2SAFE)	nearline (2)	13 TiB		13.08 TiB	30780
ENES	B2SAFE for IPSL	IPSL-SR	CINES	ariane CINES (B2SAFE)	nearline (2)	520 TiB		8.73 TiB	586
EPOS	B2SAFE for EPOS PP WG7	PPWG7-SR	CINECA	data repo CINECA (B2SAFE)	nearline (1)	45 TiB		33.67 TiB	8421096
EPOS	B2SAFE for ORFEUS@KNMI	ORFEUS	SURFsara	irods1 storage SURFsara (B2SAFE)	nearline (2)	30 TiB		27.51 TiB	8671052
EUDAT	B2DROP	B2DROP	JUELICH	B2DROP	online (2)	15 TiB		207.9 GiB	90856
EUDAT	B2SHARE	B2SHARE (generic)	CSC	B2SHARE CSC	online+ (1)	10 TiB		20.66 GiB	156
EUON	B2SHARE	B2SHARE (EUON)	CSC	B2SHARE CSC	online+ (1)	10 TiB		364.56 MiB	10
GBIF	B2SHARE	B2SHARE (GBIF)	CSC	B2SHARE CSC	online+ (1)	10 TiB		1.02 KiB	1
INCF	B2SAFE for INCF	INCF-SR	CSC	eudat-res01 CSC (B2SAFE), eudat-res02 CSC (B2SAFE)	nearline (2)	8 TiB		3.33 TiB	127814
INCF	B2SAFE for INCF	INCF-SR	SNIC	eudat-icat PDC (B2SAFE)	nearline (2)	8 TiB		3.36 TiB	129381
MIN	B2SAFE for MIN	MFN-SR	RZG	irods0-eudat RZG (B2SAFE)	nearline (2)	12 TiB		4.22 TiB	153601
NRM	B2SHARE	B2SHARE (NRM)	CSC	B2SHARE CSC	online+ (1)	10 TiB		18.91 KiB	1
RDA	B2SHARE	B2SHARE (RDA)	CSC	B2SHARE CSC	online+ (1)	10 TiB		179.27 KiB	1
VPH	B2STAGE+B2SAFE for VPH	VPH-SR	PSNC	eudat1.eudat PSNC (B2SAFE)	nearline (2)	20 TiB		570.84 GiB	491285
VPH	B2STAGE+B2SAFE for VPH	VPH-SR	RZG	irods3-eudat RZG (B2SAFE)	nearline (2)	20 TiB		610.15 GiB	495777

Figure 9: Default Storage Accounting Panel from the RCT

### 3.4. Operational Tools

This section recapitulates the main operational tools used for Services and Resources Provisioning.

#### 3.4.1. Site- and Service-Registry

The site and service registry is indispensable for a distributed infrastructure such as EUDAT. The EUDAT instance of the GOCDDB (section 2.2) is collaboratively administrated by the site administrators and service group managers, supported and supervised by the GOCDDB administrator. All sites in WP6 are

<sup>38</sup> The collector agents are available at <https://svn.eudat.eu/EUDAT/Services/ResourceProvisioning/frontend/samples/>

<sup>39</sup> The EUDAT storage accounting information is available under: [https://rct.eudat.eu/?path=/storage\\_space\\_resources/](https://rct.eudat.eu/?path=/storage_space_resources/)

registered as indicated by Figure 3. The relevant service endpoints are registered, grouped and supervisors/service managers are assigned (Annex B). The service types which are assigned to the service endpoints now follow the naming convention <service>.<product> (e.g. *b2safe.irods* instead of *irods*). Also the naming of the service groups has been consolidated using the B2-Services nomenclature.

The downtime scheduling is used by the site and/or service managers. The information about downtimes is propagated via the Nagios monitoring system and notifications are automatically sent to a dedicated EUDAT operations mailing list.

### 3.4.2. Resource Coordination Tool (RCT)

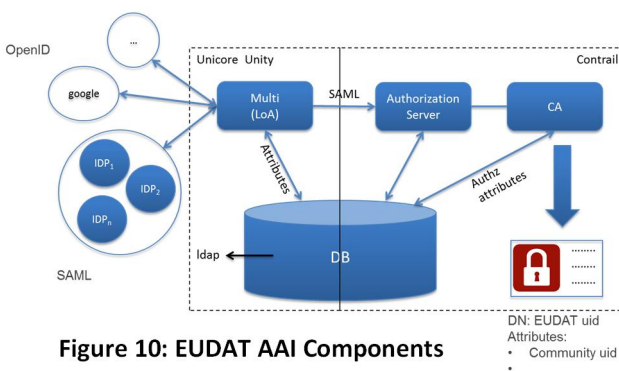
The functionality of the Resource (Provisioning) Coordination Tool has been further developed (new views, e.g., the storage accounting table; a tables that list the provided resources grouped by communities or by the data projects). The RCT became increasingly important not only for the project-wise capacity management (coordinating resources), but also for the coordination and management of the data project to which the resources are dedicated; a purpose that was originally not in the scope of the RCT. Actually a Data Project Coordination Tool is needed which aids planning, coordinating and managing not only the capabilities (services) and capacities (resources) but also coordinating the human resources which are often needed for realizing the community projects. The RCT is also used to generate unique Resource Identifiers when dedicated resources are registered as part of a data project. These are unique identifiers are used to identify the project-specific storage resources within the storage usage accounting system.

### 3.4.3. Monitoring

The central monitoring via <http://cmon.eudat.eu> has been supported and further documented<sup>40</sup>.

## 3.5. EUDAT AAI Services

The EUDAT AAI suite of service components is in detail described in section 3.6 of the deliverable D5.3. Core of the identity management service the Unity<sup>41</sup> multi server which acts like an IdP<sup>42</sup> that offers multiple protocols (oauth access tokens, x.509, SAML) from the perspective of service providers and it acts like a Service Provider from the perspective of SAML, OpenID or x.509 identity providers which are often part of larger national and supranational identity federations. The Unity frontend service has a user database in the backend which stores the unique EUDAT user identifiers and associated attributes. Users can register with Unity via a web portal and use it like a self-registration service. They can decide



**Figure 10: EUDAT AAI Components**

whether they want to create a new identity with a new username/password credential stored on the Unity server or whether they want to use another existing user identity and credential (from an external identity provider), e.g. one with a higher level of assurance, and bind them to their EUDAT user identity. The recent WP5 deliverable D5.3 presents a detailed viewgraph and describes the authentication flow and explains how to use and integrate the EUDAT

<sup>40</sup> <http://www.eudat.eu/User%20Documentation%20-%20The%20EUDAT%20Monitoring%20system.html>

<sup>41</sup> <http://www.unity-idm.eu/>

<sup>42</sup> Identity Provider to be understood in general terms here.

AAI service with the B2-Service components running at the EUDAT sites.

The EUDAT AAI service has been made recently available as a pre-production service<sup>43</sup> for service integration tests. First integration tests with a new version of the GOCDB were already successful. Also other operational tools such as the RCT, the helpdesk system and the central monitoring service will be connected then. Figure 10 shows the main components which are to be provided in an auditable secured and trustworthy environment when the software becomes production ready<sup>44</sup>. The Unity component is already centrally monitored. Other user and provider facing components such as the OAUTH and the EUDAT CA service (for short lived credentials) and the opened bridge operated at BSC will be monitored.

It is task of WP6 to ensure that the EUDAT AAI components are made available and securely operated in a trustworthy environment, that the necessary policy documents (Terms of Use, Terms of Conduct) are available and that the EUDAT AAI components are made acceptable for larger identity management and service provider federations. WP6 is also coordinating the 2<sup>nd</sup> level support for users and service providers. The appointed production service provider site is Juelich.

### 3.6. Coordinating the Provisioning of EPIC/Handle PID services within EUDAT

EPIC<sup>45</sup>/Handle PID services (version 2)<sup>46</sup> have been provided from the first project year. Initially, the PID service was provided by SURFsara only. SURFsara purchased a number of prefixes from CNRI for community services and for the general purpose data centres.

A trustworthy, highly reliable long-term sustainable data infrastructure for registered data, such as the EUDAT CDI intends to be, which itself relies essentially on the trustworthiness, reliability and the long-term sustainability of the providers of persistent identifiers, requires a distributed network of PID service providers which are operated in compliance with auditable policies which ensure that the credibility of each of the PID service providers remains on a high level. The reputation of the PID service providers can depend on the credibility/trustworthiness of each single PID service provider: if one PID provider fails, others if not the whole PID provisioning framework can be discredited.

In order to mitigate this risk and to ensure and sustain a high credibility of EPIC and EPIC PID service providers, EUDAT is supporting the EPIC consortium by providing EPIC PID services in compliance with the EPIC policies which, at the same time, are to be improved and strengthened. A corresponding coordination of Handle PID services within EUDAT has been introduced. This activity comprises the coordinated installation and change management of Handle PID services within EUDAT, and it takes care for and contributes to the collaboration with EPIC. Contributions can be: the EPIC membership, evaluating or specifying policies and operational guidelines, writing or improving the user and deployment documentation and providing EUDAT mirrors of PID services.

The coordination of the PID services in EUDAT comprises:

- acting as point of contact for requests from PID prefix managers and EUDAT PID service providers. This includes the 2<sup>nd</sup> level support with regard to the EPIC/Handle service stack.
- monitoring and eventually coordinating the provisioning of Handle prefixes.

---

<sup>43</sup> Beta service operated currently in the development environment at the Juelich; an openid bridge is running at BSC.

<sup>44</sup> EUDAT AAI productions readiness is expected in the middle of 2015.

<sup>45</sup> <http://www.pidconsortium.eu/>

<sup>46</sup> <https://github.com/CatchPlus/EPIC-API-v2>

- supervising the operation of the EPIC PID services that are provided via the EUDAT partners. This includes to check
  - that all the relevant PID service endpoints are registered and the monitored.
  - that all the PID services providers within EUDAT are performing, available, reliable and that their quality of service is in compliance with the EPIC policies.
- acting as EUDATs contact for EPIC and as such contribute to a high standard of the EPIC policies.

The following activities have been conducted:

- The documentation of the EPIC PID v2 service for users and administrators<sup>47</sup> has been written;
- Two additional PID services (PDC, DKRZ) have been enabled and integrated in the monitoring.
- The PID services have been registered in the GOCDB.
- Every PID prefix managed by EUDAT has a mirror.

Since the Handle PID service is developed and provided by the EPIC consortium, EUDAT also relies on support provided by EPIC. In 2014 the number of Handle PID service providers increased to four. Beside SURFsara there are now RZG, PDC and DKRZ available as Handle PID service providers. Three of them are EPIC members (SURFsara, PDC and DKRZ).

The two further candidate sites, EPCC and BSC, are in the software deployment and testing phase. Other sites expressed their intentions to join this distributed service infrastructure in the future.

---

<sup>47</sup> <http://www.eudat.eu/User%20Documentation%20-%20PIDs%20in%20EUDAT.html>,  
<http://www.eudat.eu/User%20Documentation%20-%20How%20to%20deploy%20an%20EPIC%20service.html>

## 4. CORE SERVICE ENABLING

The enabling of the B2-Services, the task to ensure that B2SAFE & B2STAGE, B2SHARE and B2FIND are utilized in the context of concrete data projects<sup>48</sup>, was in the focus of the activities of WP6 during the last 18 Months. New data projects from different research areas have been enabled and the existing (running) projects have been further supervised, supported and updated by new releases of the B2-Service packages. These projects were often joint efforts in close collaboration with the development team in WP5, particularly from B2SAFE and B2STAGE. But the enabling for B2SHARE and B2FIND also required the expertise from WP5. Enabling projects which turned out to become rather development-centric have been first handed over to WP5 but remains on the list (indicated as “planned”).

The enabling of core services conducted and supported by a network of experts has been introduced in the 2<sup>nd</sup> project year. The initial concept of the service enabling was presented in the deliverable D6.3 (Section 3.2.1, p.19). Since then the core services (B2-Services) have been enhanced, new ways to configure B2SAFE and B2STAGE have been explored, and the interactions with the community representatives in the context of the enabling projects have been further optimized. This was and still is a fruitful learning process for communities and IT-centric service providers.

The service enabling framework, the procedures and the network of experts for the service rollout and configuration, was also very useful for evaluating and implementing new projects from the first call of collaboration.

### 4.1. What is a Data Project and where do Data Projects come from?

A data project is owned by a community or a society that concerns a defined set of data collections<sup>49</sup>. In the scope of a data project are *Data* and *Users* – *Users* in terms of data depositors and data consumers. A data project that can be handled by WP6 must comprise a more detailed requirements specification that indicates which B2-Services are to be employed for the project. It must be clear how data has to be handled, stored, registered, replicated, transferred, indexed, made findable, accessible and auditable. It must also be clarified which capabilities and capacities regarding services, resources and the predicted data access pattern are necessary to implement and operate the data project over a certain span of time. The Resource Provisioning Coordination Tool was used to register the capability and capacity requirements (and inform the resource providers) particularly for the data projects with B2SAFE, B2STAGE and B2SHARE.

A data project has a start date and an end date. The start date is the date when the project was adequately defined, e.g., via a registered (accepted) project proposal. However, a project can start with a kind of “pre-sales” interaction when it is not yet clear whether the project, as it has been described, can already be enabled via WP6. This can depend on several factors such as the provisioning of sufficient scalable capacities and capabilities at the time when they are needed. The end date of a data project is typically a date in the distant future. It is important to note that data projects need to be prolonged at certain points in time which implies a decision of committed stakeholders also with regard to their investments to keep data and data facilities accessible and useful (latter often requires data curation activities).

---

<sup>48</sup> See the table in Annex B.

<sup>49</sup> See also the description in the Glossary (Annex A).

The realization (implementation, enabling) of data projects in the context of EUDAT can either be done by using the B2-Services as they are already available, or the project needs to be enabled, which in fact means that the project-relevant B2-Services are to be adequately configured across several sites<sup>50</sup>. Since many projects were initially requiring just one specific B2-Service we speak about “service enabling”.

The first data projects resulted from the service case descriptions of the communities which are (associated) partners in EUDAT. Many of these projects started already in 2012 and 2013.

Other collaboration projects were initiated during the frequent encounters of EUDAT with other communities. New projects were initiated during or after the EUDAT conferences<sup>51</sup>, User Forums<sup>52</sup>, Training events<sup>53</sup> and the working group activities<sup>54</sup>.

Annex B gives an overview about all B2-Service related enabling projects which WP6 have started to be plan (“pre-sales”), enable and put into (pre)production. 14 enabling projects are based on the collaboration proposals from this call.

## 4.2. What is a Service Enabling Team?

The service enablers are in contact with the users and they are aware of the community requirements, experienced with service development aspects as well as with the deployment and daily operation. Since each of the enabling projects can already become a demanding endeavour which can overburden individual experts, teams have been formed for each B2-Service. These teams were complemented by experts at from the community sides. The teams used the EUDAT wiki, EUDAT mailing lists, TTS queues and virtual meetings as shared platforms and to channel the flow of information, to keep the progress of the enabling projects tracked and to exchange experiences under the supervision of team leaders<sup>55</sup>.

Each enabling team has 2 – 10 members and is

1. acting as *point of contact* for any service-related requests from communities or EUDAT centres to provide enabling support or 2nd-level support for the specific B2-Service whose service components can be successively installed at the EUDAT sites.
2. monitoring the operation of “running” B2-Service instances after they have been installed and configured. The team leader is actually in charge to supervise the operations. The services should generally be operated under the responsibility of the local staff at each site. The team is only responsible to check the availability of the B2-Services for their data projects. And the service enabling team members need to take action only if issues occur which affect the quality of service and which is not or cannot be solved timely by the resource and service provider site. This includes

- to check that the service endpoints (of the relevant B2-Services) are properly registered in the site and services registry (<https://creg.eudat.eu>) and that the information is up-to-date.

---

<sup>50</sup> See [hdl:11304/bc0eafe4-f0dd-11e4-ac7e-860aa0063d1f](https://hdl.handle.net/11304/bc0eafe4-f0dd-11e4-ac7e-860aa0063d1f) as an example.

<sup>51</sup> <http://www.eudat.eu/events/eudat-conferences>

<sup>52</sup> <http://www.eudat.eu/events/event-userforum>

<sup>53</sup> <http://www.eudat.eu/events/trainings>

<sup>54</sup> <http://www.eudat.eu/working-groups>

<sup>55</sup> The largest team was the one for B2SAFE and B2STAGE which required particular expertise on PID and AAI services. The team was led by CINECA. The smaller team for the B2SHARE was led CSC and for B2FIND by DKRZ. Both sites are also hosting the corresponding services.

- to check that the adequate storage capacity is available also for the near future and otherwise has be requested using the RCT (<https://rct.eudat.eu>).
- to validate that all the service endpoints which are relevant for the data project are monitored (via <https://cmon.eudat.eu>) and that the quality of service is appropriate.

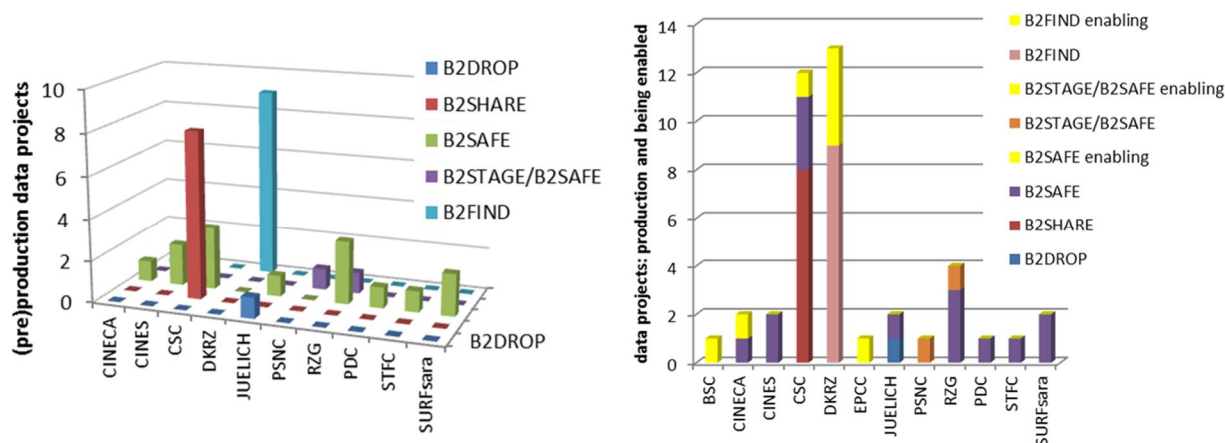


Figure 11: a) # of Projects in Production/Service/Resource b) Total Number of Data Projects per Site

Figure 11a gives an overview about which Resources Providers are currently involved per B2-Service in pre-production and production data projects. The amount of data projects does not reflect the total efforts spent per indicated site; the figure just indicates the involvement of the Resource Provider. For instance, the enabling effort per B2SAFE-related project is typically larger than for other B2-Services. The large number of projects in which DKRZ is involved is due the fact that this site is hosting the B2FIND service. The enabling work (the setup of the metadata harvesters for the different data sources) was shared across the team and supported by WP5<sup>56</sup>. The same holds for B2SHARE that is hosted by CSC. The B2SHARE-related projects concern the setup of particular community/project metadata extensions in close collaboration with WP5<sup>57</sup>.

Figure 11b indicates the site and their involvement as Resource Provider in production and ongoing data projects. Also in this case it is important to mention that the amount of projects per Resource Provider does not represent the total efforts that the partners spent on the enabling work.

### 4.3. B2SAFE and B2STAGE

The B2SAFE & B2STAGE enabling team has been prototypic for service enabling in general. Not only because it has been the first established enabling team, also because the workflow of the enabling process can be complex and very different from project to project. For instance, the configuration requirements for the data replication from the CLARIN centre MPI-PL (TLA) are different from the requirements of the CLARIN centre CUNI (LINDAT). Table 1 provides some details about the enabling projects: the start date, the repository from where the data has to be replicated, the storage providers for the 1<sup>st</sup> and 2<sup>nd</sup> replica, the storage class, the actual used volume and the number of registered data

<sup>56</sup> See deliverable D5.3, section 3.4

<sup>57</sup> see deliverable D5.3, section 3.3

objects (i.e. the number of PIDs). In case of multi-step replications, the primary replica site is listed first<sup>58</sup>.

The descriptions of the configurations for B2SAFE and B2STAGE in the projects are shared via the EUDAT wiki during the planning and enabling phase. Projects which have been enabled can be also described in the User Documentation, become part of success stories and partly they have been presented at conferences such as, for instance, a poster that shows examples of the B2SAFE configuration<sup>59</sup>.

Table 1: Data projects regarding B2SAFE and B2STAGE (Status March 2015)

#	Start	CfC <sup>60</sup>	Status	Enabler	Discipline	Repository	Provider	Class	MaxV	Volume	#PIDs
1	Dec 12		prod	E. Erastova, RZG	Linguistics	CLARIN: MPI-PL:TLA	RZG	nearline (2)	100TiB	96.68TiB	1.4M
							SURFsara	nearline (2)	80TiB	73.00TiB*	1M*
2	Dec 12		prod	S. Apweiler, Juelich	Climatology	ENES: DKRZ:CMIP5	JUELICH	nearline (2)	13TiB	13.08TiB	30.8k
							CSC	nearline (2)	13TiB	12.92TiB	30.5k
3	Dec 12		prod	C. Cacciari, CINECA	Earth System Science	EPOS: INGV-PP-WP7	CINECA	nearline (1)	45TiB	33.67TiB	8.4M
4	Dec 12		prod	J. Heikkinen, CSC	Neuro- science	INCF: [KTH] <sup>61</sup>	PDC	nearline (2)	8TiB	3.36TiB	129K
							CSC	nearline (2)	8TiB	3.33TiB	128k
5	Dec 12		enabling	E.Erastova, RZG	Lifescience, Bio medicine	VPH: [UCL]	EPCC				
6	Feb 13		pre-prod	R. Verkerk, SURFsara	Earth System Science	EPOS: KNMI:ORFEUS	SURFsara	nearline (2)	30TiB	27.51TiB	8.7M
7	Apr 13		Prod	E. Erastova, RZG	Linguistics	CLARIN: [EKUT]	RZG	online(1)	200GiB	67GiB	1k
8	Jul 13		prod	P. Dugenie, CINES	Linguistics	CLARIN: CUNI:LINDAT	CINES	nearline (2)	10TiB	274GiB	105
9	Jan 14		prod	E. Erastova, RZG	Biodiversity	[MfN Berlin]	RZG	nearline (2)	12TiB	4.22TiB	154k
A	Mar 14	#415	planned	G. Morelli, CINECA	Biodiversity	[iMarine]	CINECA				
10	Mar 14	#390	prod	J. Heikkinen, CSC	Toxicology	diXa: [EBI]	CSC	nearline (2)	1TiB	158GiB	16k
							STFC	nearline (2)	1TiB	158GiB	16k
B	Apr 14	#407	planned	S. Coutin, CINES	University	[IST]					
11	Apr 14	#402	pre-prod	S. Coutin, CINES	Climatology, Meteorology	ENES: [IPSL]	CINES	nearline(2)	520TiB	8.73TiB	586
12	Apr 14		pre-prod	E. Erastova, RZG	Lifescience, Bio medicine	VPH: [UCL]	RZG	nearline (2)	20TiB	610GiB	496k
							PSNC	nearline (2)	20TiB	562GiB	482k

<sup>58</sup> The numbers are taken from the actual storage accounting system with the exception that the replica volume at SURFsara (project 1 was much smaller at the time of writing because the replication from RZG to SURFsara had to be repeated. The given volume and number of objects (tagged by \*) are expected to be achieved soon.

<sup>59</sup> Elena Erastova, 2014, "B2SAFE Production Instances - Some Examples", [hdl:11304/bc0eafe4-f0dd-11e4-ac7e-860aa0063d1f](https://hdl.handle.net/11304/bc0eafe4-f0dd-11e4-ac7e-860aa0063d1f)

<sup>60</sup> Call for Collaboration: if applicable, the number identifies a related collaboration proposal.

<sup>61</sup> Brackets indicate names of the repository provider/owner which are not necessarily the concrete names of the repositories.

13	Jul 14	#406	enabling	H. Widmann, DKRZ G. Morelli, CINECA	High Energy Physics	ALEPH/DPHEP: [INFN]	CINECA
14	Jun 14	#422	enabling	J. Heikkinen, CSC	Lifescience	ELIXIR: EBI:PRIDE	CSC
C	Sep 14	#395	planned	D. Vicente, BSC	Biodiversity	LifeWatch [VZL]	BSC
15	Dec 14	#394	enabling	J. Heikkinen, CSC	Lifescience	ABC: [BSC]	BSC nearline (2) 10TiB

#### 4.4. B2SHARE

B2SHARE is EUDAT's repository service for storing, describing, sharing and publishing so-called "long tail" data. These are small and medium scale research data collections in various formats which are usually not covered by institutional data preservation policies. Technical details about the service are presented in section 3.3 of the deliverable D5.3.

EUDAT offers the B2SHARE service for European researchers and scientists. The service is hosted at the Finnish IT Center for Science, Datacenter CSC Kajaani, where also the data is stored. The centre is certified by the ISO/IEC 27001:2005 standard for its information security management system. More details about the conditions of this service can be taken from the FAQ<sup>62</sup>.

EUDAT retains the right of archiving the data, i.e. creating replicas at trusted centres to take care of their long-term persistence. Before depositing research data for the first time, the researcher must register to the service and accept the ToU<sup>63</sup>. Communities can obtain a customized interface for the description of

The screenshot shows the B2SHARE web interface. At the top, there is a navigation bar with the EUDAT logo and a B2SHARE banner. Below this, a user is logged in as 'jreetz'. The main area is divided into two steps:

**Step 01: Drag and drop files here**

This step includes a file upload area with 'Select files' and 'Start upload' buttons. Below the upload area is a table listing the files:

Filename	Size	Status	Remove
fig2.png	118 kb	[Progress bar]	[Remove icon]
EUDAT-DEL-WP6-D6.4-4.4-B2SHARE.docx	163 kb	[Progress bar]	[Remove icon]

**Step 02: Select a domain or project**

This step shows a grid of domain selection buttons:

- Linguistics (CLARIN)
- Herbarium
- Biodiversity
- Generic (EUDAT)
- Ontology (euon)
- RDA
- Hydro-Meteorology (DixHM)
- Biomedical Research

At the bottom, there is a footer with navigation links (Home, About, Terms of Use, FAQ, User Guide, REST API, Contact) and the INVENJO logo.

the deposited data before it is committed to the service. Requests for these specific metadata extensions are normally treated as (small) projects which are handled by a small B2SHARE service enabling team. Data which is described using one of these project-specific MD-templates can be accounted accordingly. The service is connected to the storage accounting system, so that information about the usage of the B2SHARE storage space and the number of registered objects can be reviewed on a daily basis.

The B2SHARE service was deployed as pre-production service in January 2014 and version 1.0 was released and deployed in production as the first stable version in late September 2014. Since then more than 180 different data sets

Figure 12: B2SHARE Upload Panel with Customizations

<sup>62</sup> <https://b2share.eudat.eu/docs/b2share-faq>

<sup>63</sup> <https://b2share.eudat.eu/docs/b2share-tou>

have been uploaded which corresponds to 21 GiB of research data.

The following community extensions have been enabled are supported by the service:

- Linguistics (CLARIN)
- Herbarium (NRM)
- Biodiversity (GBIF)
- Ontology (EUON)
- Hydro-meteorology (DRIHM)
- Biomedical research (BBMRI)
- Research Data Alliance (RDA)

## 4.5. B2FIND

B2FIND offers an interdisciplinary EUDAT metadata catalogue which allows users to easily find collections of scientific data that are managed by EUDAT services like B2SHARE or are described by metadata made available via external community archives and repositories, both from EUDAT and non-EUDAT communities.

The B2FIND service was released and deployed in production as the first stable version in February 2014. The service has been kept open for metadata providers since its first deployment. Metadata ingestion is based on a modular architecture and workflow. At the beginning of 2015 from ten communities more than 300.000 data entities are currently harvested, by default via the OAI-PMH protocol. Interaction with other communities exists and we expect further expansion w.r.t. the number of communities and the volume of harvested data.

In the following we present a summary of these communities and their specific issues and requirements. The community-specific needs lead not only to necessary adaptations of the service, but as well to more flexibility and better operability of the system. We split the communities according to their status and stage with regard to 'enabling for B2FIND':

### 4.5.1. Collaboration Projects and Related Communities

B2FIND production (community repositories are already continuously harvested)

1. ENES (research community of Earth System Modelling): ca. 1365 records harvested in geo-referenced specific MD format ISO 19115. The small number of records is caused by the low granularity, i.e. each metadata refers to high volume collections of data resources that comprise several TB's.
2. CLARIN TLA (Linguistics): more than 138000 records with specific MD-format 'IMDI' are harvested from the Text Language Archive (very fine-granular, i.e. each metadata record refers to a small data volume).
3. B2SHARE: currently ca. 180 records using MarcXML are regularly harvested from the generic B2SHARE service operated by CSC.
4. DataCite (data archive for citable research data): ca. 240000 cross-discipline records (proprietary MD schema, provided via OAI-PMH using the DublinCore format) harvested.
5. NARCIS/DANS (Dutch aggregator of social sciences): 175000 records harvested.
6. GBIF (research community of biodiversity): 13800 records harvested. Recently via JSON API, because former OAI provider (specific MDformat 'EML') is deprecated
7. The European Library (archive of cultural heritage): 27000 data sets as DublinCore harvested via OAI. Differentiated mapping of a wide field of disciplines needed.

8. PanData (repository of nuclear physics research data): 930 records and via STFC there will be an extended amount of data set taken into account.
9. Aleph (HEP data archaeology project at INFN): 185 records provided as MarcXML data. Together with the service B2STAGE prototype workflow 'Stage data selected by a B2FIND script for further HPC processing' has been set up.
10. SDL (Indian library, focused on agricultural data): 24000 records harvested.

#### B2FIND enabling started or planned (development of mapping)

1. IVOA (research community and repositories of Astronomy): records provided in DublinCore format (currently 177 data sets at the enabling start time).
2. CESSDA (Repository of social sciences): Part of the project DASISH, scheduled to end in summer 2015 and to be migrated to B2FIND within next weeks. Up to a million of records formatted in different metadata formats expected (current number of records at the start of the project: 740).
3. DARIAH (European Digital Arts and Humanities): The enabling of harvesting datasets from DARIAH has been started (5400 records so far). Many more records formatted in different metadata formats are expected.
4. BBMRI (Biomedical databases): Cooperation started, BBMRI has its own elaborated MD schema and model (MIABIS).
5. WDC-Terra and EPOS: (Seismological and geophysical data): Request by GFZ for publishing data in EUDAT B2FIND. In this context as well ongoing discussions with related EUDAT community EPOS.

#### B2FIND projects in the pipeline (contacts established but collaboration has not yet started)

1. ICOS currently building a Carbon Portal with metadata. Harvesting right now not yet possible.
2. agINFRA (agricultural data): Mapping tool MINT is used.
3. EISCAT (data from radar systems for studying the Earth's polar atmosphere): could provide data in ENVRI or ESPAS XML formats, but not yet via OAI-PMH.

For all harvested repositories and the actual number of records, see <http://b2find.eudat.eu/group>.

#### 4.5.2. 2nd Level Support

Especially in the initial phase of the collaboration project with new communities an intensive exchange and support is needed, to deploy stable harvesting services at the metadata provider side and to define and adapt the semantic metadata mapping to community-specific fields and formats. Especially with B2SHARE, which is operated at CSC, an intensive cooperation is ongoing and agreements on the mapping according to EUDAT central categories such as 'domains', 'disciplines' and specific data identifiers have been achieved. A metadata schema for B2FIND has been developed, which assures sustainable and reliable operability.

## 5. OPERATIONAL SECURITY

### 5.1. Security Policy

As described in the previous deliverable D6.3, a comprehensive security policy according to best information security practices was approved by EUDAT. The EUDAT security policy<sup>64</sup> describes a comprehensive framework for organising security related measures.

The key security roles appointed are the EUDAT Security Officer (Urpo Kaila, CSC), the EUDAT Deputy Security Officer (Ralph Niederberger, JUELICH), the EUDAT CSIRT Team and EUDAT Site Security Officers.

### 5.2. Acceptable Use Policy, Terms of Use

An acceptable use policy, conforming to the AUPs of other scientific infrastructures, has been included in the EUDAT Terms of Use<sup>65</sup>. The AUP comprises fundamental disclaimers and limitations of the role of EUDAT in order to mitigate potential technical, operational and legal risks. All core security related roles, obligations and rights have been included in EUDAT Terms of Use according to best information security practices conforming to those in other scientific infrastructures. The ToU has generally to be accepted by the users of EUDAT services. Acceptance of the ToU is anticipated if the EUDAT services are used.

### 5.3. Joint Security Framework with other e-infrastructures

A substantial effort has been made to create a joint Security Frameworks together with other e-infrastructures in Europe (EGI, PRACE) and abroad (e.g. OSG)<sup>66</sup>. Regular security meetings have been held with other e-infrastructures for developing a common framework and to handle security threats.

### 5.4. Security Training

Security trainings have been delivered during each EUDAT Technical Meeting (2012, 2013). A joint security training workshop together with other European e-infrastructures was held<sup>67</sup>. At the end of the training a pilot Site Security Officer Certification exam was arranged for the participants.

### 5.5. Security Assessments

A site and infrastructure based security review, done during 2012-2013, showed that security items still are to be implemented. The sites are progressing to improve their security measures. In addition, several technical security assessments and vulnerability scans of hosts and service endpoints, listed on the central registry, were performed in 2014 to proactively detect insecure configurations or unpatched services, including 'Heartbleed' vulnerability checks, on the EUDAT CDI. The EUDAT B2SHARE service was scanned frequently in 2014 for vulnerabilities. A special procedure was defined, on how the EUDAT Security Officer is alerting site deputies and site security officers about security vulnerabilities.

---

<sup>64</sup> see D6.3 Annex K

<sup>65</sup> see D6.3 Annex B

<sup>66</sup> see D6.3 Annex K

<sup>67</sup> as reported in D6.3, see also <http://www.nsc.liu.se/joint-sec-training/>.

## 5.6. Risk Management

Operational risk management has been established and reacted on newly discovered security vulnerabilities. A risk assessment and a continuity plan is available particularly for the B2SHARE service.

## 6. QUALITY ASSURANCE

The aim of this activity was to support partners to get the DSA granted by the end of the project. This goal was not achievable yet because the “business as usual” is transforming slowly at the generic data centers. After the conclusions drawn from the first assessment campaign in 2013, a second assessment has been undertaken end of 2014 to which mainly the generic data centers responded.

The result<sup>68</sup> indicates the increasing awareness and some progress in the uptake of DSA criteria, particularly at generic data centers which, in the meanwhile, are getting prepared to become repository providers with all the departmental responsibilities that this implies. Most partners provide a suitable level of service concerning data transfer, storage provisioning and access, which corresponds to the commitment for bit stream preservation which is typical for the generic data centers. Another positive notice is the raise of awareness for legal aspects, particularly the need to proof evidence for service portfolio management, the availability of a service catalogues, and change management procedures necessary to specify contractual agreements such as SLAs. On the other hand, the DSA criteria which are related to long-term preservation are often not well met, which is not surprising currently. In many cases data formats are not recommended by the repositories. This can be intended as, e.g., the generic B2SHARE service implies. However, as far as the provided information allows, the risk management for data curation needs to be improved at least when departments of generic data centres become responsible for the repository management. As well as the transparency of the documentation is an issue: documentation that is expected to be public according to the DSA is often claimed to exist while it is not publicly accessible.

### 6.1 Quality Assurance Plan

The purpose of the Quality Assurance Plan (see Annex E, pp.50) is to specify, implement and monitor quality metrics, standards, best practices, resources, specifications, related roles, compliance requirements and the sequence of activities with particular regard to the EUDAT services. This plan has been developed and refined in the course of the EUDAT project and according to the improved understanding of EUDATs core services, the B2-service. It refers to the FitSM library of service quality aspects, which is also discussed in the deliverable D2.3.2.

The latest version of the QA plan has been released in March 2015. It has taken also into account the lessons learned in WP6 and the needs for a future EUDAT2 project. This plan is given in the Annex E.

### 6.2 Service Level Agreement

The purpose of the Service Level Agreement (SLA) is to agree on terms to enable a service provider to commit a set of resources, means and methods ensuring the preservation and access for digital scientific data. By default, the SLA assumes a commitment for data preservation which is does not exceed the lifetime of the EUDAT project and which includes bit stream preservation with a metadata service, integrity check and delivery of a Persistent Identifier (PID).

Annex F (pp.60) presents the SLA template which has been defined for B2SHARE and B2SAFE services.

---

<sup>68</sup> The survey is not representative due to the small number of participating sites and the low response rate from repositories.

## ANNEX A. GLOSSARY

AAA	Authentication, Authorization and Accounting
AAI	Authentication and Authorization Infrastructure
a.i.	ad interim
API	Application programming interface
AtP	Attribute Provider: assigned specific roles and properties to an identity
AuP	Acceptable Use Policy
AuthN	Authentication – the process of identifying an entity by means of a credential that the entity is exposing via a secure mechanism
AuthZ	Authorization – the process of granting access to a service
Backend Storage	The Backend Storage (or Local Long Term Archive, i.e., Local LTA) includes storage technologies (e.g. dCache) and hardware (e.g. tapes, disks) but also file systems (e.g. GPFS). Closely related is also the network interconnectivity to these elements that should not be forgotten.
B2FIND	Brand of the EUDAT Joint Metadata Service
B2SAFE	Brand of the EUDAT Safe Replication Service
B2SHARE	Brand of the EUDAT Simple Store Service
B2STAGE	Brand of the EUDAT Data Staging Service
CDI	EUDAT Collaborative Data Infrastructure
CERT	Computer Emergency Response Team
CI type	Configuration Item type, denotes a class of configurable entities
CLARIN	Common Language Resources and technology Initiative. An ESFRI project in the Social Sciences and Humanities domain.
CMS	Content Management System

Community	Research or Scientific Community: a diverse network of interacting scientists and/or researchers. It can include "sub-communities" working on particular scientific fields or research areas, and within particular institutions. A community can have many representatives which are – ideally – affiliated with one or several institutions (legal entities) which are legally responsible for particular data collections. These data collections are typically managed via repositories.
CROWD	Centralized Identity Management solution from Atlassian
Curation	Provision of domain-dependent contextual support for permanent access to the meaning of data – including metadata, lexica, etc
Curation and Preservation	The process of ensuring that data can be re-used over time.
CSIRT	Computer Security Incident Response Team
Data Project	<p>A project owned by a community or a society that concerns a defined set of data collections (the definition characterizes the data domain; it does not constraint the extension of the data or the data domain, but it specifies the character or the data, particularly the original purpose of the data acquisition and/or collection and how it is organized, in the best case the ontology or the data model). Thus, in the scope of a data project are typically <i>Data</i> and <i>Users</i> – and <i>Users</i> in terms of <i>data depositors</i> and <i>data consumers</i>. The data project can be owned by those who want to deposit data as well as those who want to consume data.</p> <p>The owner of the data project must be represented by a <i>responsive</i> project principal who can explain the requirements for the project and who has the authority to organize and take decisions on priorities and commitments regarding the data project. A data project comprises a requirements specification that describes how the data has to be handled, stored (where, when, how, how long), registered (create identifiable entities with PIDs assigned), replicated, transferred, indexed, findable, accessible and auditable. It must also be clear which capabilities and capacities regarding services, resources and the predicted data access pattern are necessary to implement and operate the data project over a certain span of time (e.g. data growth with regard to data volume, granularity and access), and how the capacities are funded.</p> <p>A project has a start date and an end date. The start date is the date when a project has been adequately defined; this can be a project proposal that has been registered. The end date can be a date in the distant future (in case of a long-term project). The realization (implementation) of a data project in the context of EUDAT can either be done by just using the B2-Services as needed, or the project needs to be enabled, which actually means that the project-</p>

	relevant B2-Services are to be enabled. Many projects were requiring just one B2-Service. In this case, Service and Project enabling is the same.
DoW	Description of work – usually the EUDAT description of work is meant.
DSA	Data Seal of Approval – <a href="http://www.datasealofapproval.org">http://www.datasealofapproval.org</a>
DSI	GridFTP Data Storage Interface, a library developed by CINECA
EB	EUDAT Executive Board
EGEE	Enabling Grids for E-sciencE
EGI	European Grid Initiative
e-IRG	e-Infrastructure Reflection Group
EISCAT_3D	The next generation European incoherent scatter radar system. An ESFRI project in the Environmental Sciences domain.
ENES	European Network for Earth System Modelling
EPIC	European Persistent Identifier Consortium <a href="http://www.pidconsortium.eu">http://www.pidconsortium.eu</a>
EPOS	European Plate Observing System. An ESFRI project in the Environmental Sciences domain.
FP7	Seventh Framework Program
GC	EUDAT General Council
GÉANT	European multi-gigabit computer network for research and education purposes.
high-level infrastructure service	A service which is based on a service oriented architecture across different administrative domains. i.e. several services are required to provide the high-level infrastructure service. For instance, the Safe Replication service requires several iRODS-Services operated at different sites and a PID service which requires a distributed service infrastructure in the background (mirrors).
HPC	High Performance Computing
IaaS	Infrastructure as a Service
IdP	Identity Provider (can be a generalized IdP, not necessarily a Shibboleth IdP)
iRODS	Integrated Rule-Oriented Data System <a href="http://www.irods.org">http://www.irods.org</a>

ISO	International Organization for Standardization
LifeWatch	E-Science and Technology Infrastructure for Biodiversity Data and Observatories. An ESFRI project in the Environmental Sciences domain.
Local Long Term Archive (Local LTA)	See Backend Storage.
MoU	Memorandum of Understanding
NREN	National Research and Education Network
OAIS	Open Archival Information System
OASIS	Organization for the Advancement of Structured Information Standards
OSCT	Operational Security Coordination Team
PaaS	Platform as a Service
PID	Persistent Identifier
PM	EUDAT Project Manager
PMO	EUDAT Project Management Office
PRACE	Partnership for Advanced Computing in Europe
Pre-Production	A service provisioning state “pre-production” is also called “beta”. A pre-production service can often be safely used, particularly with regard to data loss and data corruption. But no statement can be made regarding the quality of service, its reliability, stability, accessibility and (long-term) availability.
Preservation	Provision of generic support for permanent access to ‘physical’ data – the bits and bytes – including storage, replication, provenance, etc
QA	Quality assurance
QoS	Quality of Service

Repository	<p>An administrated directory for storing, describing, managing identifiable registered digital objects. Digital objects have an extension (have a size and can be enumerated) and they can contain any kind of data of any kind of binary representation: binary data of any format, software in binary form or as source code, virtual machine images, configuration files, archival information packages, digital audio-visual stream chunks, digital text corpora.</p> <p>The level of trust (the trustworthiness) may depend on the way how repositories are managed, the quality and completeness of the description (incl. provenance) of the data objects, the level of service that is guaranteed, the way how authenticity and provenance is ensured, and finally the auditable measures taken to ensure the reliability, security, availability, stability and sustainability of the repository.</p>
Resource Coordination Tool (RCT)	<p>An operational tool to coordinate the provisioning of requested resources; sometimes also called: Resources Provisioning Coordination Tool, in order to emphasize the fact that the tool is intended to support the coordination or the provisioning of resources, not necessarily the resources themselves. This let each resource provider the freedom to decide on the adequate underlying technology that optimally fits to the specific strategic plans of a site. In other words: the tool is used to ask for bids on specific resources and potential provider sites can decide whether and on which time scale they are able to provide the demanded resources, and what does it will cost.</p> <p>See Resource also Provisioning Framework.</p>
Resource Provisioning	<p>Making available resources in terms of system platforms which consist of virtual and/or physical compute, storage and network components including firmware and a basic execution environment (i.e. operation system). It is not relevant whether resources are provided as physical or virtual instances, but these instances are located at a site.</p>
Resource Provisioning Framework	<p>Administrative solution to handle the Resource Provisioning management. The solution is until now based on a database with a web front-end where service providers can enter resource requests, resource queries can be done, and site administrators can manage the information about their provided resources. This platform is also referred as Resource Coordination Tool or RCT.</p>
RI	Research Infrastructure
RP, Resource Provider	An organizational instance that provides maintains and operates compute, storage and network resources at a site.
RP coordinator	Resource Provisioning Coordinator: an employee (staff member) at a resource provider site who is coordinating the timely integration and provisioning of resources (see also SHF).

RT	Request Tracker, commonly abbreviated to RT, is a ticket-tracking system written in Perl used to coordinate tasks and manage requests among a community of users.
SCI	A Trust Framework for <i>Security Collaboration among Infrastructures</i>
Service Provisioning	Make services available which are deployed on top of resources. Services can be deployed and operate on top of resources which are provided at another site.
Site	Administrative Domain, a legal entity, able to sign legal contracts such as Service Level Agreements. A site can provides different kind of resources and services.
SHF	Service Hosting Framework: collaborative framework for the coordinated provisioning of processing, storage and network resources for hosting (data-oriented) services. This framework allows local or remote service providers to deploy and to operate their services.
SLA	Service Level Agreement
SP, Service Provider	<p>a person, team or organizational instance (site) in charge of deploying, maintaining and operating (providing) a service on top of resources provided by a resource provider (RP). A service provider does not necessarily provide the resources which are used. However, after a deployment and pre-production phase, services can be handed over to service operators at the Resource Provider Site which are providing the resources. In this case, Service and Resource Provider are located at the same site (administrative domain).</p> <p>A Service Provider that must comply with an SLA, has to be part of an organization that is a legal entity (site, administrative domain).</p>
SP coordinator	Service Provisioning Coordinator: a person who is in charge of coordinating the timely provisioning of requested services.
SSO	Single Sign On, i.e. access many services or resources after a single login
SVN	Subversion – a version control system <a href="http://subversion.apache.org">http://subversion.apache.org</a>
TF	Taskforce
TTS	Trouble Ticket System
VO	Virtual Organisation

## ANNEX B. DATA PROJECTS

Table 2: Data Projects indicating Start Time (“pre sales”), Call for Collaboration Proposal and Status

	Project	CfC #	Status	B2-Service Contact	Community	Discipline/Topic
Dec 12	B2STAGE+B2SAFE & VPH (1)		enabling	E. Erastova, RZG	VPH	Lifescience, Bio Medicine
Dec 12	B2SAFE & EPOS PP WG7		prod	C. Cacciari, CINECA	EPOS	Earth System Science
Dec 12	B2SAFE & CMIP5		prod	S. Apweiler, Juelich	ENES	Climatology
Dec 12	B2SAFE & INCF		prod	J. Heikkinen, CSC	INCF	Neuroscience
Dec 12	B2SAFE & TLA		prod	E. Erastova, RZG	CLARIN	Linguistics, Humanities
Feb 13	B2SAFE & ORFEUS@KNMI		pre-prod	R. Verkerk, SURFsara	EPOS	Earth System Science
Apr 13	B2SAFE & EKUT		prod	E. Erastova, RZG	CLARIN	Linguistics, Humanities
Jul 13	B2SAFE & LINDAT		prod	P. Dugenie, CINES	CLARIN	Linguistics, Humanities
Oct 13	B2SHARE generic instance		prod	P. Hyvarinen, CSC	EUDAT generic	<i>multiple science domains</i>
Jan 14	B2SHARE & CLARIN		pre-prod	P. Hyvarinen, CSC	CLARIN	Linguistics, Humanities
Jan 14	B2FIND & ENES		prod	H. Widmann, DKRZ	ENES	Climatology
Jan 14	B2FIND & generic B2SHARE		prod	H. Widmann, DKRZ	EUDAT B2SHARE	any research data
Jan 14	B2FIND & PanData		prod	H. Widmann, DKRZ	PanData	Nuclear Physics
Jan 14	B2SAFE & MfN Berlin		prod	E. Erastova, RZG	Museum f. Naturkunde Berlin	Biodiversity, Museum
Feb 14	B2FIND & CLARIN		prod	H. Widmann, DKRZ	CLARIN	Linguistics
Mar 14	B2SHARE & DRIHM		pre-prod	P. Hyvarinen, CSC	DRIHM	Earth System Science
Mar 14	B2SHARE & BBMRI		pre-prod	P. Hyvarinen, CSC	BBMRI, ELIXIR NL	Biobanking, Medical Sc.
Mar 14	B2SAFE & diXa	#390	prod	J. Heikkinen, CSC	diXa	Lifescience, Toxicology
Mar 14	B2SHARE & EUON		prod	P. Hyvarinen, CSC	EUON	Ontology
Mar 14	B2SAFE & iMarine	#415	planned	G. Morelli, CINECA	iMarine	Biodiversity
Apr 14	B2FIND & EISCAT	#411	planned	H. Widmann, DKRZ	EISCAT	Earth System Sciences
Apr 14	B2FIND & ICOS	#410	planned	H. Widmann, DKRZ	ICOS	Earth System Sciences
Apr 14	B2SAFE & IST	#407	planned	S. Coutin, CINES	University IST	University
Apr 14	B2SAFE & IPSL	#402	pre-prod	S. Coutin, CINES	ENES	Climatology, Meteorology
Apr 14	B2STAGE+B2SAFE & VPH (2)		pre-prod	E. Erastova, RZG	VPH	Lifescience, Bio Medicine
Apr 14	B2FIND & SDL		prod	H. Widmann, DKRZ	SDL	Agriculture
Apr 14	B2FIND & TheEuropeanLibrary	#397	prod	H. Widmann, DKRZ	European Library	Cultural Heritage
Jun 14	B2SAFE & ELIXIR	#422	enabling	J. Heikkinen, CSC	ELIXIR	Lifescience, Biobanking
Jun 14	B2FIND & DataCite		prod	H. Widmann, DKRZ	DataCite	<i>multiple science domains</i>
Jun 14	B2FIND & NARCIS/DANS		prod	H. Widmann, DKRZ	DANS	Social Sciences
Jun 14	B2FIND & GBIF		prod	H. Widmann, DKRZ	GBIF	Biodiversity
Jul 14	B2STAGE+B2FIND & ALEPH	#406	enabling	H. Widmann, DKRZ G. Morelli, CINECA	ALEPH/INFN/ DPHEP	High Energy Physics
Aug 14	B2SHARE & RDA		pre-prod	P. Hyvarinen, CSC	RDA	RDA documents
Sep 14	B2FIND & BBMRI	#421	enabling	H. Widmann, DKRZ	BBMRI	Biobanking, Medical Sc.
Sep 14	B2SAFE & Lifewatch VZL	#395	planned	D. Vicente, BSC	LIFEWATCH VZL	Biodiversity
Oct 14	B2FIND & WDC-Terra/EPOS	#412	enabling	H. Widmann, DKRZ	WDC-Terra, EPOS	Earth System Sciences
Oct 14	B2SHARE & NRM	#420	pre-prod	P. Hyvarinen, CSC	Swedish Museum of Natural History	Biodiversity, Museum
Oct 14	B2SHARE & GBIF	#420	pre-prod	P. Hyvarinen, CSC	Finnish Museum of Natural History	Biodiversity, Museum
Dec 14	B2SAFE & ABC	#394	enabling	J. Heikkinen, CSC	Ascona B.DNA Consortium	Lifescience, Molecular Dynamics
Jan 15	B2FIND & IVOA		enabling	H. Widmann, DKRZ	IVOA	Astronomy
Jan 15	B2FIND & CESSDA		enabling	H. Widmann, DKRZ	CESSDA	Social Sciences
Jan 15	B2FIND & agINFRA		planned	H. Widmann, DKRZ	agINFRA	Agricultural, Environmental

The start dates indicate when the collaboration on concrete B2-Services started. Colours for "Project" are used to distinguish between the different B2-Services involved. The column "CfC" contains the proposal numbers from the first Call for Collaboration. The "Enabling" can have the focus on "community infrastructure service enabling" and "CDI service enabling".

## ANNEX C. CONFIGURATION MANAGEMENT USING THE CENTRAL REGISTRY

The following section illustrates the usage of the GOCDDB instance for configuration management by showing exemplary entries in the various sections of the registry.

The purpose of the central site registry in EUDAT is to provide information about the configuration of the Collaborative Data Infrastructure, its organisational and technical components and how they are grouped.

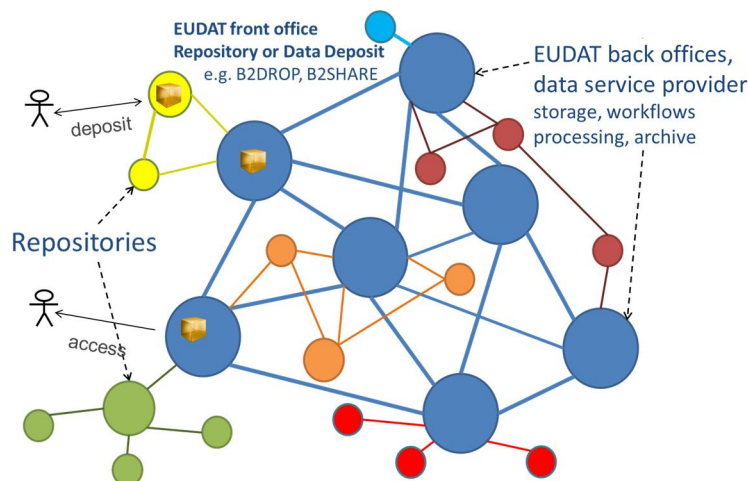


Figure 13: The CDI as Network of Generic and Thematic Centres (according to WP2)

The (static) information is used for different scenarios such as monitoring, notification services, data policy management services, other information services, for building contact lists of sites, site security contacts, site service endpoint contacts, service group contacts and additional information such as the certification status of sites and services. The information model is extensible by adding further properties to the service endpoints and/or groups of service endpoints. Extension properties are e.g. used to specify for each service endpoint the target software version or additional information needed by remote components (see Figure 17).

### Requirements (overview<sup>69</sup>)

- (1) Provide detailed configuration information about sites, service components, service groups with the flexibility to extend the list of attributes (e.g. further certificates)
- (2) Extensible information model
- (3) Security and safety: Role management with a fine grained access and authorization
- (4) Ability to define separate administrative domains
- (5) Reliable and highly available
- (6) Browsable web interface
- (7) API for queries – preferably REST style
- (8) Ability to schedule services for downtimes
- (9) Production-ready sustainable solution, strong user base
- (10) The tool should comply with SQM requirements such as Configuration Management (FITSM PR11, p.54)

<sup>69</sup> See D6.1, pp.35, for more details.

- (11) Registry service must be trustworthy (auditable): the authenticity of information and its sources must be ensured.
- (12) PIDs for service components: stable, unique IDs are needed e.g. for accounting records, service stable (persistent) configuration descriptions.
- (13) Additional standard attributes required: certification (trust) levels, Licenses, applicable SLA/OLAs
- (14) Some critical database entries, e.g. quality indicators, require genuine signatures and authentic binding to the authoritative source of information (e.g. via PKI mechanisms)
- (15) Some dynamic information should be made available (e.g. for workflows): quotas, available resources
- (16) Modification of entries should be possible via REST API (e.g. scheduled downtimes, rather dynamic information). Sites, which are registered in more than one registry instance should be facilitated to keep particularly dynamic information up-to-date.
- (17) History (log) about changes (to be able to track who changed when which entry)

The screenshot shows the EUDAT Sites management interface. The header includes the EUDAT logo and the text 'All Sites in GOCDB'. Below the header is a filter section with dropdown menus for Project (all), Certification (all), Infrastructure (all), and Scope (all). There is also a checkbox for 'Include Closed Sites' and a dropdown for 'Extension Name' (none). The main content area displays a table of 19 sites, each with a 'Name' column, a 'Project' column (all 'EUDAT\_REGISTRY'), an 'Infrastructure' column (all 'Production'), a 'Certification Status' column (all 'Candidate'), and a 'Scope(s)' column (all 'EUDAT').

Name	Project	Infrastructure	Certification Status	Scope(s)
» BSC	EUDAT_REGISTRY	Production	Candidate	EUDAT
» CINECA	EUDAT_REGISTRY	Production	Candidate	EUDAT
» CINES	EUDAT_REGISTRY	Production	Candidate	EUDAT
» CSC	EUDAT_REGISTRY	Production	Candidate	EUDAT
» CUNI	EUDAT_REGISTRY	Production	Candidate	EUDAT
» DKRZ	EUDAT_REGISTRY	Production	Candidate	EUDAT
» EKUT	EUDAT_REGISTRY	Production	Candidate	EUDAT
» EPCC	EUDAT_REGISTRY	Production	Candidate	EUDAT
» INGV	EUDAT_REGISTRY	Production	Candidate	EUDAT
» JUELICH	EUDAT_REGISTRY	Production	Candidate	EUDAT
» KIT	EUDAT_REGISTRY	Production	Candidate	EUDAT
» MPIPL	EUDAT_REGISTRY	Production	Candidate	EUDAT
» PDC	EUDAT_REGISTRY	Production	Candidate	EUDAT
» PSNC	EUDAT_REGISTRY	Production	Candidate	EUDAT
» RZG	EUDAT_REGISTRY	Production	Candidate	EUDAT
» STFC	EUDAT_REGISTRY	Production	Candidate	EUDAT
» SURFsara	EUDAT_REGISTRY	Production	Candidate	EUDAT
» UCL	EUDAT_REGISTRY	Production	Candidate	EUDAT

Figure 14: EUDAT Instance of the GOCDB showing the List of Registered Sites

## Solution

As already described in deliverables D6.1 and D6.2 an early evaluation showed that the EGI GOCDB is a good candidate and the experiences with GOCDB v4 and, in the meanwhile, with GOCDB v5 released by STFC are promising.

### GOCDB facts:

- Developed by EGEE, STFC, EGI.eu, mainly influenced by requirements from the HEP community.

- Provides (static) information about Service Providers (Sites, i.e. administrative domains), Service Endpoints (service components accessible via sockets) and the Services (generally expressed as group of service endpoints which have to be available to run the overall service).
- Provides a comprehensive hierarchical role model that allows any Service (endpoint) Provider and its Service Endpoints to be registered and described.
- Future releases will allow representing configurations of infrastructures with different administrative domains.
- GLUE2<sup>70</sup> is used as basic information model. Although dominated by the concepts of Grid Computing, is also useful for describing relevant properties of sites and the services within the EUDAT CDI. Additional required properties, e.g. QA certificates on the level of services or service endpoints, can be added by property extensions.
- EUDAT was and is able to contribute to the future evolution.
- EGI uses the GOCDB and other (community) infrastructures are interested, use or plan to use this solution as well as to manage the configuration of their own e-Infrastructures. This might facilitate the interoperation of GOCDB instances in the near future.

The screenshot displays the upper part of the EUDAT Site Panel, which is organized into several sections:

- Contact:** A table listing various contact points for the site.
 

E-Mail	jkennedy@rzg.mpg.de
Telephone	+49 89 3299
Emergency Telephone	+49 89 3299
CSIRT Telephone	+49 89 3299
CSIRT E-Mail	eudat-csirt@rzg.mpg.de
Emergency E-Mail	eudat-csirt@rzg.mpg.de
Helpdesk E-Mail	eudat-support@rzg.mpg.de
- Project Data:** A table providing details about the project and infrastructure.
 

Project	EUDAT_REGISTRY
Infrastructure	Production
Certification Status	Candidate <a href="#">Change</a>
Scope(s)	EUDAT
- Networking:** A table listing network-related information.
 

Home URL	http://www.rzg.mpg.de
IP Range	
Domain	rzg.mpg.de
- Location:** A table providing geographical coordinates and location details.
 

Country	Germany
Latitude	48.26726
Longitude	11.6734599
Time Zone	Europe/Berlin
Location	Garching near Munich, Germany
- Site Extension Properties:** A table with columns for Name, Value, Edit, and Remove. It includes a '+ Add Properties' button.
- Services:** A table with columns for Hostname (service type), URL, Production, Monitored, and Scope(s).

Figure 15: Site Panel (upper part) with the Site Description

<sup>70</sup> <https://www.ogf.org/documents/GFD.209.pdf>

Services 				
Hostname (service type)	URL	Production	Monitored	Scope(s)
epic.rzg.mpg.de (handle.epic)	https://epic.rzg.mpg.de/epic_v2/	✓	✓	EUDAT
irods0-eudat.esc.rzg.mpg.de (b2safe.irods)	irods://irods0-eudat.esc.rzg.mpg.de:1247	✓	✓	EUDAT
cmon.eudat.eu (NSCA-client)	https://cmon.eudat.eu	✓	✓	EUDAT
mon-int.esc.rzg.mpg.de (NSCA-client)	https://mon-int.esc.rzg.mpg.de/hagios/	✓	✗	EUDAT
con01.rzg.mpg.de (b2host.ssh-gateway)	ssh://con01.rzg.mpg.de	✓	✓	EUDAT
weblicht-eudat.esc.rzg.mpg.de (WebServer)		✗	✓	EUDAT
creg.eudat.eu (GOCDB)	https://creg.eudat.eu/view_portal	✓	✓	EUDAT
irods3-eudat.esc.rzg.mpg.de (b2safe.dsi)	gsiftp://irods3-eudat.esc.rzg.mpg.de:2811	✓	✓	EUDAT
irods3-eudat.esc.rzg.mpg.de (b2safe.irods)	irods://irods3-eudat.esc.rzg.mpg.de:1247	✓	✓	EUDAT
irods-replix.esc.rzg.mpg.de (b2safe.irods)	irods://irods-replix.esc.rzg.mpg.de:1247	✓	✓	EUDAT

 Add Service

Figure 16: Site Panel (lower part) with Local Service Endpoints (Service Type, Production Level)

The screenshot shows the EUDAT management interface for a service endpoint. The main content area is titled "Service: irods0-eudat.esc.rzg.mpg.de - b2safe.irods". It contains several panels:

- System:** Host name: irods0-eudat.esc.rzg.mpg.de, IP Address: 130.183.206.64, Operating System: Linux, Architecture: amd64, Contact E-Mail: elena.erasova@rzg.mpg.de.
- Grid Information:** Host DN, URL: irods://irods0-eudat.esc.rzg.mpg.de:1247, Parent Site: RZG, Scope(s): EUDAT.
- Project Data:** Production Level (checked), Beta (checked), Monitored (checked).
- Service Groups this Service Belongs To:** B2SAFE, B2SAFE\_CLARIN, B2SAFE\_CLARIN\_EKJT, B2SAFE\_CLARIN\_REPLIX, B2SAFE\_MIN.
- Recent Downtimes:** A table with columns Description, From, and To, and an "Add Downtime" button.
- Service Extension Properties:** A table with columns Name, Value, Edit, and Remove. This table is circled in red.

Name	Value	Edit	Remove
irods_path	/vzRZGE/eudat/		
irods_zone	vzRZGE		
irods_resource	cacheResc		
irods_version	3.3.1		
b2safe_version	2.3		

Figure 17: Service Endpoint Panel with Extended Properties

Name	Description	Scope(s)
» EUDAT_COLLABORATIVE_SERVICES	EUDAT Collaborative Services	EUDAT
» B2ACCESS	All service endpoints which belong to the EUDAT AAI	EUDAT
» B2FIND	EUDAT Joint Metadata Indexing and Search Service	EUDAT
» B2SAFE	All service endpoints which belong or are related to the EUDAT B2SAFE service	EUDAT
» B2SHARE	EUDAT B2SHARE Services	EUDAT
» B2STAGE	Service endpoints related to the B2STAGE service	EUDAT
» B2HANDLE	All the EUDAT PID service endpoints	EUDAT
» B2ACCESS_MYPROXY	EUDAT coordinated core MyProxy services	EUDAT
» B2SAFE_CLARIN	Service endpoints which belong to the EUDAT B2SHARE for all CLARIN repositories	EUDAT
» B2SAFE_CLARIN_LINDAT	Service endpoints which belong to the B2SAFE service for CLARIN Lindat	EUDAT
» B2SAFE_CLARIN_EKUT	Service endpoints which belong to the B2SAFE service for EKUT	EUDAT
» B2SAFE_CLARIN_REPLIX	Service endpoints which belong to the B2SAFE service for CLARIN Replix	EUDAT
» B2SAFE_ENES	Service endpoints which belong to the EUDAT core service	EUDAT
» B2SAFE_EPOS	Service endpoints related to the B2SAFE island for EPOS repositories (INGV, KNMI, ...)	EUDAT
» B2SAFE_INCF	Service endpoints which belong to B2SAFE for the INCF dataspace	EUDAT
» B2SAFE_VPH	Service endpoints which belong to the B2SAFE service for VPH	EUDAT
» EUDAT_HOUSED_COMMUNITY_SERVICES	Community services which are running at EUDAT sites	EUDAT
» EUDAT_OPERATIONAL_MONITORING	EUDAT local and central monitoring services	EUDAT
» EUDAT_OPERATIONAL_SERVICES	Service endpoints which belong to the EUDAT Operational Services	EUDAT
» B2HOST	EUDAT Service Hosting Framework (ssh gateways)	EUDAT
» EXTERNAL_COMMUNITY_SERVICES	Collection of community services that are provided by non-EUDAT sites	EUDAT
» B2SAFE_MfN	B2SAFE for the Berlin Museum of Natural Science	EUDAT
» B2SAFE_VPH_RZG	B2SAFE for VPH project data with RZG	EUDAT
» B2DROP	EUDAT B2DROP Services	EUDAT

Figure 18: Service Groups Panel showing the available Service Groups in EUDAT

**Properties**

Monitored	✓
Scope(s)	EUDAT
Contact E-Mail	Elena.Erastova@rzg.mpg.de

**Services**

Hostname (service type)	Description	Production	Scope(s)
epic.rzg.mpg.de (handle.epic)	PID / EPIC registration service	✓	EUDAT
lux115.mpi.nl (b2safe.irods)	IRODS service at MPI-TLA	✓	EUDAT
irods1.storage.sara.nl (b2safe.irods)	IRODS server at SURFsara	✓	EUDAT
ariane.cines.fr (b2safe.irods)	IRODS Server in production at CINES	✓	EUDAT
darin.uni-tuebingen.de (b2safe.irods)	IRODS endpoint for CLARIN-EKJT project	✓	EUDAT
irods0-eudat.esc.rzg.mpg.de (b2safe.irods)	IRODS v3.3 at RZG	✓	EUDAT
epic3.storage.surfsara.nl (handle.epic)	PID / EPIC registration server	✓	EUDAT
irods-replix.esc.rzg.mpg.de (b2safe.irods)	IRODS server for the replix project	✓	EUDAT

**Roles**

Name	Role	Revoke
Elena Erastova	Service Group Administrator	Revoke
Johannes Reetz	Service Group Administrator	Revoke

**Service Group Extension Properties**

Name	Value	Edit	Remove
+ Add Properties			

**Recent Downtimes (View all Downtimes)**

Description	From	To
Maintenance on the storage infrastructure	11-Feb-15 14:00:00	13-Feb-15 10:00:00

Figure 19: Particular Service Group Panel for CLARIN with the Related Service Endpoints

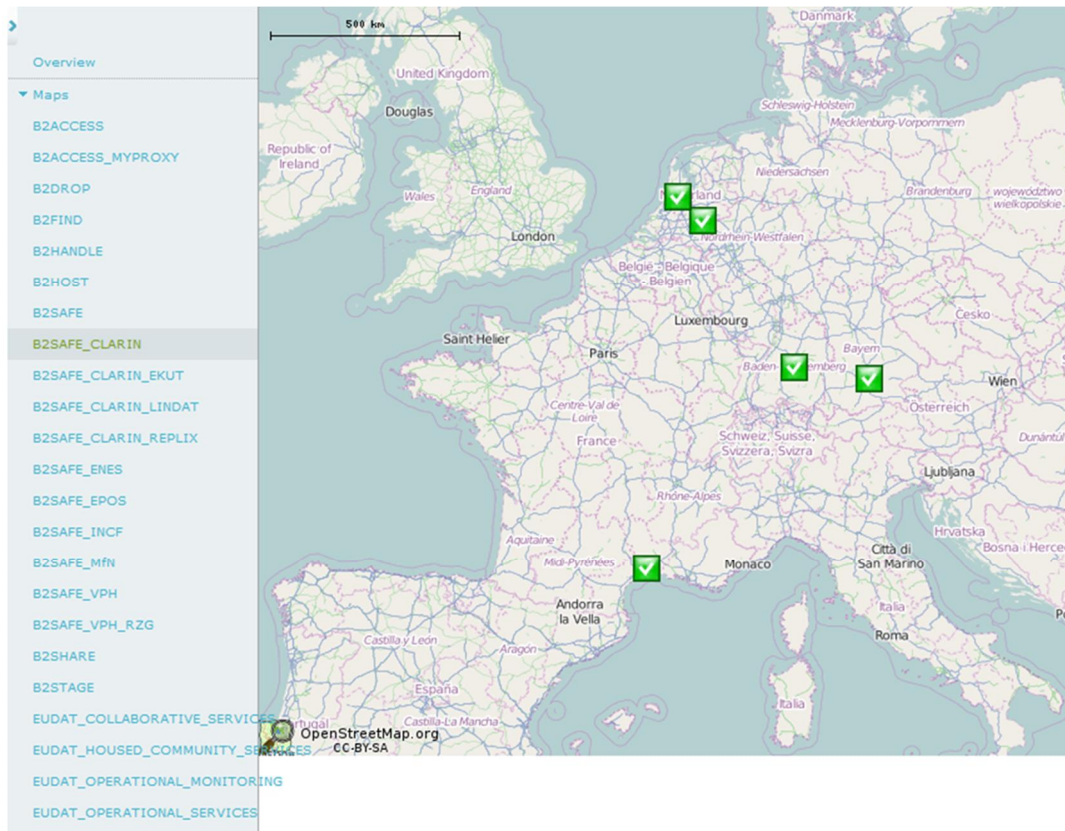


Figure 20: Monitoring of the Service Group (CLARIN) showing the Related Service Endpoints

Table 3: GOCDB RESTful API used in EUDAT<sup>71</sup>

Method	Description	Protected
<u>get_cert_status_changes</u>	changes to certification statuses	Y
<u>get_cert_status_date</u>	certification status for production sites and change date	Y
<u>get_downtime</u>	downtimes for sites and services	N
<u>get_service_endpoint</u>	service endpoints and associated information	N
<u>get_service_types</u>	valid service types and associated description	N
<u>get_service_group</u>	service groups and the related service endpoints	Y
<u>get_service_group_role</u>	service groups and the roles on these groups	Y
<u>get_site</u>	site information including contacts, grouped by site	Y
<u>get_site_contacts</u>	persons (and associated info) having a role at site level	Y
<u>get_site_list</u>	sites with minimal associated information	N
<u>get_site_security_info</u>	Returns security contact information for sites	Y
<u>get_user</u>	Returns a user or a list of users with associated details	Y
<u>get_project_contacts</u>	persons (and associated info) having a role in a project.	Y

`$wget --no-check-certificate -O result.xml 'https://creg.eudat.eu/gocdbpi/public/?method=get_service_types'`

<sup>71</sup> more about the GOCDB API: [https://wiki.egi.eu/wiki/GOCDB/PI/Technical\\_Documentation](https://wiki.egi.eu/wiki/GOCDB/PI/Technical_Documentation)

## ANNEX D. USER DOCUMENTATION TEAM AND DOCUMENTS

The objective of the User Documentation team is to create documentation for the EUDAT users at all levels and publish it on the EUDAT website. The intended audience, the process followed and the sources of information for the User Documentation are described in deliverable D6.3, Annex F. The User Documentation team could rely on the expertise of colleagues from across EUDAT so as to come up with high quality, fit for purpose documentation. The method to split drafts from production also worked well, although the port from Confluence to Drupal has occasionally created formatting problems. It is true that this part of the process could be smoother, but in hindsight the issues were not that difficult to deal with, so EUDAT2020 will need to bear in mind return on investment if it considers changing this part of the methodology.

### Presentation

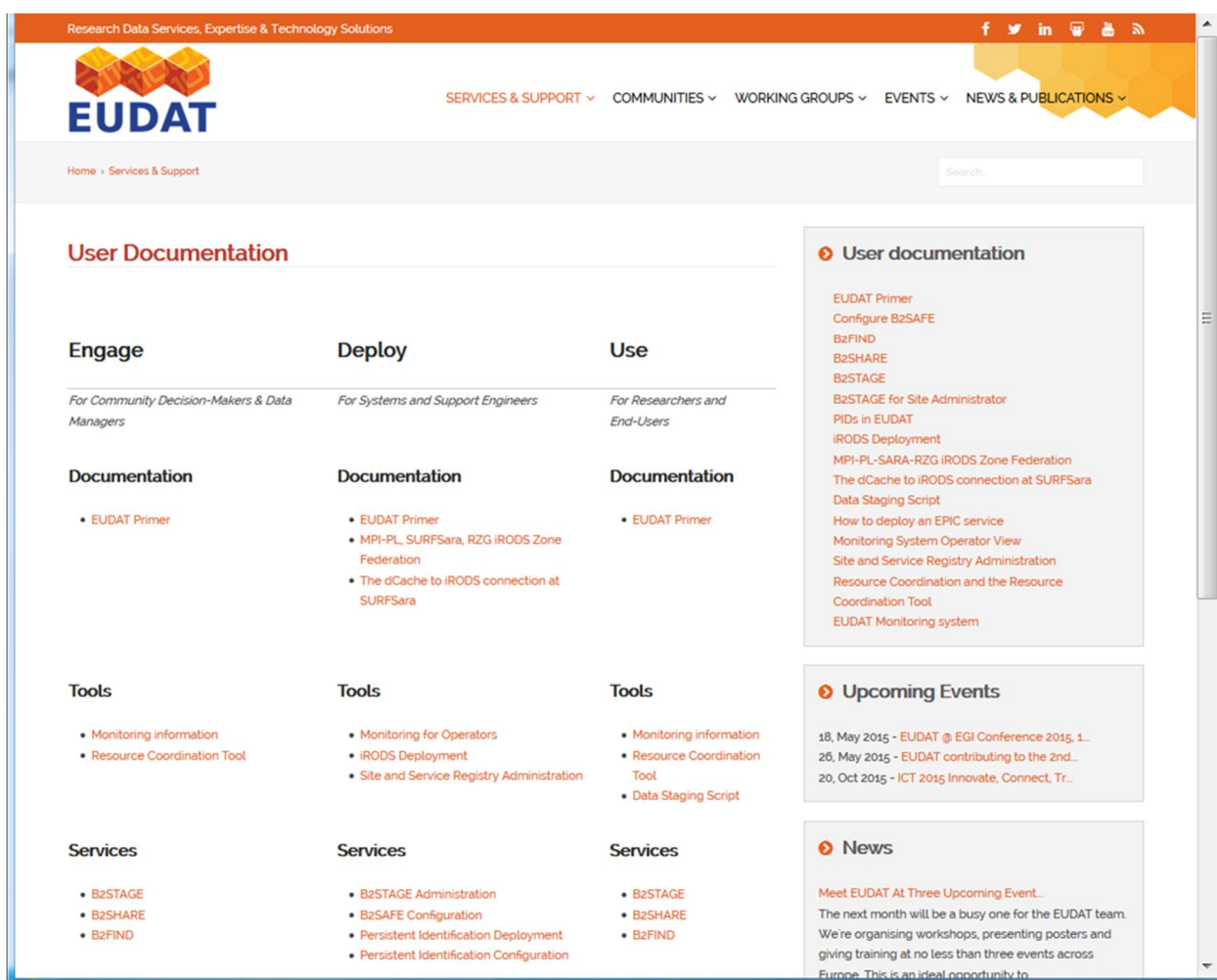


Figure 21: EUDAT User Documentation Entry Point (Mar 2015)

The documentation is available on the EUDAT website from the navigation bar assigned to Services and Support. The documents on the website are grouped according to the three main audiences: community managers, system administrators, and end-users from various disciplines. Roles which want to

- *engage* with EUDAT for specific services
- *deploy* software components in order to provide and operate a service
- *use* EUDAT services.

Figure 21 shows the User Documentation webpage with the documents that have been published.

Table 4 lists the User Documentation which is published on the EUDAT webpage<sup>72</sup> as on 28 February 2015. 16 documents are published, many of which have been updated over the course of the last two years. A small number of editors have been involved, in most cases (but not necessarily) taking care of a set of services in line with their expertise. The process to employ a member of the team that was not involved in the authoring as a reviewer was followed in all cases, and it was deemed very fruitful.

Table 4: User Documentation published on the EUDAT website <http://eudat.eu/userdoc>

Title	Version	Date of Publication	Editor	Approved by	Audience
EUDAT Primer	1.6	Feb-15	Kostas Kavoussanakis, Rob Baxter, Cristina Manzano, Emanuel Dima, Carl Johan Håkansson		EC, community data managers, some end users, site data managers
PIDs in EUDAT and PID services	1.0	Feb-15	Robert Verkerk, Kostas Kavoussanakis	Carl Johan Håkansson	Community managers, site admins
User Documentation - B2FIND	2.0	Feb-15	Hannes Thiemann, Heinrich Widmann	Massimo Fares	End users, community managers
User Documentation - B2SHARE	1.2	Feb-15	Carl Johan Håkansson	Massimo Fares	End users of B2SHARE, EC, interested public
User Documentation - Site and Service Registry Administration	1.0	Oct-14	Massimo Fares	Carl Johan Håkansson	Site admins
User Documentation - Monitoring System Operator View	1.0	Oct-14	Roger Downing	Massimo Fares	EUDAT operators and site admins
User Documentation - How to deploy an EPIC service	1.0	Sep-14	Robert Verkerk	Kostas Kavoussanakis	Site data managers
User Documentation - Data Staging Script	1.2	Sep-14	Mariani Giacomo, Giovanni Morelli	Carl Johan Håkansson	End-users

<sup>72</sup> <http://www.eudat.eu/userdoc>

Title	Version	Date of Publication	Editor	Approved by	Audience
User Documentation - iRODS Deployment	1.7	Sep-14	Roger Downing, Kostas Kavoussanakis, Jdrzej Rybicki, Benedikt von St. Vieth, Willem Elbers (tbc), ...	Kostas Kavoussanakis	Community data managers, site data managers, the interested public
User Documentation - B2STAGE for Site Administrator	1.1	Jul-14	Mariani Giacomo, Giovanni Morelli	Carl Johan Håkansson	Site admins
User Documentation - Configure B2SAFE	2.2	Jul-14	Elena Erastova, Roger Downing	Kostas Kavoussanakis	Site Admin, maybe Community Data Manager
User Documentation - B2STAGE	1.0	Oct-14	Mariani Giacomo, Giovanni Morelli	Carl Johan Håkansson	End users
Resource Coordination Tool (RCT)	1.1	Oct-14	Cristina Manzano, Florian Kaiser	Emanuel Dima	Community Data managers, site admins, interested stakeholders
User Documentation - MPI-PL-SARA-RZG iRODS Zone Federation	1.5	Dec-13	Kostas Kavoussanakis, Elena Erastova, Robert Verkerk, Willem Elbers	Elena Erastova, Willem Elbers	Community Data Managers, Site Admins
User Documentation - The dCache to iRODS connection at SURFSara	0.3	Dec-13	Kostas Kavoussanakis, Robert Verkerk	Robert Verkerk	Community Data Managers, Site Admins
User Documentation - The EUDAT Monitoring system	1.0	Apr-14	Roger Downing	Carl Johan Håkansson	End-users

## ANNEX E. SERVICE QUALITY ASSURANCE PLAN

# SQA Plan

## Quality Assurance Plan for the EUDAT services

Author(s)	Pascal Dugénie (CINES), Ari Lukkarinen (CSC)
Contributor(s)	Urpo Kaila, (CSC), Johannes Reetz (RZG)
Status	Draft
Version	v.0.9
Date	15 March 2015
References	IEEE Standard 730, Software Quality Assurance Plan (Rev. of IEEE Std 730-1998), IEEE, 345 E. 47 <sup>th</sup> Street , NY 10017 IEEE Standard 730.1 , Guide for Software Quality Assurance Planning , IEEE, 345 E. 47 <sup>th</sup> Street , NY 10017 IEEE Standard 610.12, Standard Glossary of Software Engineering, IEEE, 345 E. 47 <sup>th</sup> Street , NY 10017

## Scope and Purpose

This document is a SQA (Service Quality Assurance) plan which purpose is to specify, implement and monitor quality metrics, standards, best practices, resources, specifications, related roles, compliance requirements and the sequence of activities relevant to EUDAT services.

The SQA plan assumes that EUDAT parties and stakeholders explicitly agree on common efforts to qualify services and data related to EUDAT. Part of the scope of this SQA plan is also to determine the SQA activities to be performed, and who is performing the activity and to identify the results to be achieved throughout the life cycle of the EUDAT project. This plan is a living document as it will continuously evolve.

## Goals and Functions

The SQA plan consists of three main goals:

1. the SQA plan must identify the quality criteria, metrics, standards, roles and practices for evaluating the quality;
2. the SQA plan must ensure that EUDAT services and related documents have an acceptable level of quality prior to release according to the criteria identified above;
3. the SQA plan must provide methods to monitor the quality in order to detect risks of failure at an early stage and apply corrections.

In order to achieve these goals, the SQA plan has several functions:

- to assist EUDAT project teams in tailoring the process to fit the needs;
- to review and audit services to ensure that they conform to specified requirements;
- to ensure the visibility to management on software process compliance.

## Service Management Processes

The SQA plan is based on a subset of the FitSM requirements. FitSM is an open lightweight standards family aimed at facilitating service management in IT service provisioning, including federated scenarios (see <http://www.fedsm.eu>).

High level actions to be done for each process requirement are described below. Detailed requirements have been taken out.

The objective for EUDAT 2020 is to increase these requirements as the services become more mature.

Action plans are also considered, such as:

- FIT-SM Capability and Maturity assessment at the beginning of the project. Based on the results consider doing it again later;
- follow-up DSA assessment cycles;
- the possibility to use ISO 16363 (Space data and information transfer systems -- Audit and certification of trustworthy digital repositories) as additional framework.

## Service Portfolio Management (PR1)

- PR1.1 A service portfolio shall be maintained. All services shall be specified as part of the service portfolio.
- PR1.2 Design and transition of new or changed services shall be planned.
- PR1.3 Plans for the design and transition of new or changed services shall consider timescales, responsibilities, new or changed technology and communication.
- PR1.4 The organizational structure supporting the delivery of services shall be identified, including a potential federation structure as well as contact points for all parties involved.

## Service Level Management (PR2)

- PR2.1 For all services delivered to customers, SLAs shall be in place.
- PR2.2 A service catalogue shall be maintained.
- PR2.3 SLAs shall be reviewed at planned intervals.
- PR2.4 Service performance shall be monitored against service targets defined in SLAs.
- PR2.5 For supporting services or service components provided by federation members or groups belonging to the same organization as the service provider, OLAs shall be agreed.
- PR2.6 OLAs shall be reviewed at planned intervals.
- PR2.7 Performance of service components shall be monitored against operational targets defined in OLAs.

## Service Reporting (PR3)

- PR3.1 Service reports shall be specified and agreed with their recipients.
- PR3.2 The specification of each service report shall include its identity, purpose, audience, frequency and content.
- PR3.3 Service reports shall be produced. Service reporting shall include performance against service targets, information about significant events, workload characteristics and detected nonconformities.

## Service Availability & Continuity Management (PR4)

- PR4.1 Service availability and continuity requirements shall be identified taking into consideration SLAs.
- PR4.2 Service availability and continuity requirements shall include end-to-end availability of services.
- PR4.3 Service availability and continuity plans shall be created. FitSM--1:2013 – Requirements Version 1.2 7

## Capacity Management (PR5)

- PR5.1 Service capacity and performance requirements shall be identified taking into consideration SLAs.
- PR5.2 Capacity plans shall be created.
- PR5.3 Capacity planning shall consider human, technical and financial resources.

## Information Security Management (PR6)

- PR6.1 Information security policies shall be maintained.
- PR6.2 Physical, technical and organizational information security controls shall be implemented, based on identified information security risks.
- PR6.3 Information security controls shall be reviewed at planned intervals.
- PR6.4 Information security incidents shall be given an appropriate priority and managed accordingly.

## Customer Relationship Management (PR7)

- PR7.1 Service customers shall be identified.
- PR7.2 For each customer, there shall be a designated contact responsible for managing the customer relationship and customer satisfaction.
- PR7.3 Communication mechanisms with customers shall be established.
- PR7.4 Service reviews with the customers shall be conducted at planned intervals.
- PR7.5 Service complaints from customers shall be managed.
- PR7.6 Customer satisfaction shall be managed.

## Supplier Relationship Management (PR8)

- PR8.1 Suppliers shall be identified.
- PR8.2 Communication mechanisms with suppliers shall be established.
- PR8.3 (Contractual) disputes with suppliers shall be managed.
- PR8.4 Supplier performance shall be monitored. FitSM--1:2013 – Requirements Version 1.2 8

## Incident & Service Request Management (PR9)

- PR9.1 All incidents shall be recorded according to a defined procedure.
- PR9.2 All incidents shall be classified according to a defined procedure.
- PR9.3 All incidents shall be prioritized according to a defined procedure, taking into account service targets from SLAs.
- PR9.4 Escalation of incidents shall follow a defined procedure.
- PR9.5 Closure of incidents shall follow a defined procedure.
- PR9.6 Personnel involved in the incident and service request management process shall have access to relevant information including configuration and release information.
- PR9.7 Users shall be kept informed of the progress of incidents and service requests they have reported.
- PR9.8 There shall be a definition of a major incident and a documented procedure for classifying and managing them.

## Problem Management (PR10)

- PR10.1 Problems shall be identified and recorded based on analysing trends on incidents.
- PR10.2 If a problem's root cause has been identified, but the problem has not been permanently resolved, problem management shall try to identify actions to eliminate or reduce the impact on the services such as effective workarounds.
- PR10.3 Known errors shall be identified and recorded together with associated workarounds and temporary fixes.
- PR10.4 Problem management shall maintain up-to-date information on known errors and effective workarounds and make them available to the incident and service request management process.

## Configuration Management (PR11)

- PR11.1 CI types and relationship types shall be defined and documented.
- PR11.2 The level of detail of configuration information recorded shall be sufficient to support effective control over CIs.
- PR11.3 Each CI and its relationships with other CIs shall be recorded in a CMDB.
- PR11.4 CIs shall be controlled and changes to CIs tracked in the CMDB.
- PR11.5 The records stored in the CMDB shall be audited at planned intervals.
- PR11.6 Before a new release into a live environment, a configuration baseline of the affected CIs shall be taken. FitSM--1:2013 – Requirements Version 1.2 9

## Change Management (PR12)

- PR12.1 All changes shall be recorded according to a defined procedure.
- PR12.2 All changes shall be classified according to a defined procedure.
- PR12.3 All changes shall be assessed and approved according to a defined procedure.
- PR12.4 There shall be a documented procedure for classifying and managing emergency changes.
- PR12.5 In making decisions on the acceptance of requests for change, the benefits, risks, potential impact to services and customers and technical feasibility shall be taken into consideration.
- PR12.6 A schedule of change shall be maintained. It shall contain details of approved changes, and proposed deployment dates, which shall be communicated to interested parties.
- PR12.7 For changes of high impact or high risk, the steps required to reverse an unsuccessful change or remedy any negative effects shall be planned and tested.

## Release & Deployment Management (PR13)

- PR13.1 A release policy shall be defined.
- PR13.2 The deployment of new or changed services and service components to the live environment shall be planned with all relevant parties including affected customers.
- PR13.3 Releases shall be built and tested prior to being deployed.

- PR13.4 Acceptance criteria for each release shall be agreed with the customers and any other relevant parties. Before deployment the release shall be verified against the agreed acceptance criteria and approved.
- PR13.5 Activities shall be planned and tested for reversing an unsuccessful deployment of a release or remedying any damage caused.
- PR13.6 Releases shall be monitored for success or failure and the results shall be analysed.

## Continual Service Improvement Management (PR14)

- PR14.1 Opportunities for improvement shall be identified and recorded.
- PR14.2 Opportunities for improvement shall be prioritized.
- PR14.3 Opportunities for improvement shall be evaluated and approved.

## Description

### *SQA coverage*

The SQA plan defines a scope of requirements for the Collaborative Data Infrastructure which allow defining compliance tests for each EUDAT service.

### *EUDAT services*

1. B2FIND is a metadata harvesting and cataloging service for research data collections which are stored in EUDAT and external repositories.
2. B2SHARE is a reliable highly available repository service for researchers, scientific communities and citizen scientists, who can upload, describe, register and share their small- and medium scale research data from diverse contexts.
3. B2STAGE is a robust and efficient service for transferring research data sets between EUDAT long-term storage resources and workspaces of HPC or HTC facilities for computations, data analysis or processing.
4. B2SAFE is a robust and highly available distributed service which allows community and departmental repositories to implement and apply data management policies on large research data sets which are safely distributed across multiple administrative domains, ensuring the long-term availability, persistence, authenticity and integrity of the ingested data. One prominent class of use cases is the safe replication of data across multiple geographic, administrative and technology domains.
5. B2DROP is a data exchange workspace service for researchers and scientists in Europe which allows them to keep their research data synchronized with local resources, to exchange it with other researchers, and which provides the possibility to select, transfer and deposit data collections into B2SHARE or B2SAFE.

## Service Requirements Scope

Requirements	B2DROP	B2FIND	B2SAFE	B2SHARE	B2STAGE
discover descriptive information about the data stored through repository services		X			
harvest metadata from a variety of repositories to build a joint catalogue		X			
use standardized facets to search for data		X			
assign Persistent Identifiers to ensure long-lasting reference and access based on auditable rules			X	X	
deposit and release data via standard interfaces and protocols	X		X	X	X
reliable trustworthy environment	X	X	X	X	X

## SQA within the EUDAT project

The SQA activity was performed by Task 6.4 as a part of WP6 (Operations Work Package). Task 6.4.2, in particular, was in charge for planning and assessing the SQA during the whole EUDAT project life time. SQA criteria for the EUDAT services were identified in accordance with the DSA (Data Seal of Approval). Several iterations of SQA assessment of these criteria and a gap analysis have been conducted in order to detect compliance issues at an early stage in order to apply corrections.

## SQA Documentation

The SQA plan refers to a set of documents, each covers different purposes: policy documents, SQA plan and SQA assessments.

- SQA plan description document
- Documents for the SQA assessments: feedbacks from site deputies and reviews to address gaps
- Service Level Agreement (SLA) templates and appendices
- Terms of Use (ToU) document

## SQA Assessments Audits and Reviews

SQA assessments, audits and reviews are necessary to measure the compliance to standards, the reactivity for solving problems, and the quality of the procedures. The SQA plan must ensure that these processes are performed regularly and follow a rigorous canvas. These audits can be organized internally.

At least two self-assessments are needed during the project in order to measure the gap between the two rounds. The assessment questionnaire is submitted to the service providers. The collected answers are analyzed and results are provided via separate documents.

On a shorter time scale, more frequent reviews of the ticketing service used within the project such as JIRA and the TTS allows anticipating the risk of service failure. Ticketing reviews allows also prioritizing requirements such as bug correction or service improvements.

## SQA Tools, Techniques and Methods

There are several SQA tools, techniques and methods that can be applied for assessing and monitoring SQA processes. Some of them have already been introduced throughout the EUDAT project. The following table summarizes the purpose, intended use and limitations within the EUDAT project.

Tools	Purpose	Intended use	Limitations
Deviation database	aims to collect information and produce synthesis reports that are used to suggest to perform some corrective actions.	To be used as a decision aid during Operation Coordination and Service Enabling Team meetings.	This database must be kept constantly up to date to provide accurate and useful information.
Debugging aids	allow to track the origin of a service failure and assign corrective actions to the most suitable persons.	To be used to manage projects by updating ticket status.	none
Documentation aids	aim to provide a common editing place repository to deliver technical and managerial information.	A documentation aid can take place as a collaborative environment based on an enhanced wiki service.	There is no proper standardized versioning mechanism; therefore this aspect must be managed internally.
Checklists	help to plan, identify dependencies and assign priorities to actions.	The ticketing service can be use as a checklist during review meetings (actions done, actions to be done).	Actions are seen as independent. There is no mechanism to identify quickly chains of dependencies.
Performance monitors	This tool allows to predict or detect weakness in the CDI in order to make improvements.	The CDI can be monitored by a central monitoring system.	There is no way to detect failure of a central monitoring system.
Software traceability	This tool allows to evaluate the quality of the software and software-related components, including code analyzer and software inspection, to evaluate performance and robustness as well as testing modules.	Before any release, the software are evaluated by several partners.	There must be a place to centralize test results and a technical committee to validate the level of quality obtained before any release in production.
Information tools	This tool aims to manage lists of diffusion in order to deliver information to the responsible people.	Mailing list are used to exchange information as well as to send notifications	
Standards used	Using established standards, in particular for data access and transfer, metadata management and persistent identification of data is a prerequisite for long term	EPIC Handle, transfer protocols iRODS and GridFTP. HTTP protocol is use for access.	These standards are in constant evolution. Therefore it must be necessary to survey these standards.

	preservation.		
--	---------------	--	--

## Media Management

This section describes labeled and inventoried media filed in a storage area in accordance with security requirements and in a controlled environment that prevents degradation or damage to the media.

Media management	Purpose	Target
Media description	The description of each media ensures better traceability and prevents risks of failure.	
Media documentation	The media documentation provided by the supplier of written later indicates all specifications or operations that can be done on the media.	Storing this documentation is a safe place.
Number of copies	More than one copy of the data highly minimize the risk of losing data.	At least one copy must be kept in a fair distance from the mater copy. The medium can be disk as well as tape according to some storage rules that optimize cost and access speed.
Access protection	Access protection allows distributing data only to the people who are allowed.	In a multiuser environment, several security mechanisms can be achieved from a simple ID/password method to a more sophisticated certificate exchange.
Regular backup	Regular backup of data allows recovering data in case of a disaster.	Data and metadata must be kept synchronized.
Damage prevention	Limited life duration of disks and tape oblige to enable a continuous media management technique.	Continuous follow-up of media ages and apply a risk calculation of failure.

## Document Management and Control

Documents are the main method of communication between the team members, project management, and the communities within the EUDAT project. Therefore, the quality plan aims to manage and control the documents used in the whole project. This includes:

- a common documentation repository with controlled access for EUDAT project members;
- a versioning system dedicated for documents;
- a general validation process for every typology of documents;
- a notification mechanism to warn when a new document has been released.

## Service Design and Development Control and Procedures

The SQA plan recommends applying controls and procedures during each phase of the design, the

development and the service deployment activities of the EUDAT project.

Design reviews should be led by senior designers, software architects of the respective domain in collaboration with the community of users. A functional analysis aims to determine the correctness of the proposed design in order to validate the technical choices.

#### *Design changes*

In case of changes to the requirements, that impacts design choices, the changes should be reviewed and signed-off between projects managers and the community of users.

#### *Design monitoring*

Every design process, procedure and activity should be closely monitored and measured. By this type of control, the project management can make identify early whether the developments are progressing in the correct path.

#### *Training*

For this, a skill gap analysis is done to identify the training requirements at the project initiation phase. The project quality plan should indicate these training requirements and necessary steps to get the staff trained.

#### *Risk management*

The SQA plan identifies a number of quality risks. The project management team is responsible to come up with appropriate mitigation plans in order to address each quality risk.

#### *Managing service defects*

Service defects may occur at different stages of the service deployment. The SQA plan provides guidelines and instructions on how to manage the defects depending the stage of the service deployment.

## ANNEX F. SERVICE LEVEL AGREEMENT FOR B2SAFE AND B2SHARE

# SLA template

## Service Level Agreement

### governing the EUDAT services

Author(s) Pascal Dugénie, Marion Massol (CINES)  
Contributor(s) Ari Lukkarinen, Urpo Kaila, (CSC)  
Magchiel Bijsterbosch (SURFSARA)  
Johannes Reetz (RZG)  
Lorène Béchard (CINES)  
Status Draft  
Version v.0.9  
Date 15 March 2015  
References

#### legends

cross references (examples to be replaced when the template is filled)

text to be modified or inserted by corresponding lawyers

text to be filled by the contractors

#### About the SLA

The following document is an update of the Service Level Agreement (SLA) template for the EUDAT services B2SAFE and B2SHARE. In particular, it takes into account aspects for the B2SHARE service.

This version, delivered by task 6.4, can be regarded as template for the setup of bi- or multilateral contractual agreements (SLAs) and as basis for the further development of SLAs for B2-services which may be subject of future tasks and projects dealing with Service Portfolio management and SLAs.

## TABLE OF CONTENT

SLA front page for [ ] B2SAFE or [ ] B2SHARE

ART. 1: DEFINITIONS

ART. 2: SCOPE AND PURPOSE

ART. 3: LIAISON COMMITTEE

ART. 4: SUBMITTED DATA PROPERTY

ART. 5: SUBMITTED DATA INTEGRITY AND AUTHENTICITY

ART. 6: BACKEND SERVICE PROVIDER LIABILITY

ART. 7: BACKEND SERVICE PROVIDER OBLIGATION OF MEANS

ART. 8: BACKEND SERVICE PROVIDER LIMITS OF LIABILITY

ART. 9: SUBMISSION SERVICE LIABILITY

ART. 10: SUBMISSION SERVICE SUCCESSION

ART. 11: ACCESS RIGHTS

ART. 12: DATA RESTITUTION

ART. 13: DATA ELIMINATION

ART. 14: DURATION OF SERVICE

ART. 15: FINANCIAL CONDITIONS

ART. 16: FORCE MAJEURE

ART. 17: INSURANCE

ART. 18: INVALIDITY

ART. 19: CONTRACT DOCUMENTS

ART. 20: DISPUTES

## APPENDICES

SLA ANNEX 1: TERMS OF SERVICE (TOS)

SLA ANNEX 2: EUDAT TERMS OF USE (TOU)

SLA ANNEX 3: SPECIFICATIONS OF THE EUDAT SERVICES

STANDARDS AND POLICIES

ERRORS REPORTING PROCEDURES

RECOVERY PLAN

SLA ANNEX 4: VOLUME AND PRICE

SLA ANNEX 5: DESCRIPTION OF THE DATA TO BE PRESERVED

SLA ANNEX 6: ACCESS RIGHTS

SLA ANNEX 7: BACKEND SERVICE PROVIDER INSURANCE

## SLA ANNEX 8: COMPOSITION OF THE LIAISON COMMITTEE

## Service Level Agreement governing the EUDAT service

B2SAFE service

B2SHARE service

BETWEEN

*Name of backend service provider*

With headquarters located at *address - city - country*

Represented by its director *Name of the director*

hereafter identified as the BACKEND SERVICE PROVIDER

AND

*Name of the organization (legal entity) that will submit the data*

With headquarters located at *address - city - country*

Represented by its director *Name of the director*

hereafter identified as the SUBMISSION SERVICE

The BACKEND SERVICE PROVIDER and the SUBMISSION SERVICE are referred individually as PARTY and collectively as PARTIES.

THE FOLLOWING HAS BEEN AGREED

## Art. 1: Definitions

Availability	ability of a service to perform its required function at a stated instant or over a stated period of time [ITIL-V2].
BACKEND SERVICE PROVIDER	actor that has the mission and the responsibility to preserve the scientific data submitted on its platform by SUBMISSION SERVICE, and further enable the access and sharing of these data.
Bit stream preservation	basic level of preservation concerned with the maintenance of a digital resource. Its function is to ensure the continuing integrity of, and controlled access to, the digital objects which are contained within the preservation storage environment, including their associated metadata.
Confidentiality	ability of a service to ensure that information is accessible only to those authorized to have access and is protected throughout its lifecycle.
Data	Information received, preserved and communicated by a BACKEND SERVICE PROVIDER. They can include data and metadata.
Integrity	quality of data of being complete and unaltered [ISO/DIS 13008 – ISO 15489-1:2001].
Persistent identifier	an information linked to a data that identifies it persistently and uniquely within a given policy and reference system (eg. EPIC Handles are the PID referential used in the EUDAT project; they contain some fundamental properties of the referenced data and can be resolved into URI locators to access the data at specific storage locations).
Restitution	Data restitution is an operation of definitive return to the SUBMISSION SERVICE of data that has been stored to a BACKEND SERVICE PROVIDER.
SUBMISSION SERVICE	actor that is responsible in packaging data before transferring it to a BACKEND SERVICE PROVIDER
Transfer	operation initiated by a SUBMISSION SERVICE that consists in transmitting a package of data to a BACKEND SERVICE PROVIDER

## Art. 2: SCOPE and PURPOSE

The purpose of this Service Level Agreement (SLA) is to agree on terms to enable the BACKEND SERVICE PROVIDER to commit a set of professional and efficient resources, means and methods ensuring the SUBMISSION SERVICE preservation, access for digital scientific data.

This SLA focuses on the commitment for short term preservation. Long term preservation including a high level of data quality control is out of the scope of this SLA.

[ ] The level of the B2SAFE service includes bit stream preservation including a metadata service, integrity check and delivery of a Persistent Identifier (PID).

[ ] The level of the B2SHARE service includes bit stream preservation including a metadata service and a delivery of a Persistent Identifier (PID).

### Art. 3: LIAISON COMMITTEE

This SLA is managed by a liaison committee, in charge of establishing and administering relationships between SUBMISSION SERVICE and BACKEND SERVICE PROVIDER, and in particular to deal with activity planning and tracking.

The liaison committee is also responsible to approve phases preceding the setup of data submission. The validation of these phases are mandatory in order to ensure that data preparation, testing and development are conducted in partnership between SUBMISSION SERVICE and BACKEND SERVICE PROVIDER.

The members of this liaison committee is listed in [Annex 8](#).

### Art. 4: SUBMITTED DATA PROPERTY

Data preservation by the BACKEND SERVICE PROVIDER doesn't include the transfer of the property of those data.

The SUBMISSION SERVICE ensures BACKEND SERVICE PROVIDER a prior agreement with producers (see Terms of Use in [Annex 2](#)), authors or rights owners for the data is available, whereby the SUBMISSION SERVICE is duly authorised to transfer the relevant documents to the BACKEND SERVICE PROVIDER.

### Art. 5: SUBMITTED DATA INTEGRITY AND AUTHENTICITY

Through this SLA, the SUBMISSION SERVICE declares:

- (i) to submit to the BACKEND SERVICE PROVIDER, the data in their digital, original and genuine format;
- (ii) to comply with a secure means of transfer.

The data submitted is provided with relevant information concerning its identification and properties as well as any information required for its exploitation.

### Art. 6: BACKEND SERVICE PROVIDER LIABILITY

#### *Basic functionalities*

Through this SLA, the BACKEND SERVICE PROVIDER:

- is responsible for preservation of data transferred to its platform in the level of service specified in **Art. 2** of the present SLA;
- commits to proceed to storage and to ensure the AVAILIBILITY of data entrusted to him;
- guarantees data preservation in conformance with the functional and technical specifications of the BACKEND SERVICE PROVIDER system (see **Annex 3**);
- commits to Terms of Service (ToS) that indicates the level of service related to data integrity confidentiality, availability and accessibility in an as efficient as possible response time (see **Annex 1**);
- commits to provide appropriate physical support migrations when needed;
- undertakes to provide, upon SUBMISSION SERVICE request, a periodical status of the preserved data.

[ ] **[B2SAFE]** Also, the BACKEND SERVICE PROVIDER guarantees to the SUBMISSION SERVICE:

- that its policies and procedures are compliant with national and international law and regulations;
- takes the responsibility of maintaining the consistency of the PID and references used to identify all data stored in its systems;
- that all the operations performed on the data preserved are traceable;
- that processes for incident response are documented.

[ ] **[B2SHARE]** Also, the BACKEND SERVICE PROVIDER guarantees to the SUBMISSION SERVICE:

- that its policies and procedures are compliant with local, national and international law and rules;
- takes the responsibility about maintaining the consistency of the PID and metadata used to identify each data stored in its systems;
- that all the operations done on the data preserved can be traceable;
- that processes for incident response are documented.

By default, the will not guarantee the preservation of non-compliant data integrity, authenticity and possible confidentiality, or reserves the right to refuse them.

### *Complementary EUDAT services*

The BACKEND SERVICE PROVIDER offers to the SUBMISSION SERVICE a set of complementary EUDAT services such as:

- A Resource Provisioning Coordination Tool: available at <https://rct.eudat.eu/> to manage the resource request.
- A helpdesk is operated (<http://helpdesk.eudat.eu>) and can be used by the SUBMISSION SERVICE via <http://www.eudat.eu/support-request> to raise issues.
- A service monitoring tool is available at <https://cmon.eudat.eu/> and shows services availability real time status.

## Art. 7: BACKEND SERVICE PROVIDER OBLIGATION OF MEANS

The obligations undertaken hereunder by the BACKEND SERVICE PROVIDER have the nature of obligations of means.

The BACKEND SERVICE PROVIDER commits to provide to SUBMISSION SERVICE all the means at his disposal to ensure SUBMISSION SERVICE entrusted data preservation and access, in particular:

- Ensure management of submitted data
- Setup service level, with physical and logical security, compliant with the general Terms of Service (ToS) conditions (see [Annex 1](#))
- Preserve data received (except in case of elimination or return) in compliance with the functional and technical specifications of the BACKEND SERVICE PROVIDER system (available in the [Annex 3](#));
- Record, at a place or a method accessible to SUBMISSION SERVICE, information and technical descriptions allowing data recovery.
- Allow a controlled access to data in accordance with defined conditions and to consumers duly defined in [Annex 6](#) of this SLA or in a subsequent amendment.
- Inform SUBMISSION SERVICE at least  days before, about any major technical modification deployed on the system.
- Inform SUBMISSION SERVICE about any error on its data in accordance with the reporting procedure defined in the [Annex 3](#).
- Have suitable written disaster preparedness and recovery plan(s) (cf. [Annex 3](#)), including at least one off-site backup of all preserved information together with an offsite copy of the recovery plan(s).
- Inform SUBMISSION SERVICE about any use of external suppliers by BACKEND SERVICE PROVIDER, about outsourced service type and about agreements made for this use to guarantee an equivalent service level to SUBMISSION SERVICE.
- The forecasted volume of data to be stored should be provided by the SUBMISSION SERVICE in [Annex 4](#).

These figures will be reviewed yearly by mutual agreement between both parties.

## Art. 8: BACKEND SERVICE PROVIDER LIMITS OF LIABILITY

### *Data description*

The BACKEND SERVICE PROVIDER cannot be held liable for associating data and description of these data. The SUBMISSION SERVICE must provide to the BACKEND SERVICE PROVIDER such a definition of the more relevant description level that is associated with the data.

#### *Bit stream preservation*

The BACKEND SERVICE PROVIDER undertakes only the bit stream preservation and integrity control checks (at ingest as well as periodically). No further control of the format of the data is done.

#### *Data sustainability*

The BACKEND SERVICE PROVIDER will not commit itself to submitted data sustainability and understanding.

#### *Data content*

The BACKEND SERVICE PROVIDER is not responsible for the content of the submitted data. The SUBMISSION SERVICE must take all actions to ensure compliance with legislation in force.

The BACKEND SERVICE PROVIDER will not perform any data selection, content verification, sorting or cross indexation.

In no event shall BACKEND SERVICE PROVIDER be liable in case of a fault, negligence, omission or failure from SUBMISSION SERVICE, cases of force majeure or events beyond the BACKEND SERVICE PROVIDER control.

The BACKEND SERVICE PROVIDER liability under the terms of this SLA may under no circumstances be invoked except when a material damaged is directly caused by a proven failure from the BACKEND SERVICE PROVIDER. Under no circumstances the BACKEND SERVICE PROVIDER may be held accountable for neither indirect nor immaterial damages nor operating losses caused directly or indirectly by using BACKEND SERVICE PROVIDER.

## Art. 9: SUBMISSION SERVICE LIABILITY

Through this SLA, the SUBMISSION SERVICE

- is liable for associating data and description of submitted data and to provide a definition of the more relevant description level that is associated with the data;
- commit itself to submitted data sustainability and understanding;
- initiate the data transfer to BACKEND SERVICE PROVIDER;
- is responsible for the content of the submitted data and must take all actions to ensure compliance with **legislation in force**;
- is responsible of data selection, content verification, sorting or cross indexation before any transfer.

## Art. 10: SUBMISSION SERVICE SUCCESSION

Any natural person, legal person or beneficiary who could succeed the SUBMISSION SERVICE must prove his right to succession and in particular the powers referred to in this agreement, and establish a new contract with the BACKEND SERVICE PROVIDER.

## Art. 11: ACCESS RIGHTS

The policy that governs access rights is defined by the SUBMISSION SERVICE after validation by BACKEND SERVICE PROVIDER. This policy is attached as [Annex 6](#).

In no event shall BACKEND SERVICE PROVIDER be liable neither in case of communication to users as soon as they have been granted access by SUBMISSION SERVICE, nor usage (holding, preservation and communication) of data that can be made by SUBMISSION SERVICE or granted users.

In no event shall BACKEND SERVICE PROVIDER be held liable for access, or any usage (holding, preservation, communication) that follows from said access, that adheres to the access policy as governed by Annex 6 and was successfully validated as such by the BACKEND SERVICE PROVIDER.

In accordance with this policy by BACKEND SERVICE PROVIDER, the SUBMISSION SERVICE holds the BACKEND SERVICE PROVIDER free from harm from all third-party remedies.

## Art. 12: DATA RESTITUTION

Any claim for restitution of data to the SUBMISSION SERVICE must be made to BACKEND SERVICE PROVIDER taking into account a six months' notice before the effective date of data return. It can take place at the end of the agreement, in case of cessation of activity of the BACKEND SERVICE PROVIDER or any case of force majeure. The Parties shall agree in time on a return protocol.

The BACKEND SERVICE PROVIDER agrees to return all electronic records and associated elements in the technical condition they were in when submitted to the system.

This return may result in financial compensation for the BACKEND SERVICE PROVIDER based in particular on time, volume to be processed and return protocol to the SUBMISSION SERVICE. The calculation of this compensation will be made based on a minimum effort required and the possible purchase of equipment required for the process of data restitution.

## Art. 13: DATA ELIMINATION

Data elimination can be initiated in any of the following conditions:

1. on request of the SUBMISSION SERVICE for deletion of specific data;
2. when data has reached the end of utility, this limit can be decided with the agreement of both parties;
3. after restitution as mentioned in [Art. 12](#)

The BACKEND SERVICE PROVIDER ensures that, once the elimination is completed, the corresponding data on its platform will be fully destroyed without any copies left.

The BACKEND SERVICE PROVIDER will send a notification to SUBMISSION SERVICE prior to elimination of data. The data will be automatically destroyed within [redacted] days after notification.

The BACKEND SERVICE PROVIDER can keep track and show the path of data on its platform.

After an operation of elimination, the BACKEND SERVICE PROVIDER is relieved from any obligation.

## Art. 14: DURATION OF SERVICE

During the life of this SLA, BACKEND SERVICE PROVIDER is committed to preservation duration defined at the data submission time.

By default, data preservation duration is defined as the duration of EUDAT project (until [redacted]).

Either party may terminate the agreement, subject to six months' notice. In this case, the records will be returned to the SUBMISSION SERVICE in the terms described in [Art. 12](#).

## Art. 15: FINANCIAL CONDITIONS

As both parties are partners of the EUDAT project, the service is provided is free for the duration of the EUDAT project.

After this date, the cost will be determined in accordance with economic models proposed by the EUDAT WP2. Three costs models have been proposed and described in [Annex 4](#).

## Art. 16: FORCE MAJEURE

Any event of force majeure as defined in [ad hoc reference in the EU texts of laws to be inserted by a legal consultant](#) has the effect to interrupting or even discontinue if the continuation of the contract is not possible even after interruption, the effects of this agreement, in particular regarding responsibilities and liabilities of each party.

## Art. 17: INSURANCE

The BACKEND SERVICE PROVIDER must have contracted an insurance policy (attached in [Annex 7](#)), in particular to cover damages and losses that may incur the deposited data.

The SUBMISSION SERVICE may choose to be insured for this type of damages.

## Art. 18: INVALIDITY

In the event that any stipulation of this SLA is invalid or unenforceable by law, regulation, competent jurisdiction after a final decision, all other stipulations of this SLA will continue to apply and [disability effect only apply to the portion of the SLA](#) immediately in question, unless otherwise agreed by the parties.

In addition, the Parties undertake, in good faith negotiations to replace the invalid or unenforceable stipulations with other stipulations whose effects are comparable.

Failure by either party to achieve invalid or unenforceable stipulations replacement, should not affect the validity of the remaining stipulations neither the valid part of a partly invalid stipulation which will take effect to the extent permitted by law.

## Art. 19: CONTRACT DOCUMENTS

No other document will render obligations under this SLA unless it has previously been the subject of an amendment signed by the parties and attached to this agreement.

This SLA and its annexes supersede any verbal or written agreements that the parties could previously have made.

## Art. 20: DISPUTES

In case of dispute and should the parties fail to reach an agreement, the parties expressly conferred exclusive jurisdiction on the court of **name of the European court to be inserted by a legal consultant**.

## APPENDICES

The following additional documents are attached. They will be maintained as necessary following the annual meeting of the Liaison Committee.

- Annex 1: Terms of Service (ToS) for the BACKEND SERVICE PROVIDER
- Annex 2: EUDAT Terms of Use (ToU)
- Annex 3: Functional and technical specifications of the backend service provider system:  
This annex includes also:
  - additional standards and policies for the preservation of digital data,
  - reporting procedure for errors on the data preserved,
  - disaster preparedness and recovery plan(s).
- Annex 4: Maximum volumetric and price
- Annex 5: Description of the data to be preserved
- Annex 6: Access rights and policies
- Annex 7: BACKEND SERVICE PROVIDER insurance contract
- Annex 8: Composition of the liaison committee

Done in two (2) original copies,

<date>

For the SUBMISSION SERVICE	For the BACKEND SERVICE PROVIDER
<Name>	<Name>
<Function>	<Function>

## SLA annex 1: Terms of Service (ToS)

The Terms of Service (ToS) specifies the technical environment at the BACKEND SERVICE PROVIDER.

This document must contain the following sections:

1. EQUIPMENT  
including hardware, software, applications and system environment
2. NETWORKS
3. GENERAL OPERATING CONDITIONS  
including the system availability and average yearly service outages
4. APPLICATIONS AND SERVICES
5. USER ADMINISTRATION
6. SUPPORT  
including support for end-users, technical assistance, support for the SUBMISSION SERVICE
7. DATA MANAGEMENT  
including disk/tape space, backup, recovery procedures, conditions for long term archiving
8. SECURITY
9. DATA CONFIDENTIALITY
10. UPGRADES
11. OPERATIONAL SUMMARY
12. DOCUMENTS AND ACCESS TO USE DATA
13. QUALITY OF SERVICE, LIMITS OF LIABILITY

## SLA annex 2: EUDAT Terms of Use

*The Terms of Use (ToU) must have been agreed by the SUBMISSION SERVICE*

## SLA annex 3: Specifications of the EUDAT services

Service	links to the service with general specifications
B2SAFE	<a href="http://www.eudat.eu/b2safe">http://www.eudat.eu/b2safe</a>
B2SHARE	<a href="http://www.eudat.eu/b2share">http://www.eudat.eu/b2share</a> , <a href="https://b2share.eudat.eu">https://b2share.eudat.eu</a>

### *Standards and policies*

Table of standards and policies in the domain of preservation of digital data should be inserted here:


### *Errors reporting procedures*

Any reporting procedure for errors on the data preserved should be inserted here:


### *Recovery plan*

Any relevant information about disaster preparedness and recovery plan should be inserted here.


## SLA Annex 4: Volume and price

An estimate of forecast of the data volume is made at the SLA signature. It will be reviewed yearly by the liaison committee during follow-up meetings.

year	Number of files		Cumulative nb of files		Volume (in Gb)		Cumulative volume	
	Min.	Max.	Min	Max.	Min.	Max.	Min.	Max.
2015								
2016								
2017								
2018								

The price associated to the service is calculated using the following formula.

The BACKEND SERVICE PROVIDER shall insert here the formula for price calculation.

## SLA Annex 5: Description of the data to be preserved

*The SUBMISSION SERVICE shall insert here the description of the data to be preserved.*

## SLA Annex 6: Access rights

As part of this SLA for the service:

B2SAFE

B2SHARE

between

the BACKEND SERVICE PROVIDER *Name of backend service provider*

and

the SUBMISSION SERVICE *Name of the organisation that will submit the data.*

level of access right	yes/no
access to the stored data at the BACKEND SERVICE PROVIDER is offered to a wide audience (ie. public with anonymous access)	
access to the stored data at the BACKEND SERVICE PROVIDER is offered to a restricted audience (ie. list of users with identification)	
access to the data stored at the BACKEND SERVICE PROVIDER is strictly limited to the SUBMISSION SERVICE by using a dedicated account provided by the BACKEND SERVICE PROVIDER to the SUBMISSION SERVICE	

## SLA Annex 7: Backend service provider insurance

*The BACKEND SERVICE PROVIDER shall insert here the insurance contract.*

## SLA Annex 8: Composition of the liaison committee

The liaison committee is composed of the following members:

Name of the representative	Institution	Role in the committee