


## D6.1: Report on Status and Progress of Operations - Year 1

Author(s)	Johannes Reetz (MPCDF), Pavel Weber (KIT), Giovanni Morelli (CINECA), David Vicente (BSC), Urpo Kaila (CSC), Ioannis Liabotis (GRNET), Christos Kanellopoulos (GRNET), Dejan Vitlacil (KTH)
Status	Final
Version	v1.0
Date	20/07/2016

**Abstract:** This document about WP6 “Operations” reports on the service management by and within the EUDAT Common Data Infrastructure (CDI), the coordinated provisioning of resources and services, the operational tools that were used, and the support services (PID and AAI services) that were provided during the first year of the EUDAT2020 project. The instantiation and operation of services and suites of services in the context of data projects is another important task that is handled by WP6. This document therefore also describes the underlying conceptual approach that was used for implementing data projects and the “recipes” that were developed for service enabling, and reports EUDAT’s achievements in implementing research data projects and pilots that were requested by research communities during the first year of EUDAT2020. Data projects that were enabled during the first phase of the EUDAT project, and that EUDAT has operated since then, were reviewed and are still being managed. Among the main achievements in operations are the formalising of clear processes, roles and responsibilities according to the FitSM standard, the design and implementation of operational tools for managing the service portfolio and the data projects, the establishment of a CDI identity management system which allows users to register with EUDAT using identities provided by other identity providers, the extension of the network of PID service providers within the EUDAT CDI, the set-up of an availability and reliability monitoring system, consolidating the coordination of the daily core operations (provisioning and management of resources and services), providing intensified 1<sup>st</sup> and 2<sup>nd</sup> level user support, developing “recipes” for service enabling, and the beginning of planning for and the enabling of over 30 new data projects and pilots for customers with an aggregated total of about 40,000 users.

Document identifier: EUDAT2020-DEL-WP6-D6.1	
Deliverable lead	MPCDF
Related work package	WP6
Author(s)	Johannes Reetz (MPCDF), Pavel Weber (KIT), Giovanni Morelli (CINECA), David Vicente (BSC), Urpo Kaila (CSC), Ioannis Liabotis (GRNET), Christos Kanellopoulos (GRNET), Dejan Vitlacil (KTH)
Contributor(s)	Raphael Ritz (MPCDF), Debora Testi (CINECA), Rob Baxter (EPCC), Margareta Hellström (ULUND)
Due date	29/02/2016
Actual submission date	20/07/2016
Reviewed by	Erwin Laure (KTH), Rob Baxter (EPCC)
Approved by	PMO
Dissemination level	PUBLIC
Website	www.eudat.eu
Call	H2020-EINFRA-2014-2
Project Number	654065
Start date of Project	01/03/2015
Duration	36 months
License	Creative Commons CC-BY 4.0
Keywords	CDI, Service Management, Operations, Resource and Service Provisioning, Data Projects, Data Project Management Tool, Service Portfolio Management Tool, User Support, Persistent Identifiers, Identity Management, Security and Data Privacy

*Copyright notice:* This work is licensed under the Creative Commons CC-BY 4.0 licence. To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0>. 

*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.

## TABLE OF CONTENT

<b>EXECUTIVE SUMMARY .....</b>	<b>6</b>
<b>1. INTRODUCTION .....</b>	<b>8</b>
<b>2. THE SERVICE MANAGEMENT FRAMEWORK .....</b>	<b>11</b>
2.1. The Service Management Scope .....	11
2.1.1. CDI Actors .....	11
2.1.2. Scope: CDI Common Services .....	11
2.1.3. Scope: Central B2 Service Instances .....	12
2.2. Service Management Principles .....	13
2.3. Service Management Tools .....	13
2.4. Service Catalogue .....	14
2.5. Service Management Processes .....	15
<b>3. CENTRAL SERVICES AND OPERATIONAL TOOLS .....</b>	<b>20</b>
3.1. Data Project Management .....	20
3.1.1. The DPMT between Service Management and Data Management Planning .....	22
3.1.2. Data Project Use Cases and Workflows .....	23
3.1.3. The DPMT Data Model .....	24
3.1.4. Data Project Phases .....	26
3.1.5. Actors and Roles for Data Project and Service Management .....	27
3.1.6. DPMT Implementation Status .....	27
3.2. Service Portfolio and Catalogue Management Tool .....	29
3.2.1. SPMT Architecture .....	30
3.2.2. Model/Database Design .....	30
3.2.3. Main Use Cases .....	31
3.2.4. API Basics .....	31
3.3. Monitoring .....	32
3.3.1. Migration and Operation of the Legacy CMON Service .....	32
3.3.2. Design and Implementation .....	32
3.3.3. Status of the Availability and Reliability Monitoring Service .....	33
3.4. B2ACCESS – A Federated Identity and Access Management Solution for EUDAT .....	35
3.4.1. The Production Environment for B2ACCESS .....	36
3.4.2. eduGAIN SAML IdPs .....	36
3.5. B2HANDLE Setup and Operation .....	37
<b>4. SERVICE AND RESOURCE PROVISIONING .....</b>	<b>39</b>
4.1. Operations Coordination .....	39
4.2. Service Installation Support for Release and Deployment .....	40
4.3. Change management .....	41
4.4. Software Version Monitoring .....	42
4.5. Service Management and Operational Tools .....	43
4.6. Service Hosting Framework .....	44
4.7. Ensuring Continuing Availability and Reliability .....	45
<b>5. DATA PROJECT ENABLING AND SERVICE INTEGRATION .....</b>	<b>46</b>
5.1. Introduction .....	46
5.2. Review of existing data projects .....	48
5.3. Uptake Plans and Service Integration .....	49
5.3.1. EPOS .....	49
5.3.2. ICOS .....	51
5.3.3. LTER .....	53

5.3.4.	CLARIN .....	53
5.4.	Data Pilots and Project Enabling Activities .....	53
5.5.	Templates and Recipes for the Project enablers.....	54
<b>6.</b>	<b>HELPDESK AND SUPPORT SERVICE.....</b>	<b>57</b>
6.1.	First level support .....	57
6.2.	Second Level Support .....	57
6.3.	Third Level Support.....	58
6.4.	Helpdesk System and Statistics .....	58
6.5.	Support activities.....	59
<b>7.</b>	<b>SECURITY.....</b>	<b>60</b>
7.1.	Vulnerability Management.....	60
7.2.	Security Scans .....	61
7.3.	E-infrastructure Security.....	61
7.4.	Security in Software Development.....	61
7.5.	Building Operational Trust and Compliance - DSA .....	61
7.6.	WISE - Liaison with other Research Infrastructures .....	62
7.7.	Security Policy and AUP.....	62
7.8.	Joint Security Training .....	62
7.9.	Risk Management.....	62
<b>8.</b>	<b>FUTURE WORK .....</b>	<b>63</b>
8.1.	Central Services and Operational Tools.....	63
8.2.	Service and Resource Provisioning.....	63
8.3.	Project enabling and Service Integration .....	64
8.4.	Helpdesk and Support .....	64
8.5.	Security .....	64
<b>ANNEX A.</b>	<b>GLOSSARY.....</b>	<b>65</b>
<b>ANNEX B.</b>	<b>CDI SERVICE MANAGEMENT PROCESSES AND ROLES.....</b>	<b>70</b>
<b>ANNEX C.</b>	<b>DATA PROJECT MANAGEMENT TOOL - USE CASES AND WORKFLOWS .....</b>	<b>72</b>
<b>ANNEX D.</b>	<b>A &amp; R MONITORING AND THE TOPOLOGY OF B2 SERVICES .....</b>	<b>79</b>
<b>ANNEX E.</b>	<b>SOFTWARE VERSIONS MONITORING FRAMEWORK .....</b>	<b>84</b>
<b>ANNEX F.</b>	<b>DATA PROJECTS .....</b>	<b>85</b>

## LIST OF FIGURES

Figure 1: Tasks within the Operations work package.....	9
Figure 2: The classes of services within the EUDAT CDI, corresponding B2 services are indicated.....	12
Figure 3: The DPMT connected to the SPMT, the DMP Tool and the Configuration Information Service .....	22
Figure 4: High level schematics (UML) of the DPMT Data Model. ....	24
Figure 5: Distributed Services as Compositions of Service Components .....	25
Figure 6: Data Project Phases .....	26
Figure 7: DPMT Projects overview panel.....	27
Figure 8: DPMT Registered Services overview panel .....	28
Figure 9: DPMT Registered Service Components overview panel .....	28
Figure 10: The SPM Tool – High-Level Model (UML).....	31
Figure 11: Centralized monitoring of B2 services from the user perspective .....	33
Figure 12: Architecture of the availability and reliability monitoring .....	34

Figure 13: Availability & Reliability monitoring panel. The probes for the service instances are preliminary.	35
Figure 14 B2ACCESS schematic showing the integration status in Feb 2016.	36
Figure 15 Evolution of the number of users signed in with B2ACCESS (snapshot from April 2016).	37
Figure 16: B2SAFE automatic installation via Puppet.	40
Figure 17: Product versions monitoring system.	42
Figure 18: Project enabling workflow.	48
Figure 19: Schematic integration plan for EUDAT services in the EPOS-ORFEUS-EIDA pilot.	50
Figure 20: EPOS-INGV use case.	51
Figure 21: Data flow in ICOS together with the relevant B2 services that ICOS intends to use.	52
Figure 22: Schematic ICOS data flow (preliminary).	52
Figure 23: Distribution of service requests by the data pilots.	54
Figure 24: Storage Service Provider requested by the data pilots.	54
Figure 25: Decisions at the beginning of the B2 service enabling process.	55
Figure 26: Decision graph for the enabling of B2FIND.	55
Figure 27 User tickets per service queue.	58
Figure 28: Support Requests managed per service per month (cleaned statistics).	59
Figure 29: View on the panel that informs about installed software versions.	84

## LIST OF TABLES

Table 1: CDI Central B2 Service Instances and the current Service Providers.	13
Table 2: Service Management Tools.	14
Table 3: Service Portfolio indicating the contents of the EUDAT Service Catalogue.	14
Table 4: Data Projects from the first EUDAT2020 project year.	21
Table 5: Example of the SPMT API.	32
Table 6: Summary of activities related to automatic iRODS/B2SAFE installation procedure.	41
Table 7: Statistics of currently registered service components.	43
Table 8: Services installed during the 2 <sup>nd</sup> phase of EUDAT.	43
Table 9: The Service Hosting Environment.	44
Table 10: Result of the review of the existing data projects from EUDAT 1.	49
Table 11: DPMT Use Cases and Workflows.	72
Table 12: Project Phases.	77
Table 13: EUDAT Products for monitoring and Service Type.	80
Table 14: Status Update existing Data Projects from EUDAT1.	85
Table 15: New Data Projects and Pilots in EUDAT2020.	86

## EXECUTIVE SUMMARY

The EUDAT2020 project under the Horizon 2020 programme is the continuation of the first phase of the EUDAT project under the FP7 programme. The work by EUDAT's Work Package 6 (WP6) on *Operations* (and the related work on *Data Project Enabling*) in this phase of EUDAT is based on the achievements described in the deliverables about *Operations* from the previous phase of EUDAT and also relies on the deliverables from several other EUDAT work packages, namely the *Community Requirements and Engagement* work package (WP4) where the master data management plans of the research communities (called service uptake plans) and the data pilots from the first call were particularly relevant, the *Service Building* work package (WP5), and the *Service Strategy, Policy and Sustainability* work package (WP2). The architecture of the *EUDAT Collaborative Data Infrastructure* (CDI)<sup>1</sup> and the B2 data management services offered by EUDAT (in the form of software packages) have served as the foundation for detailed planning and the implementation of data projects (and data pilots) and the related service management within the EUDAT CDI. This deliverable provides an overview of the activities relating to operations, including the work on planning and enabling data projects, during the first year of this phase of the EUDAT project.

Progress was made on consolidating the operational environment of the EUDAT CDI – this was achieved by implementing the service management processes according to the FitSM-2015<sup>2</sup> standard. This is an ongoing activity that started in the first year of this phase of EUDAT and the work will continue in the following years, particularly in relation to improving the service and operations level agreements and the various service management policies. WP6 organises the cooperation between the EUDAT CDI *Service Providers* via an *Operations Coordination Team (OPCT)* that comprises senior site deputies representing the providers. The provisioning of services and resources by the provider is coordinated by the manager for EUDAT's Task "Service and Resource Provisioning" as lean as possible and the OPCT facilitates the communication if situations arise where expectations or problems concerning the timely provisioning of resources and services need to be managed. The OPCT is also responsible for maintaining and operating the CDI services according to the relevant EUDAT service management policies and guidelines for information security and data privacy. Policies and guidelines were created, and will be maintained and further developed, following the FITSM approach to the appropriate degree.

The specific tasks undertaken by WP6 are:

- integrating, commissioning and operating instances of EUDAT services (including the components of distributed services that are part of the CDI), which is handled on a daily basis by the operations coordination team (The OPCT is in charge of coordinating and supervising the EUDAT resource and service provisioning, along with any maintenance that is needed by data projects from when they start being enabled until they are terminated; in particular this includes handling calls for resource pledges and providing storage capacities for European researchers and their research organisations upon request.),
- providing service hosting capacities for running EUDAT data services at data centres, and harmonizing the procedures for requesting and obtaining data processing capacities in the context of data projects,
- supporting the research communities that are involved with EUDAT and their users when it comes to using the EUDAT services – particularly the user-facing services B2DROP, B2SHARE, B2STAGE, B2FIND and B2ACCESS, which is EUDAT's federated AAI service; also helping to enable B2HANDLE PID services and supporting the process of connecting research data repositories to the EUDAT CDI with B2SAFE, and
- ensuring the operational and infrastructure security of the EUDAT CDI, which is achieved via consultation with the security team during both service building and the configuration of the services in the course of implementing data projects, and also through performing regular security vulnerability scans.

<sup>1</sup> EUDAT2020 D5.1 deliverable: Report on Service Building Status and Progress, year 1

<sup>2</sup> FitSM-2015, <http://fitsm.itemo.org/fitsm-standard>, referenced April 2016.

The work package's main achievements include the following.

- The operations work package has been set up according to the new EUDAT task structure, with bi-weekly WP task leader meetings, and at least monthly meetings of the operations coordination team and the project enabling team.
- WP6 has started establishing the EUDAT service management framework (that is, the relevant service management processes are defined according to the FitSM-2015 recommendations) and has started specifying service management policies that reflect many of the current practises in operations.
- A production instance of B2ACCESS (<https://b2access.eudat.eu>) has been made available – it is being operated by FZ Juelich in compliance with the policy statements (GEANT Code of Conduct, Data Privacy Statement) demanded by eduGain. B2ACCESS has been integrated into the B2SHARE instance <https://b2share.eudat.eu>, the collaboration platform B2GETHER <https://b2gether.eudat.eu>, the pre-release of the EUDAT data project management tool (<https://dp.eudat.eu>) and the EUDAT central site and service registry (<https://creg.eudat.eu>). B2ACCESS has been developed so that it accepts primary identities from various identity providers (IdPs) such as eduGain IdPs and social IdPs (Google, Facebook, MS Live, Github), and so it accepts personal x.509 certificates (IGTF, Geant, DFN Global) and allows users to register directly on the B2ACCESS server (B2ACCESS account, self-registration).
- With GRNET and BSC as additional B2HANDLE service providers, there are now six ePIC<sup>3</sup> PID service instances in production; these PID service providers are mirroring their PIDs to other ePIC partners (including GWDG which is one of EUDAT's associated partners).
- WP6 designed the data and workflow model for the EUDAT Data Project Management and CDI information management and provided a pre-release of the DPMT (<https://dp.eudat.eu>) while it is still being developed. This tool offers a comprehensive CDI information and data project management facility that provides authoritative information about customers (research communities), service providers (sites), resource and service offers, instantiated (registered) services and related components, as well as the data projects at different stages. The DPMT can be connected to the EUDAT Service Portfolio Management Tool and makes it possible to use actual information about existing service definitions.
- WP6 has worked on the design of the data model for, and the implementation of, the Service Portfolio Management Tool that provides authoritative information about the EUDAT service definitions and the service catalogue. The latter includes controlled names (defined by the service developers) of the configuration parameters that are relevant for the configuration of the service instances (e.g. software versions). The tool will be mainly used by WP2 and WP5.
- WP6 has upgraded the EUDAT Helpdesk system and has consolidated the first and second level user support with reduced response times – this has been achieved through having external requests handled by a dedicated first level support team who have access to a recently compiled catalogue of answers to frequently asked questions.
- The Data Project Enabling Team provides consultancy for implementing data management plans using the EUDAT CDI services – this includes coordinating the provision of the necessary capacities.
- WP6 has worked on enabling data projects from research community uptake plans as well as data pilots.
- Under the guidance of the OPC team, WP6 has continuously rolled-out and maintained the services and provided the resources needed for the data projects and pilots.
- WP6 developed a unified configuration management and deployment procedure based on Open Source tool *Puppet*. Currently the B2SAFE installation on several operating systems is supported.
- WP6 developed and provides a tool for monitoring the versions of actually deployed B2 service packages and component software at the different sites. Sites are notified if new releases of software for services that they are hosting are available.

<sup>3</sup> <http://www.pidconsortium.eu>



## 1. INTRODUCTION

The strategic vision of the EUDAT project is to enable European researchers and practitioners from any research discipline to preserve, find, access, and process research data in a trusted environment. EUDAT is realising this vision by enriching the pan-European e-infrastructure landscape with a collaborative data infrastructure. The EUDAT Collaborative Data Infrastructure (CDI) is a network of collaborating, cooperating centres, combining the richness of numerous data repositories that hold data specific to particular research disciplines with the permanence of some of Europe's largest data and computing centres for research. Together these centres provide a layer of interoperable data services based on an architecture (explained in the D5.1 deliverable<sup>4</sup>) and adhering to a set of common technical standards and policies.

The current phase of the EUDAT project, EUDAT2020, is the continuation of the initial successful first phase of EUDAT, which finished on the 31<sup>st</sup> of March 2015. EUDAT2020 started on the 1<sup>st</sup> of March 2015. The strategy that this phase of the EUDAT project has followed during its first year, and that it will continue follow over the course of the next two years, aims at 1) bringing more instances of the EUDAT services into production, 2) integrating those instances with each other by bundling them into suites of service instances in the course of data projects, and in this context by supporting their integration with complementary e-infrastructure services that are offered via service catalogues, and 3) providing high-quality trustworthy data platforms that make it easier for researchers and research communities to collaborate with each other across Europe and also globally.

The purpose of EUDAT's Operations work package (WP6) is to further develop and consolidate the operational environment of the EUDAT CDI, and particularly to implement the service management framework, ensuring the authenticity, integrity, confidentiality, retention and preservation of data deposited by its users or on their behalf, especially those marked for long-term archiving. WP6 organises the co-operation of the different actors within the CDI, relying on an operations coordination team (OPCT) with senior site deputies supported by their staff to ensure that the required resources and capacities are provided on time and maintained according to operational and policy guidelines, following the FitSM IT service management principles.

The Operations work package has these particular objectives:

- to integrate, commission and operate the service instances of the EUDAT CDI,
- to coordinate the provision of storage capacities to European researchers and their research infrastructures based on a variety of technologies,
- to provide service hosting capacities to run data services at data centres,
- to provide support to research communities and users in connecting their repositories to the CDI and using the EUDAT services, and
- to ensure policy compliance, along with the operational and infrastructure security of the CDI.

Specifically this work package establishes and runs the service management framework that defines processes, roles and responsibilities, although several aspects of this fall primarily in the realm of other work packages: WP2 (general service process management), WP4 (customer relationship management) and WP5 (management of the improvement of services and third level support).

There is a strong link between WP6 and the other work packages. The *Service Strategy, Policy and Sustainability* work package (WP2) orchestrates the service management process within EUDAT via the EUDAT *service portfolio* and the EUDAT *service catalogue*<sup>5</sup>. A corresponding Service Portfolio Management Tool has been designed and developed within WP6 and is provided to all the organisations that have a role in and responsibility for managing the service portfolio (and catalogue) or that just use the information in the portfolio.

---

<sup>4</sup> See EUDAT2020 WP5 deliverable D5.1, 2016, Chapter 3.

<sup>5</sup> The concepts of the Service portfolio and Service catalogue are explained in EUDAT2020 D2.1 "Service Portfolio Processes Definition and SLA Template Set"



The purpose of this document is to provide an overview of the operational part of the EUDAT service management framework and of the activities that were conducted in the course of operating the EUDAT CDI, such as managing data projects and providing helpdesk services.

The work of the operations work package is organised into five tasks which are shown in Figure 1.

The first task (6.1) *Central Services and Operational Tools* involves developing and providing the central EUDAT operational services and also takes care of the provisioning of the PID and AAI services. These services (or tools) provide functions that are essential for running and sustaining the EUDAT CDI, such as information management and reporting services, e.g. for registering service providers, their resources and services, or registering customers and managing their data projects. These services provide monitoring and accounting information and they make it possible to manage the service portfolio.

The second task (6.2) involves *Service and Resource Provisioning*. The work under this task coordinates the provision of the resource capacities, service components and service hosting facilities that are needed to deliver the EUDAT services which are required by the customer data projects. The priority in the first year of this phase of the EUDAT project was to deploy the services for the research communities directly involved in the EUDAT consortium, based on the agreed uptake plans. The site deputies and staff involved in the Operations Coordination Team have developed service management policies (e.g. the configuration and change management in the CDI) as well as ensuring both that the software of the service instances is up-to-date and safe and that risks of disasters (service disruptions, data loss) in the CDI are mitigated.

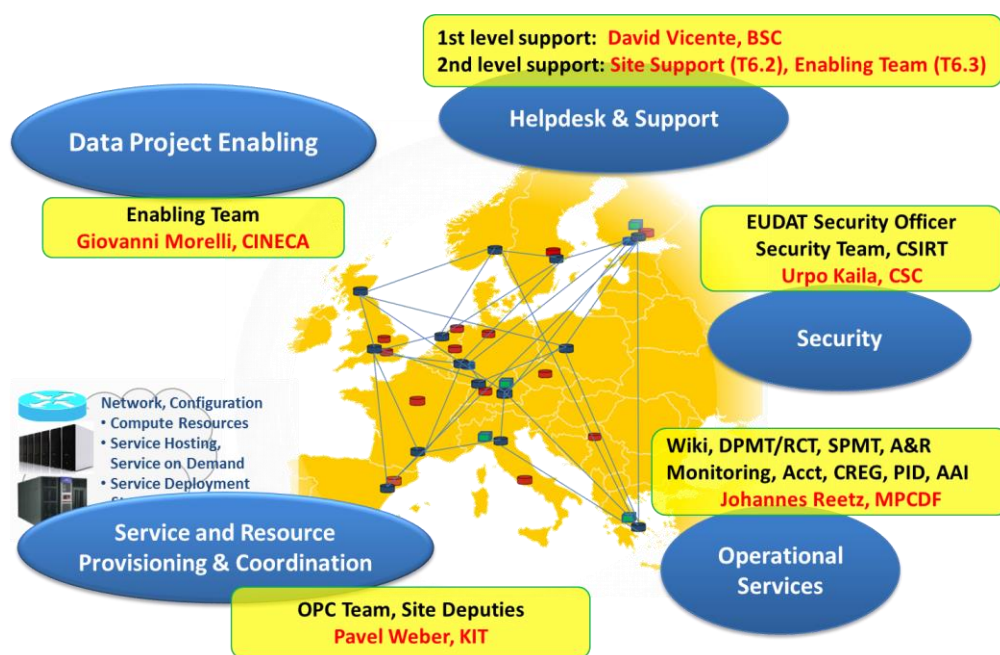


Figure 1: Tasks within the Operations work package

The third task (6.3) *Project and Services Enabling* provides the core service-specific expertise and enabling support required by the research communities – which are regarded as customers – and their users in order to implement and supervise the operation of their data projects within the CDI. Unlike in the initial phase of the EUDAT project, a data project in this phase of EUDAT can comprise, for example, a request to configure and provide a suite of service instances distributed over different service provider sites. The corresponding coordination is undertaken by a data project enabler or a team of data project enablers who are responsible for planning the configuration details with the community or customer that requested the project, in collaboration with the people involved in task 6.2 (especially the site deputies of the CDI service providers). The project enablers work on enabling and supervising the further implementation of the data projects during the enabling phase, while the service integration is strongly supported by service integration experts (who typically come from the customer side, e.g. research data repository experts). The project enablers and the service integrators together share various use case-specific experiences that are documented through

the supporting data project management facilities. These shared experiences also help to give feedback to the service developers in WP5.

The fourth task (6.4) *User Support and Helpdesk* provides and manages the EUDAT helpdesk system, coordinates the support infrastructure for the CDI, and is responsible for providing first level support to customers and users from the EUDAT stakeholder research communities.

The fifth task (6.5) *Security* takes care of operational security, e.g. it defines and supervises compliance with the security and data privacy policies and implements the Computer Security Incident Response Team (CSIRT) team. This task provides consultancy in the areas of security (security by design) and data protection, and it coordinates EUDAT's collaboration on security with other e-Infrastructures (for example, in relation to information exchange and the harmonisation of security policies across different e-Infrastructures including IT-centric research infrastructures).

The document is organised as follows.

Section 2 provides an overview of the service management framework based on the FitSM standard and introduces the associated processes, roles and responsibilities within EUDAT. This section focuses on the description of responsibilities and the tools that are provided, and also demonstrates that almost all the FitSM requirements are already met by the existing EUDAT operational environment.

Section 0 is about the EUDAT central services and operational tools. It describes the concepts of the Data Project Management Tool (DPMT, formerly named Data Project Coordination Portal - DPCP) and the Service Portfolio Management Tool along with their current status, provides news about the EUDAT central site and service registry (that serves as a central information service) and reports on the status of the availability and reliability monitoring system. This section also describes the operational work on the federated identity management service, B2ACCESS, and the B2HANDLE PID service.

Section 0 reports on the work and achievements in the area of resource and service provisioning.

Section 5 describes the conceptual approach used in EUDAT's data project enabling and presents the status of the various data projects and pilots. The data pilots are proposals from research communities and institutional departments that resulted from the first EUDAT call for data pilots, which was conducted by WP4. This section gives a brief example of the "recipes" for the enabling of the B2 services as a first result from the analysis of the requirements of the data projects and pilots.

Section 6 describes the EUDAT Helpdesk and User Support services and provides some statistics from the first year of the project.

Section 7 reports on the activities relating to security and to liaising with other research infrastructures and e-Infrastructures.

Section 8 provides an overview of the work that is planned within the different WP6 tasks.

## 2. THE SERVICE MANAGEMENT FRAMEWORK

The EUDAT Service Management Framework (SMF) describes the principles underlying the policies and structured processes of the EUDAT collaborative data infrastructure, and the implementation of those policies and processes. The primary document about the SMF is an Annex to the EUDAT CDI Collaboration Agreement and is related to the EUDAT deliverable D2.3.2<sup>6</sup> and the Service Quality Assurance Plan (EUDAT D6.4<sup>7</sup>, Annex E). The SMF document is a living document<sup>8</sup>, maintained by WP2 and WP6, which defines the operational constituents, roles and responsibilities of the EUDAT *Service Providers*<sup>9</sup>. The purpose of the document is to ensure that high quality services are delivered to the EUDAT *Customers* and their users.

This chapter is an excerpt from the SMF document and focuses on the operational aspects of the EUDAT CDI. The following sections discuss the main actors within the CDI, and provide an overview of the service portfolio, the central services and the operational tools in EUDAT. The final section in this chapter explains the existing service management processes, roles, and responsibilities.

### 2.1. The Service Management Scope

The SMF comprises guidelines for the collaborative management of all the services provided by and within the EUDAT CDI. It covers **Service Providers** and **Customers** as actors of the CDI, and specifies how the provisioning, configuration, management and support of *Resources* and *Services* for the *Customers* and the *Providers* are coordinated and which tools are being used<sup>10</sup>.

#### 2.1.1. CDI Actors

EUDAT **Service Providers** manage and deliver services to *Customers*, working together under the EUDAT CDI Collaboration Agreement. Organisations can sign the Collaboration Agreement either as *Generic Service Providers* or *Thematic Service Providers*:

- **Generic Service Providers** have regional, organisational or national mandates to support research, usually in a range of different disciplines.
- **Thematic Service Providers** are discipline-specific organisations mandated to support a defined research community or group of customers and users.

**Customers** request the delivery of services, or bundles of services, from the *Service Providers* under specific constraints. A *Customer* may or may not be a direct *User*; if not, the *Customer* purchases the services, or bundles of services, on behalf of the *User* who uses them according to the contractual arrangement with the *Service provider(s)*.

*Users* are any researchers, research groups or collaborative research organisations making use of EUDAT services provided by the Service Providers. *Users* may act on their own behalf, or on behalf of a defined *User Community*.

Both *Users* and *Service Providers* can be **Customers** for a given service or set of services. *Users* are (almost) always external to the EUDAT CDI. *Service Providers* can be *Customers* of other *Service Providers* if, for instance, they rely on certain levels of commitment or performance as part of a federated data management service, or make use of one of the CDI Central Services.

#### 2.1.2. Scope: CDI Common Services

The SMF covers services and resources provided by Service Providers as part of the EUDAT CDI. These services are split into seven classes<sup>11</sup>: data access, data discovery and metadata, persistent identification,

<sup>6</sup> Richard Blake, Community Integration Toolkit, May 2015, <http://hdl.handle.net/11304/252dffa2-f48f-11e4-ac7e-860aa0063d1f>

<sup>7</sup> [https://b2share.eudat.eu/record/216/files/EUDAT-DEL-WP6-D6%204-Final\\_Report\\_on\\_Operations.pdf](https://b2share.eudat.eu/record/216/files/EUDAT-DEL-WP6-D6%204-Final_Report_on_Operations.pdf), 2015, pp.50

<sup>8</sup> under version control: <https://svn.eudat.eu/EUDAT/SMS/trunk/EUDAT-Service-Management-Framework.pdf>

<sup>9</sup> *Service Provider* can e.g. provide storage and compute resources as a service, data as a service, or consultancy as a service.

<sup>10</sup> The EUDAT1 D6.2 Deliverable (<https://www.eudat.eu/d62-1st-year-report-status-and-progress-operations>, pp.49) provided already a systematic and comprehensive set of guidelines. Many of them are still valid.

<sup>11</sup> Compared to the service layers explained in EUDAT2020 WP5 deliverable D5.1, 2016, pp.22

authentication and authorisation, data management, service management infrastructure and storage (see Figure 2). EUDAT's B2 services, which are described in detail in the EUDAT2020 D5.1 deliverable, provide turnkey solutions for many of the requirements that are specified by data projects.

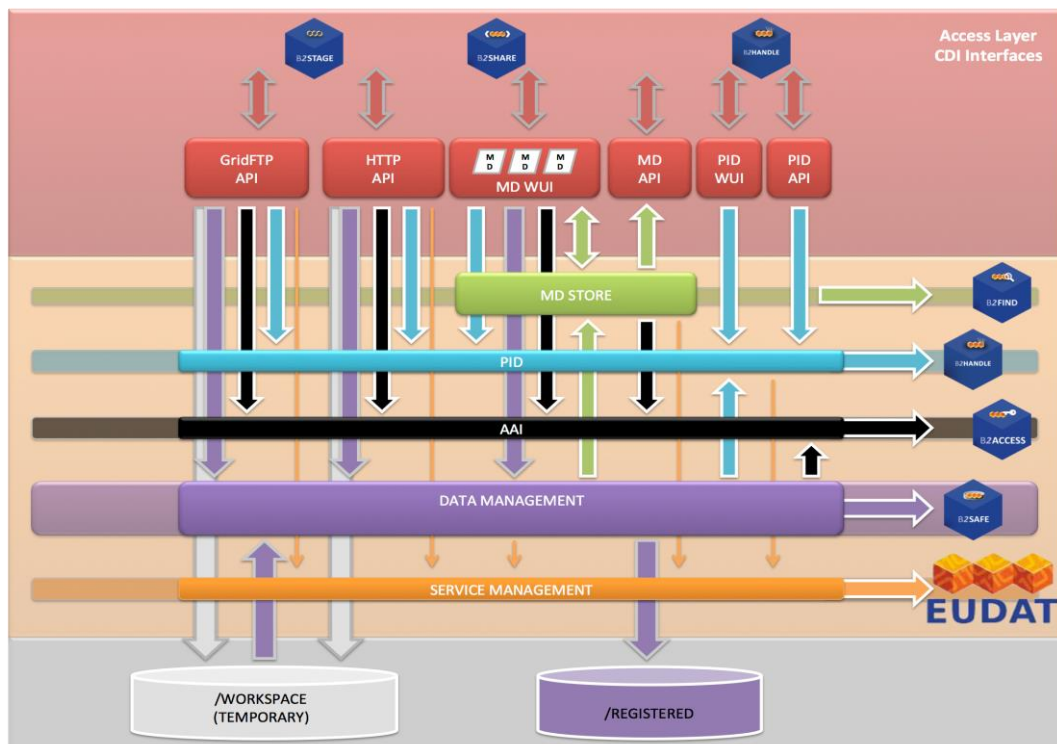


Figure 2: The classes of services within the EUDAT CDI, corresponding B2 services are indicated.

Service Providers can join the CDI Collaboration Agreement as *integrated* nodes or as *interoperable* nodes<sup>12</sup>.

1. **Interoperable nodes** must have a data repository in which they preserve or curate data from a single research community or host data from several research communities or experiments (*storage* and *access* services). Interoperable nodes must identify the data hosted in the repository via some form of persistent identifier (*persistent identification*), and it must be possible to harvest and discover the associated metadata through EUDAT's B2FIND service (*metadata*).
2. **Integrated nodes** are interoperable (as above) and also integrate their local data infrastructure with the CDI's *data management* services; they provide a common *data access* layer, integrate with a common *authentication and authorization* infrastructure (B2ACCESS AAI) and connect their services to the common CDI *service management infrastructure* and operate it according to the EUDAT service management framework.

### 2.1.3. Scope: Central B2 Service Instances

A small number of integrated *Service Providers* operate central EUDAT B2 services on behalf of the rest of the CDI (see Table 1). These services may be deployed as single instances by one *Service Provider* or as multiple instances by one or more *Service Providers*. These services are distinguished from the *Common Services* in Section 2.1.2 in that not every *Service Provider* needs to run service components of them; however, in all cases there must be *at least one* instance of each of these services in the CDI.

The provision of these Central Service Instances is covered by separate agreements.

<sup>12</sup> according to EUDAT2020 D5.1, pp.14

**Table 1: CDI Central B2 Service Instances and the current Service Providers**

Service	Required number	Purpose	Current SPs
B2FIND	1	Central metadata catalogue for the CDI	DKRZ
B2ACCESS	1	Identity and – in future - access control management for the CDI	JUELICH
B2SHARE	1	“Catch all” public data repository	CSC
B2SAFE-DPM	1	Central data policy manager	UiO <sup>13</sup>
B2DROP	1	General-purpose data sync and sharing	JUELICH
B2HANDLE	1+	Handle Prefix registration service and (ePIC) PID services	SURFsara, PDC, DKRZ, GRNET, MPCDF, BSC

## 2.2. Service Management Principles

EUDAT operates as a federation of European research data repositories and Service Providers – these are generally operated independently of each other, but some degree of coordination and some level of central operations are required to make it possible for them to collaborate with each other (which is important so that the data can be shared.)

EUDAT’s principal Customers are European researchers, with a strong emphasis on those participating in European collaborative projects. EUDAT also welcomes the participation of international research projects where European researchers are significant partners.

EUDAT follows a service-centric strategy based on service co-design, in which Customers and Users are involved wherever necessary in all phases of the service life cycle from definition, implementation and configuration to operation.

EUDAT follows a technology agnostic strategy, developing services and operational processes that can be realized across a range of complementary technologies. EUDAT promotes harmonization of service interfaces and underlying processes as a way to maximize technical interoperability.

EUDAT services promote open access to data – and research data in particular – wherever possible. All EUDAT partners are committed to publishing and sharing research data for the greater benefit of European and worldwide research. EUDAT does not assert any rights over any of the data it holds, and all data stored within the CDI retains the original rights. EUDAT promotes the release of data under standard licenses, with the long-term aspiration that all data in the CDI should become fully open access. Access to data in the CDI is free at the point of use.

EUDAT follows a process-based approach to service management. All the required processes are defined, communicated and improved based on Customer needs and feedback from the involved people and parties.

## 2.3. Service Management Tools

The following key tools (Table 2) are used to manage service management procedures across the EUDAT partner network. This is an extensible set of tools. A few of them will be ready in the second half of 2016.

<sup>13</sup> designated service provider



Table 2: Service Management Tools

Tool	Abbrev	Provider	Host	URL
Document Repository	SVN	CSC		<a href="https://svn.eudat.eu/EUDAT/SMS/">https://svn.eudat.eu/EUDAT/SMS/</a> (API, Web UI)
Service Portfolio and Service Level Management Tool	SPMT	GRNET		<a href="http://sp.eudat.eu/">http://sp.eudat.eu/</a> (under development) <a href="http://sp.eudat.eu/ui/portfolio">http://sp.eudat.eu/ui/portfolio</a> (web UI under dev.)
Data Project Management Tool	DPMT	MPCDF		<a href="https://dp.eudat.eu/">https://dp.eudat.eu/</a> (Web UI, API, under dev.) <sup>14</sup>
CDI Configuration Information Service (Site & Service Registry)	CIS	MPCDF		<a href="https://creg.eudat.eu/">https://creg.eudat.eu/</a> (Web UI, GOCDB API, to be replaced) <a href="https://dp.eudat.eu">https://dp.eudat.eu</a> (Web UI, JSON, GOCDB API; dev.)
CDI Accounting Information Service	ACCT	MPCDF		<a href="https://rct.eudat.eu/?path=/storage_space_resources/">https://rct.eudat.eu/?path=/storage_space_resources/</a> (API, Web UI, to be replaced) <a href="https://accounting.eudat.eu">https://accounting.eudat.eu</a> (under dev.)
Availability & Reliability Monitoring Tool	ARMT	GRNET		<a href="https://avail.eudat.eu">https://avail.eudat.eu</a> (Web UI, probes under dev.) <a href="https://cmon.eudat.eu">https://cmon.eudat.eu</a> (Web UI, to be replaced)
EUDAT Website	WEB	Trust-IT	CSC	<a href="https://www.eudat.eu">https://www.eudat.eu</a>
Community Collaboration Platform (CDI external)	COLLAB	Trust-IT	CSC	<a href="https://b2gether.eudat.eu">https://b2gether.eudat.eu</a> (Web UI)
Helpdesk system, Request Tracker	TTS	BSC	CINECA	<a href="https://helpdesk.eudat.eu">https://helpdesk.eudat.eu</a> (Web UI) <a href="https://www.eudat.eu/support-request">https://www.eudat.eu/support-request</a> (Web UI)
EUDAT internal Wiki	WIKI	CSC		<a href="https://confluence.csc.fi/display/EUDAT2/">https://confluence.csc.fi/display/EUDAT2/</a> (Web UI)
EUDAT internal Jira Issue Tracker	JIRA	CSC		<a href="https://jira.csc.fi/browse/EUDATTWO/">https://jira.csc.fi/browse/EUDATTWO/</a> (Web UI)

## 2.4. Service Catalogue

Table 3 lists the services currently offered to EUDAT Customers, plus a number of customer-facing services that are either under development or in even earlier stages of the service definition and production pipeline. The information about these services has been extracted from the current EUDAT Service Portfolio. Deploying service instances and their required constituents – that is, the service components - enables Service Providers to meet the requirements for being either interoperable or integrated nodes within the EUDAT CDI. Detailed descriptions of the Services and their Service Components can be found in the Service Portfolio<sup>15</sup> through the Service Portfolio Management Tool (SPMT).

Table 3: Service Portfolio indicating the contents of the EUDAT Service Catalogue.

Service	Status	Service components <sup>16</sup>
<b>Data Access</b>		
B2STAGE	Active	GridFTP, iRODS-DSI,
CDI Gateway API	under dev't	Web server (Tomcat, Jetty, Apache, nginx)
<b>Data Storage</b>		
B2SHARE	Active	Invenio v2/v3
B2DROP	Active	OwnCloud v8/v9
<b>Data Discovery and Metadata</b>		
B2FIND	Active	CKAN; OAI-PMH Server
Metadata Store	approved	Tbd

<sup>14</sup> In the first project year of EUDAT the Data Projects were registered and documented by using a combination of the RCT, the GOCDB and the EUDAT Wiki. The DPMT will unify the management of data projects, as explained in section 3.1.

<sup>15</sup> The Service Portfolio Management process is explained in the EUDAT2020 D2.1 deliverable (2015)

<sup>16</sup> Version numbers indicate the dominant supported and installed software version.

Service	Status	Service components <sup>16</sup>
Data Type Registry	under dev't	Cordra <sup>17</sup>
<b>Persistent Identification</b>		
B2HANDLE	Active	ePIC v2 & Handle System v8
<b>Data Management</b>		
B2SAFE	Active	iRODS v4; B2SAFE core v2/v3; Data Policy Manager
<b>AAI</b>		
B2ACCESS	Active	Unity <sup>18</sup> v1.6 (replaced by version 1.9 in June 2016)
<b>Other technical services</b>		
B2HOST Service Hosting	Active	various Cloudsystems (Openstack, OpenNebula, VMware)
<b>Non-technical services</b>		
Consultancy service (Data Project enabling)	Active	-
Support Service	Active	
Training service(s)	Active	-

## 2.5. Service Management Processes

The operations coordination, service and resource provisioning, the registration of customer project requests and the related service enabling as well as the user support and the operational security are overall service management tasks. Some tasks, also called *activities*, have to be done on the level of each service provider individually. For instance, software and systems are deployed, installed and maintained by the service providers individually, and each service provider can provide a data management service on the basis of specific technologies and specific storage systems. Other tasks cover most of the service providers of the infrastructure independent from the kinds of provided services. These tasks, which are about delivering CDI services to the customers, are described as service management processes.

In the following, the service management processes across the EUDAT CDI are specified by 14 process definitions according to FitSM-1-2015 v2.0<sup>19</sup>. Not all of these processes are in the responsibility domain of WP6, but most of them are relevant for data projects. The FitSM process and role model is lightweight compared to the ITIL<sup>20</sup> best practices but can be regarded as a first step towards a full ITSM standard such as ISO/IEC 20000-1. The FitSM recommendations have to be adapted to the service management requirements (“best practices”) of the EUDAT CDI which is a network of collaborating independent organisations. Each defined process spans the CDI network and is used in, or has a bearing on, most or all of the CDI services. Each process is the responsibility of a designated Process Manager, and the current organisations filling these roles are noted in the following tables. Detailed information about the current process owners will be provided by the Data Project Management Tool (DPMT)<sup>21</sup>.

The following tables present the short versions of fuller descriptions of processes, their implementation and policies which are given in the document *EUDAT Service Management Process Definitions*<sup>22</sup>.

The management of each of the processes requires different tools to be used by different actors. These tools are listed in Table 2 on p.14. Some tools indicated in Table 2 are under development (e.g. SPMT, DPMT, and ACCT) and will be introduced during 2016. The following tables refer to these tools. Each table is denoted by the FitSM process number (PR#).

<sup>17</sup> <https://www.cordra.org/>

<sup>18</sup> <http://www.unity-idm.eu/>

<sup>19</sup> [http://fitsm.itemo.org/sites/default/files/FitSM-1\\_Requirements.pdf](http://fitsm.itemo.org/sites/default/files/FitSM-1_Requirements.pdf)

<sup>20</sup> The IT Infrastructure Library; see <https://www.axelos.com/best-practice-solutions/itil/what-is-itil>

<sup>21</sup> This is motivated by the fact that the preparation, implementation and operation of data projects require all the FitSM processes.

<sup>22</sup> <https://svn.eudat.eu/EUDAT/SMS/trunk/EUDAT-Service-Management-Process-Definitions.docx>



A number of the larger, more complex processes are broken down into sets of *activities*. While processes have a single EUDAT Process Manager assigned to their overall coordination, individual activities are carried out by the Service Providers under the responsibility of the CDI Site Managers. A summary of the roles and responsibilities are provided in Annex B.

<b>PR1</b>	<b>Service Portfolio Management</b>
Responsible:	Service Portfolio Manager (CINES)
Purpose:	Maintain the EUDAT CDI Service Portfolio
Tools:	SPMT
Outputs:	EUDAT Service Portfolio

<b>PR2</b>	<b>Service Level Management</b>
Responsible:	Service Level Manager (CINES)
Purpose:	Define and maintain the CDI Service Catalogue and agreed service levels between Users and Providers (SLAs), and between Providers and Providers (OLAs) within the CDI.
Tools:	SPMT
Outputs:	EUDAT Service Catalogue; Service Level Agreements; Operation Level Agreements.

<b>PR3</b>	<b>Service Monitoring &amp; Report Management</b>
Responsible:	EUDAT Service Reporting Manager (GRNET)
Purpose:	Define and maintain service report templates and reporting mechanisms for each CDI service. Create and provide service reports.
Tools:	SPMT, DPMT, ARMT, ACCT, CIS
Outputs:	EUDAT Service Reports.
<b>Activity:</b>	<b>Availability &amp; reliability monitoring and reporting</b>
Service Provider	Monitor and report about the availability and reliability of the Services and local components of the Service Suites provided by EUDAT.
<b>Activity:</b>	<b>Resource &amp; data usage accounting and reporting</b>
Service Provider	Provide a timely overview about of the number and volume of the digital assets per data project and per registered resource that are managed within the CDI.

<b>PR4</b>	<b>Service Availability &amp; Continuity Management</b>
Responsible:	Availability & Continuity Manager (MPCDF, KIT)
Purpose:	Ensure the availability and continuity of the Service Instances and Service Component instances for the B2 services and the Service Suites enabled and deployed through the Data Projects, and for the support services like B2ACCESS or B2HANDLE on which other Services depend.
Tools:	DPMT
Outputs:	Service availability and continuity plans
<b>Activity:</b>	<b>Registration and management of service instances</b>
Service Provider	Register the availability of service instances, service components deployed and operational at service providers, and under what conditions they are available. Analyse risk factors and support the risk mitigation via disaster recovery plans. If a service provider intends to leave the CDI or if a provider intends to terminate the provisioning

	of a service, a service component or a resource, the migration of the service must be planned; at least the grace period must be long enough for the customers to react.
--	--

<b>PR5</b>	<b>Capacity Management</b>
Responsible:	Capacity Manager (KIT)
Purpose:	Manage the capacity of provided <i>Resources</i> and <i>Service Components</i> necessary to fulfil the agreed customer capacity and performance requirements. Ensure that sufficient capacities are pledged and allocated by the Resource and Service Providers.
Tools:	DPMT, ACCT
Outputs:	Capacity plans; capacity and performance monitoring plans and reports
<b>Activity:</b>	<b>Registration and capacity management of resources and service components</b>
Service Provider	Register pledged resources and offered service components. Provide and allocate resources and service components related to services which are provided for specific data projects.

<b>PR6</b>	<b>Information Security Management</b>
Responsible:	Security Manager (CSC)
Purpose:	Analyse and manage information security and privacy risks for the CDI. Maintain the security and privacy policies. Ensure e-Infrastructure security by conducting vulnerability scans. Share information with site security officers from other CDI nodes. Advise service developers about security and privacy by design. Coordinate the CSIRT and ensure the efficient flow of information between security officers if incidents or vulnerabilities are detected.
Tools:	DPMT, CIS, Wiki, Jira
Outputs:	EUDAT Information Security Policies; Information security Incident Reports; reports from security vulnerability scans
<b>Activity:</b>	Local information security management
Service Provider	Respond to and act on local security and privacy issues. Respond to and act on security and privacy advisories from the Security Manager.

<b>PR7</b>	<b>Customer Relationship Management</b>
Responsible:	Customer Relationship Manager (CSC)
Purpose:	Establish and manage the relation between CDI Service Providers and Customers; support in the preparation of Data Project requests from the customers
Tools:	DPMT, Websites
Outputs:	Up-to-date Customer database; Customer service review reports; Customer-oriented website.

<b>PR8</b>	<b>Supplier Relationship Management</b>
Responsible:	Supplier Relationship Manager (MPCDF)
Purpose:	Manage the relation between Service Providers and Resource Providers, Content Providers as well as Technology Providers.
Tools:	DPMT
Outputs:	Up-to-date supplier database.

<b>PR9</b>	<b>Incident and Service Request Management (service desk and 1<sup>st</sup> level support)</b>
Responsible:	Incident and Service Request Manager (BSC)
Purpose:	Provide a trouble ticketing system (TTS) and the 1 <sup>st</sup> level support. Organise the TTS to be useful as a service desk for the 1 <sup>st</sup> and 2 <sup>nd</sup> level support. Create and maintain a knowledge base of frequently asked/answered questions (FAQs) and on answers. Act as bridge to the 2 <sup>nd</sup> level support and clarify incoming requests if they are too vague. Facilitate the collaboration between the CDI Helpdesk and the user-facing supports lines of the Service Providers as well as of the Customers.
Tools:	TTS
Outputs:	Regular service request reports; Major incident review reports.
<b>Activity:</b>	Local incident and service request management
Service Provider	Respond to local (i.e. site-specific) 1 <sup>st</sup> level queries.
Customer	Respond to community user requests and forward issues if necessary

<b>PR10</b>	<b>Problem Management (2<sup>nd</sup> level support)</b>
Responsible:	Problem Manager (MPCDF, KIT)
Purpose:	Detailed problem investigation; solution creation particularly for issues that happen in the context of data projects. If necessary, communicate with service developers and technicians.
Tools:	TTS, Website, Wiki, Jira
Outputs:	CDI Known Error Database (KEDB) with information on problems, known errors and related workarounds and solutions. Feature requests or problem solutions requests to the service developers.
<b>Activity:</b>	Data project and service-related as well as local problem management
Service Provider	Respond to site-specific 2 <sup>nd</sup> level queries. Contribute solutions to KEDB.
Service Owner	Respond to service-specific 2 <sup>nd</sup> level queries. Contribute solutions to KEDB and FAQ.

<b>PR11</b>	<b>Configuration Management</b>
Responsible:	Configuration Manager (KIT, CINECA)
Purpose:	Provide and maintain a logical model of all configuration items and their relationships and dependencies with particular regard to the operated and planned data projects.
Tools:	DPMT, SVN, CIS
Outputs:	Up-to-date Configuration management database.
<b>Activity:</b>	Data project, service-related and resource related configuration management
Service Provider	Manage the correct configuration information about the service component instances and resources.

<b>PR12</b>	<b>Change Management</b>
Responsible:	Change Manager (KIT, MPCDF)
Purpose:	Ensure that changes to configuration items are planned, approved, implemented and reviewed in a controlled manner to avoid adverse impact of changes to services or the customers receiving services.
Tools:	Jira, Wiki ( <a href="https://confluence.csc.fi/display/EUDAT2/Change+Management">https://confluence.csc.fi/display/EUDAT2/Change+Management</a> )
Outputs:	Planned & documented changes
<b>Activity:</b>	<b>Change request management</b>
Service Provider	Create and manage change requests related to hosted service components on which other services may depend. Identify risks that the change may have for the locally provided services.

<b>PR13</b>	<b>Release and Deployment Management</b>
Responsible:	Release Manager (SURFsara, MPCDF)
Purpose:	Bundle changes of one or more configuration items into releases, so that these changes can be tested and deployed to the live environment together. Mark those releases which are unsupported. Define new releases and mark them as production-ready. Provide information about the releases via the SPMT. Prepare a release plan and deployment plan. The Service Providers are notified about the availability of new releases if the currently installed software becomes unsupported.
Tools:	SPMT, DPMT
Outputs:	Defined and successfully deployed releases.
<b>Activity:</b>	<b>Monitor release status and organise software deployments and updates</b>
Service Provider	Act on notifications (via the DPMT) of new releases of service software.

<b>PR14</b>	<b>Continual Service Improvement Management (CSI)</b>
Responsible:	CSI Manager (SURFsara)
Purpose:	Identify opportunities for improvement and register corresponding proposals. Evaluate and approve proposals for improving services and the overall customer satisfaction in a consistent manner.
Tools:	Jira, Wiki ( <a href="https://confluence.csc.fi/display/EUDAT2/Service+Building+Requirements">https://confluence.csc.fi/display/EUDAT2/Service+Building+Requirements</a> )
Outputs:	Innovations, maintaining competitiveness, new customers, improved customer satisfaction

### 3. CENTRAL SERVICES AND OPERATIONAL TOOLS

This chapter presents the concepts and design<sup>23</sup>, set-up and management of some of the EUDAT *central services and operational tools* which are *essential* for implementing the service management processes discussed in section 2.5. As outlined in section 2.3, many of these services are service management tools used for coordinating, running and sustaining the EUDAT CDI. They comprise the *Service Portfolio Management Tool* (SPMT) and the *Data Project Management Tool* (DPMT) – both of which are developed in the current phase of the EUDAT project, along with the *Availability and Reliability Monitoring* system<sup>24</sup>, the *CDI Configuration Information Service* (which is currently the site and service registry<sup>25</sup> based on a GOCDB instance) and the *Accounting Service*. The latter two services were introduced during the previous phase of the EUDAT project. All of these services are part of the EUDAT service management infrastructure<sup>26</sup> (see Figure 2) and support the coordination and monitoring of the provisioning and enabling of services and the service management activities.

Further essential services for the EUDAT CDI are the central AAI service, B2ACCESS<sup>27</sup>, and the PID service, B2HANDLE<sup>28</sup>, which is a distributed network of Handle services within the CDI. The Data Policy Manager<sup>29</sup> is another essential central CDI service, but it was not in operation during the first year of this phase of the project. Each of these services is mapped to its specific management layer of the EUDAT CDI as indicated in Figure 2.

#### 3.1. Data Project Management

A *Data Project* in the context of EUDAT is a planned set of related tasks in the domain of data management and stewardship to be executed over a defined period of time and under specific constraints (such as the cost of the project). The EUDAT data projects generally relate to tasks that the research communities need to undertake to manage their research data. Such projects can be broken down into manageable units (which are known as subprojects). Table 4 lists the data projects that started during the first year of this phase of EUDAT. The EUDAT services that are involved in each project are indicated, along with the EUDAT consultant who is responsible for each project or subproject and the general CDI provider that has the overall responsibility for the distributed service that is involved in the project. The project IDs indicate subprojects, each of which is dedicated to enabling a specific service instance. The project status information is from spring 2016.

More specifically, a EUDAT data project is regarded as the set of tasks necessary for planning, configuring, providing, managing, maintaining, monitoring and accounting for the CDI services and service components which are relevant for a particular customer, and also for executing the requisite workflows. The time span of each project must be defined in advance, but can be prolonged subsequently – this means that a data project can last as long as it is sponsored or funded by a customer. A data project can be modified and it can be moved from one customer to another. This allows projects to be switched from *data management* to *data stewardship* with potentially different project owners and actors as needed.

WP6 has been developing the *Data Project Management Tool*<sup>30</sup>. The design of the data model and the implementation of this tool are based on experience gained during the previous phase of the EUDAT project and the deliverable about Service Management Processes from WP2. Additional requirements from WP2, WP4, WP5, WP7 and the EUDAT Technical Committee were taken into account while the tool was being

<sup>23</sup> A few tools (DPMT, SPMT) have been designed and implemented by WP6.

<sup>24</sup> The advanced and integrated accounting system is under development. Currently the storage usage accounting records are collected and displayed as described in the [EUDAT1 D6.4 deliverable](#), p.16.

<sup>25</sup> [EUDAT WP6 deliverable D6.4](#), 2015, section 2.2, <http://hdl.handle.net/11304/132bacae-f8a5-11e4-8a18-f31aa6f4d448>

<sup>26</sup> EUDAT2020 WP5 deliverable D5.1, 2016, section 3.3.2.3

<sup>27</sup> EUDAT2020 WP5 deliverable D5.1, 2016, section 4.6

<sup>28</sup> EUDAT2020 WP5 deliverable D5.1, 2016, section 5.2

<sup>29</sup> EUDAT2020 WP5 deliverable D5.1, 2016, section 5.1.2

<sup>30</sup> also called *Data Project Coordination Portal (DPCP)*

developed and it was made available as early as possible so that it could be widely tested. A pre-release version of the DPMT has been available at <https://dp.eudat.eu> since January 2016.

WP6 is also developing the *Service Portfolio Management Tool* (which is described in section 3.2). In the course of planning and enabling a data project, the kinds of services that are available need to be selected from a Service Catalogue or, if a new service has to be developed, the Service Portfolio needs to be monitored. *The Service Portfolio Management Tool (SPMT)* makes it possible to manage descriptions about the *kinds of services* that are available, along with information about the development status and some implementation details about the necessary service components. The DPMT provides an inventory of *Data Projects*, which use concrete *customized suites of distributed service instances*, and lets the inventory be managed.

**Table 4: Data Projects from the first EUDAT2020 project year<sup>31</sup>**

Project / Pilot Name	Customer Community / Institution	(Sub)Project Start Date	Project ID	Services	Enabler	Status	CDI General Provider
UP-EPOS-INGV	INGV	2015-04	60.1	B2SAFE	M. Fares	pre-prod	CINECA
UP-EPOS-KNMI	KNMI	2015-04	61.1	B2SAFE	M. Fares	pre-prod	SURFsara
UP-EPOS-GFZ	GFZ	2015-04	62.1	B2SAFE	M. Fares	pre-prod	KIT
UP-LTER	LTER	2015-04	48.1	B2SAFE	C-J. Håkansson	enabling	Juelich
		2015-04	48.2	B2SHARE	C-J. Håkansson	enabling	Juelich
		2015-04	48.3	B2FIND	H. Widmann	enabling	DKRZ
DEIMS	LTER	2015-10	49.1	B2FIND	H. Widmann	enabling	DKRZ
UP-CLARIN-SOAS	SOAS / CLRAIN	2015-10	52.1	B2SAFE	W. Elbers	enabling	MPCDF
IST DataRep	IST Research Center	2016-01	1001.1	B2SAFE	E. Erastova	pre-prod	KIT
		2016-01	1001.2	B2FIND	H. Widmann	pre-prod	DKRZ
Herbadrop	Digitalium of the Finnish M	2016-01	1002.1	B2DROP	P. Dugenie	planning	CINES
		2016-01	1002.2	B2SHARE	P. Dugenie	planning	CINES
		2016-01	1002.3	B2SAFE	P. Dugenie	pre-prod	CINES
		2016-01	1002.4	B2STAGE	P. Dugenie	planning	CINES
		2016-01	1002.5	B2FIND	H. Widmann	planning	DKRZ
		2016-01	1002.6	dev:BigD	P. Dugenie	dev	CINES
		2016-01	1002.7	dev:DataT	P. Dugenie	dev	CINES
Aalto	Aalto University	2016-01	1002.8	dev:Sem	P. Dugenie	dev	CINES
		2016-01	1003.1	B2DROP	J. Heikinen	enabling	CSC
		2016-01	1003.2	B2SHARE	J. Heikinen	enabling	CSC
		2016-01	1003.3	B2SAFE	J. Heikinen	planning	CSC
		2016-01	1003.4	B2FIND	H. Widmann	planning	DKRZ
		2016-01	1003.5	dev:BigD	J. Heikinen	dev	CSC
WebCoverage	FZ Juelich	2016-01	1003.6	dev:DataT	J. Heikinen	dev	CSC
		2016-01	1005.1	B2SHARE	S. Apweiler	enabling	JUELICH
SIMCODE	Alma Mata University Bolo	2016-01	1005.2	B2FIND	H. Widmann	planning	DKRZ
		2016-01	1006.1	B2STAGE	G. Morelli	prod	CINECA
		2016-01	1006.2	B2SHARE	G. Morelli	planning	CINECA
UP-ICOS	ICOS	2016-01	1006.3	B2FIND	G. Morelli	planning	CINECA
		2016-01	50.1	B2SHARE, B2ACCE	M. Hellstrom	planning	CSC, Juelich
UP-ENES	DKRZ / ENES	2016-01	51.1	B2FIND B2SHARE	H. Thiemann	planning	DKRZ
NFFA	CNR-IOM	2016-02	1004.1	B2SHARE, B2FIND	H. v. Piggelen	planning	STFC
Westlife	STFC	2016-02	1007.1	B2DROP, B2SAFE, B	H. v. Piggelen	planning	STFC
FAIR	Dutch Techcentre for life sc	2016-02	1022.1	B2SHARE, B2SAFE,	Chr. Staiger	planning	SURFsara

Each *Data Project* has a *project lifecycle* which consists of several phases that need to be managed. As explained in section 3.1.4 and indicated in Figure 6, from the *project request phase* and the more comprehensive *planning phase* to the *enabling* and *production phases*, the description of the project becomes increasingly structured to the point where the project description can be mainly represented by a

<sup>31</sup> Colour code: The colours in the column “Services” indicate the different B2 services being planned to be used, being enabled or in another phase. Projects that depend on a service that still needs to be developed (in WP5) have a mixed colour from red to orange. Colours are used in the column “Status” to highlight those services (subprojects) which are either *enabled* or in *(pre-)production*.

structured (machine-readable) document that carries information about the topology of the related service suite which is used in the project, the configuration and workflows of the enabled and connected service instances, the service providers, the service component instances, the associated allocated storage and compute resources, along with the actors involved in the project and their roles.

When the enabling of a project is accomplished, it needs to be commissioned during a pre-production phase. The project description is a machine-readable structure that specifies how the relevant data is acquired, ingested into the CDI, processed, described, stored, registered, replicated (or distributed), transferred, indexed, and made discoverable (findable), accessible and auditable. The capabilities and capacities of the services and resources, along with the service level agreement and the envisioned data access pattern should also be specified, as well as the time span over which the data project has to be operated until it is decommissioned. These design principles are also motivated by the Research Data Alliance (RDA) Data Fabric Interest Group<sup>32,33</sup>.

For each enabled project, the topology of interacting services are defined, as well as the resources that are to be accounted for and monitored. Within this domain of services it is possible to apply data policies, for instance, using the B2SAFE *Data Policy Manager*<sup>34</sup> that uses the configuration information from the DPMT.

### 3.1.1. The DPMT between Service Management and Data Management Planning

The DPMT serves as an authoritative source of information about all the relevant service providers and service instances which are involved in an enabled project. It helps organise the instances of services within the EUDAT CDI which are – explicitly or implicitly – required by the customer to either mediate EUDAT services or to offer value added services on top of them to their user communities. The DPMT facilitates the continuation of this service delivery, makes it possible to coordinate the provision of capabilities and capacities, and is a platform for customers and service providers to keep in contact concerning specific projects.

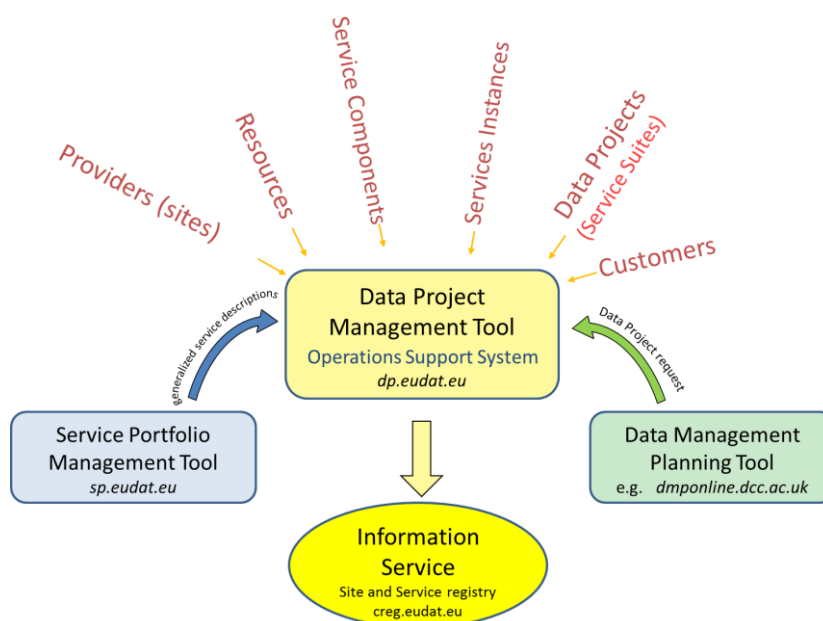


Figure 3: The DPMT connected to the SPMT, the DMP Tool and the Configuration Information Service

*Data projects* are implemented on the basis of available services in the form of service packages. The information about services being developed, as well as about those which are production-ready, is managed using the Service Portfolio Management Tool (section 3.2), which also includes a catalogue of services that

<sup>32</sup> <https://rd-alliance.org/group/data-fabric-ig.html>

<sup>33</sup> RDA DF IG: *Data Management Trends, Principles and Components*, 2015, [hdl:11304/f638f422-f619-11e4-ac7e-860aa0063d1f](https://doi.org/10.11304/f638f422-f619-11e4-ac7e-860aa0063d1f)

<sup>34</sup> EUDAT2020 WP5 deliverable D5.1, 2016, section 5.1.2



have production-ready instances. The Data Project Management Tool registers the *Service Providers* and their offers, as well as the *Customers* and their *Data Project Requests*. The Service Catalogue (via an interface to the SPMT) is used by the *Service Providers* to identify and select the kind of services and service components that they can offer. *Customers*, supported by CDI consultants (project enablers), can identify and select those services which appear to be necessary for implementing their *Data Project*.

The Data Projects largely influence how the EUDAT B2 services are configured and integrated, and thereby determine the topology of the CDI. The sustainability of the CDI, at least parts of the CDI, will depend on the sustainability of the data projects and how long they are going to be sponsored.

Figure 3 depicts the flow of information between the different tools: the Data Project Management Tool obtains descriptions of the kinds of available services and their service components (blue arrow) which are to be instantiated and configured specifically during the implementation of the project. On the other side, Data Project Requests can be based on Data Management Plans which are prepared and submitted via a Data Management Planning Tool that provides appropriate questionnaires to be filled by the project PI. Authoritative information that is captured via the SPMT, DMP tool or by the DPMT itself is made available via an information service in front of the DPMT. This Configuration Information Service that receives information from the DPMT can provide to clients<sup>35</sup> any information about the assets and configuration of the CDI.

As long as the DPMT is under development, the GOCDB has been used as *Site and Service Registry* – a registry that is not just providing configuration information but that was used also for registering and managing the information about service providers (sites), services and service groups. The DPMT is a replacement of the GOCDB as authoritative registry and the configuration information can also be made available from the DPMT via its own API (based on JSON) and an additional API that is compatible with the API of the GOCDB.

The combination of DMP tool, DPMT, SPMT and the CDI Information Service can to a certain extent be considered as a Enterprise Resource Planning tool and a Enterprise Information Management system.

### 3.1.2. Data Project Use Cases and Workflows

The Data Project (and Service) Management Tool has been designed to support the current use cases and workflows. These use cases and workflows are presented in Annex C.1 (Table 11) and comprise specific and general requirements which resulted in the data models that are presented in section 3.1.3 (Figure 4 and Figure 5).

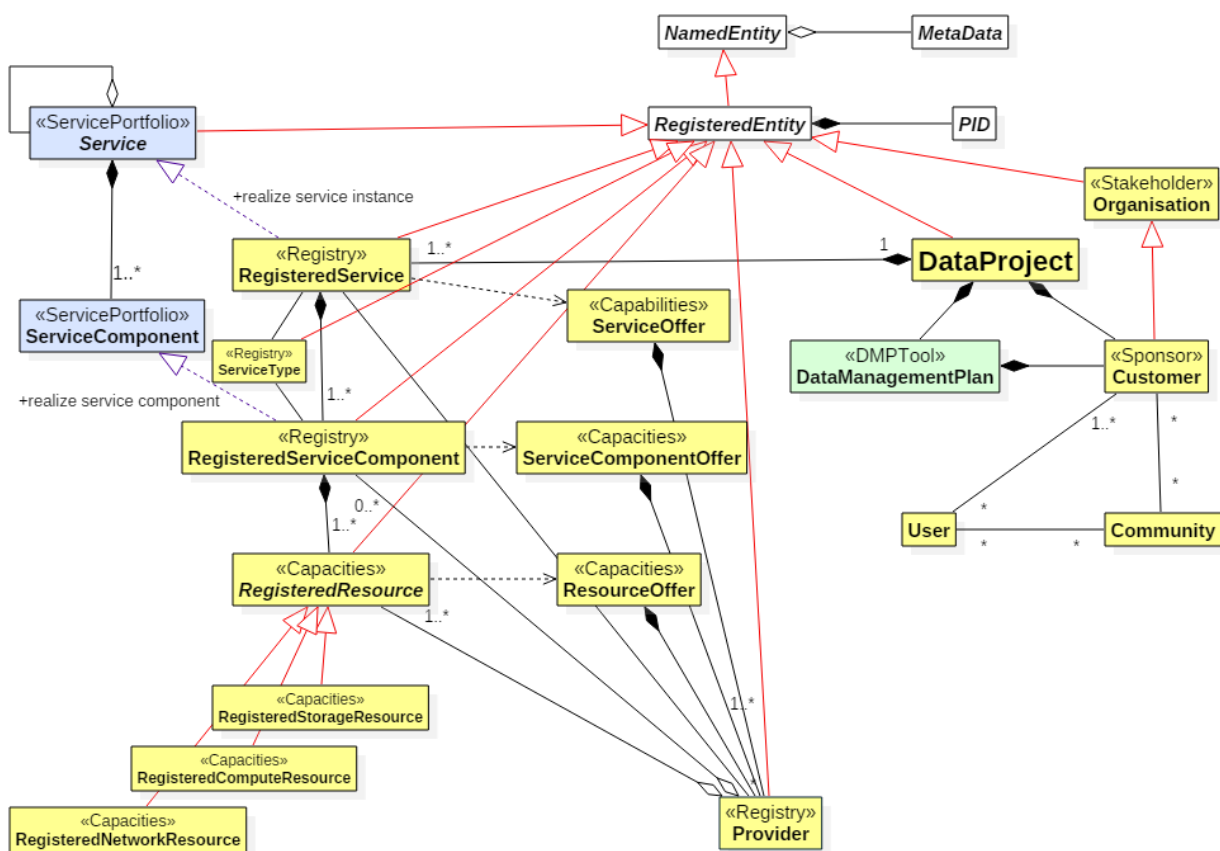
In summary, the use cases relate to the registration and coordination *of* and the information *about* the

1. (Service) *Providers*
2. *Service Offers* by General Service *Providers*
3. *Service Component Offers* by Service *Providers*  
(e.g. offers to setup and run new iRODS instances for specific kinds of B2SAFE projects)
4. *Resource Offers* by Resource *Providers*  
(e.g. storage, network, service hosting, computing)
5. *Customers* as principal contacts of *Data Projects*
6. *Data Project Requests* by *Customers*  
(e.g. a data management plan submitted via a Data Management Planning Tool)
7. *Data Projects* after the project requests have been approved  
(project structure, project description, work plan and status) during the Project Life Cycle.
8. Service Instances (*RegisteredService* in Figure 4)
9. Service Component Instances (*RegisteredServiceComponent* in Figure 4)
10. Resources (*RegisteredResource* in Figure 4)

<sup>35</sup> Clients like the Data Policy Manager, the monitoring and accounting systems, vulnerability scanner.

Other use cases and requirements:

- 11.Resource Capacity Management
- 12.API access for clients (applications, services, users) to information of the DPMT.
- 13.DPMT to be integrated with B2ACCESS, access rights to be managed by the DPMT.
- 14.Delegation of Roles and Access management
  - a Customer principal can delegate responsibilities and access rights to another customer contact;
  - a Service Provider principal can delegate responsibilities and access rights to team colleagues
- 15.Information can be imported from the Service Portfolio; the information can be displayed.
- 16.Display the Service Catalog (including version numbers of active packages and service component).
- 17.Compare the version numbers of actually deployed service instances (service packages), service component instances and notify the right contacts about new software releases if the difference between the actually installed and the available latest release is detected to be over a threshold.
- 18.Incorporate service component implementation details (attribute names) into the service component information classes



**Figure 4: High level schematics (UML) of the DPMT Data Model.**

### 3.1.3. The DPMT Data Model

The DPMT is designed to support the use cases listed in the previous section and Table 11 in the Annex. Figure 4 depicts the essential information classes from the DPMT data model. Many of these information classes are also mentioned in section 3.1.2 and indicated by using italic letters. For instance, the diagram indicates that (B2) services, service components and resources can be offered by the providers and that every valid offer must have a provider (there is no valid offer without a committed provider). CDI providers can only offer services and service components which are described in the service catalogue, which are the descriptions of active services from the service portfolio (blue classes related to Figure 5). These offers need to be specified in such a way that project manager (in agreement with the customers) can make a decision

during the planning phase<sup>36</sup> of a project about which offer should be taken into account. The providers are notified when his offer is intended to be taken within a project, at this stage a provider can retract his offer. This approach should allow making the enabling phase short and most efficient, so that the offered services, service components and resources can then be instantiated (eventually deployed if necessary and configured) and registered with the DPMT.

One of the principles for the design of the DPMT is that the information objects have unique identifiers registered by the DPMT and that the tool makes it possible to create - via B2HANDLE - *persistent identifiers* (PIDs) and assign them particularly to those named entities that are going to be used either across administrative boundaries or that will exist for a long time span. This assignment of PIDs to information objects facilitates the exchange of information objects between different tools and different organisations. By this way these persistent references can be used in other catalogues of either service descriptions or concrete service instances. This is one building block for a trustworthy and stable data fabric<sup>37</sup>. As another example, the DPMT assigns persistent identifiers to data project, sub-project, services and resources which can be used in different accounting systems.

An essential element in Figure 4 is made up of the parts labelled *Service* and *ServiceComponents* (light blue coloured). These parts point to the service descriptions provided by the SPMT that Figure 5 displays in more detail. This schematic below shows the relationships between the EUDAT B2 services and a selection of components in more detail. This model is also used by the SPMT for describing the component implementation details (compare with Figure 10).

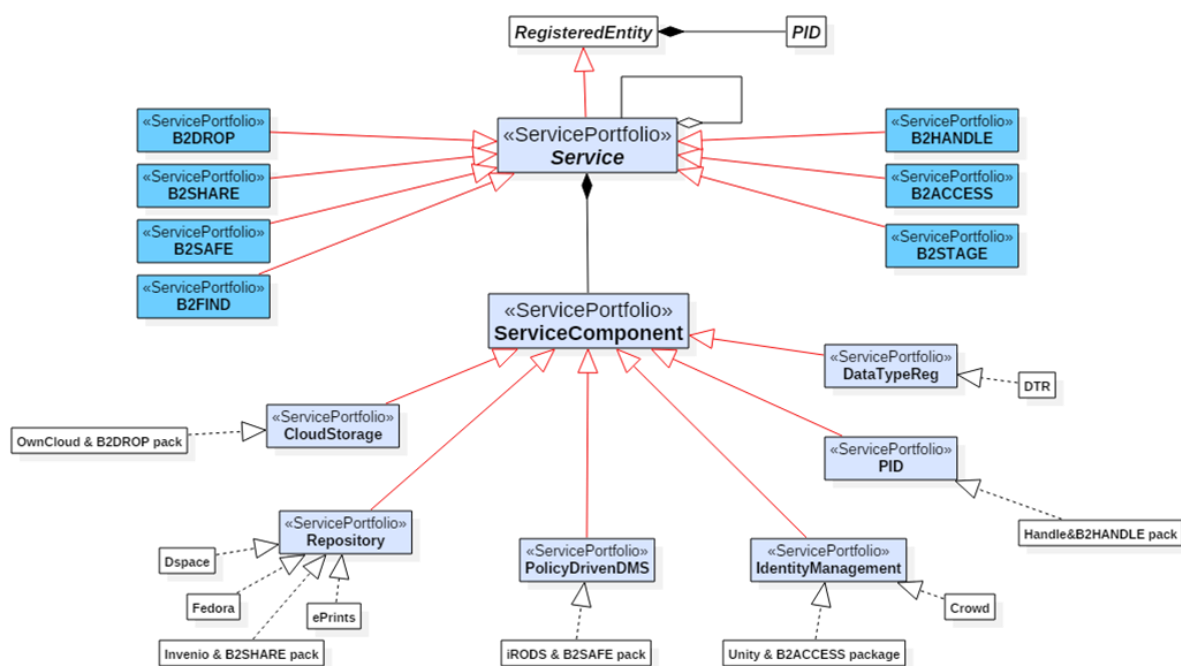


Figure 5: Distributed Services as Compositions of Service Components

Each of the information objects has a specific schema which represents the configurable properties of the corresponding service component. Both UML diagrams depict high level concepts which neither display the schemas of the classes nor the possible states and transactions. The DPMT model (Figure 4) has some classes that have different states, and each state implies a different schema. This is particularly true for the Data Project for which the different states are discussed in the following section. But also the registered services

<sup>36</sup> See section 3.1.4

<sup>37</sup> See footnote 33, p.23.

and service components, these are the instances created or reused in during the project enabling phase, have a service management life cycle and the service management will be supported by the DPMT.

### 3.1.4. Data Project Phases

As mentioned earlier, the Data Project life cycle begins with a project plan (which is usually a data project request) that can contain a high-level description of the requirements for data management and data stewardship, and it may end with the termination of the project, provided the project has not been retracted or replaced by another project before the end of its lifetime. Figure 6 shows the different stages which are described in Table 12 (Annex C.2) in more detail. It is important to note that there are no cyclic transitions, e.g. between the planning phase and the enabling phase. If a planned project is decided to be enabled, a larger amount of resources can be mobilized in order to enable a data project, e.g. depending on the size of the project procurement processes can be initiated at the different service provider sites which cannot be easily stopped without extra costs. For this reason a customer cannot easily retract a project request once the project has been planned and is now going to be enabled. However, it is possible to suspend the execution of projects for a certain amount of time.

Data projects also can comprise multiple services and service instances.<sup>38</sup> This may require a breakdown of a data project into several manageable subprojects – ideally one per service class. Each subproject then can be enabled by a different consultant for a dedicated life cycle, work plan and time line. The project phases depicted by the following diagram may apply also to subprojects.

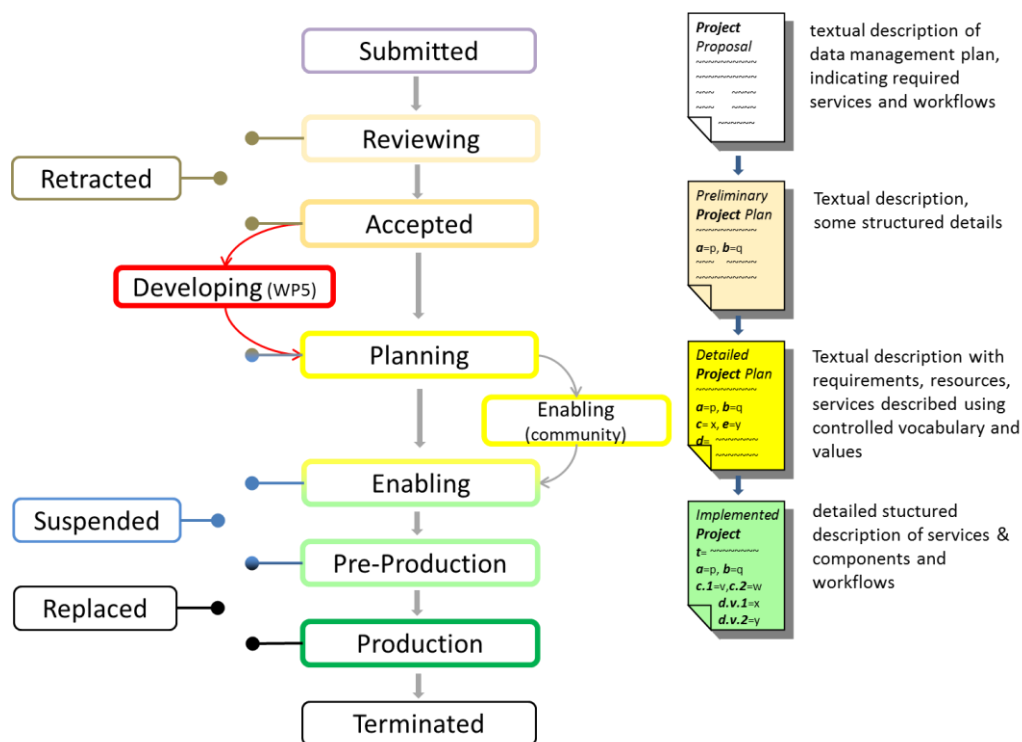


Figure 6: Data Project Phases

If a project splits into subprojects the overall project status *Pre-Production* is reached only if all sub-projects are at least on this level. When the project (or sub-project) is *production* any further modification requests are managed via procedures such as capacity management, configuration management, change management, rollout and deployment management procedures that are discussed in section 2.5.

<sup>38</sup> E.g. a customer orders multiple B2DROP instances to be provided for his community: one B2SHARE instance and several repositories are to be connected with B2SAFE. In addition the metadata from each of the community repositories shall be harvested via OAI-PMH connectors. These different services can most efficiently be enabled by concurrent processes managed by different consultants.

### 3.1.5. Actors and Roles for Data Project and Service Management

The following actors and roles specifically concerning the Data Project Management are required. These roles are also mentioned section 2.5 about the service management and they are indicated in the section and summary table (Table 11) about the DPMT use cases and workflows.

These roles are the

- CDI manager,
- Customer: the customer principal,
- Customer Admin (administrative people with restricted rights; can act on behalf of the customer),
- Provider: the provider principal (can be the site deputy mentioned in section 4.1),
- Provider Admin (administrative people with restricted rights; can act on behalf of the provider),
- Project Manager, who is responsible for planning, organising and supervising a specific data project,
- Project Enabler, the technical expert with regard to specific B2 services who interacts with the staff members from the customer as well as the provider sides in order to ensure that all the necessary components are available and are properly enabled (configured, tested) for the (sub-)project.
- Project Request Reviewer, who reviews the data project requests and who interacts with the principal investigator who submitted the data project request. The purpose of the review is to support customers to bring their data project request to an acceptable level before the request is approved or not. There should be a data project review board which is periodically reviewing incoming project requests.
- Customer Relationship Manager, who is able to organise the data project approval process and who used the DPMT to ensure a good customer relationship by providing information to the customer and by managing the customer's expectations, especially if it comes to the development of new services or new features for existing services as in case of the data pilots.

### 3.1.6. DPMT Implementation Status

The existing prototype of the DPMT already supports several of the use-cases and workflows that are described in the previous sections. Content from the GOCDB and the RCT have been imported and several workflows and overview panels have been defined. A few of them are shown in the following figures. The At the time of writing, the DPMT is not in production and therefore still the GOCDB and RCT are used.

Figure 7 shows the *Data Projects* overview panel.

Title	Service	Allocated storage	Used storage	Customer	Topics	Start date	State
B2STAGE+B2SAFE for VPH	B2SAFE, B2STAGE	5 TB	610 GB	VPH	Bio Med, Life Science	2014/05/10	production
B2SAFE for TLA	B2SAFE	100 TB	98 TB	CLARIN Center TLA (MPI PL)	Linguistic, Humanities	2012/12/01	production
B2SAFE for EKUT	B2SAFE	200 GB	68 GB	CLARIN Center Tuebingen (EKUT)	Language Resources and Technology Infrastructure, ...	2013/04/01	production
B2SAFE for EPOS PP WG7	B2SAFE	45 TB	34 TB	EPOS - INGV	Earth Science, ...	2012/12/01	production
B2SAFE for CMP5	B2SAFE	20 TB	20 TB	ENES	Climatology	2012/12/01	terminated
INCF DataSpace	B2SAFE	10 TB	8 TB	INCF	Neuroscience, Neuroinformatics	2015/03/04	production
B2SAFE for MIN	B2SAFE	12 TB	4.72 TB	MIN	Biodiversity	2014/01/01	production
B2SAFE for LINDAT	B2SAFE	200 GB	3 GB	CLARIN Center Lindat (CUNI)	Linguistics, Humanities	2013/08/01	production
IST DataRep	B2SAFE, B2FIND Template	2 TB	1 GB	IST Austria, Library Services	Life Sciences, Physical Sciences, Formal Sciences	2016/01/01	pre_production
SIMCODE-DS	B2SHARE Template	10 TB		M. Baldi, Department of Physics and Astronomy, Alma Mater University Bologna	Astronomy, Cosmology	2016/04/01	enabling

Figure 7: DPMT Projects overview panel

Figure 8 shows the Registered Services overview panel.

Title	Contact	Manager(s)	Monitored	Service component(s)	State
B2SAFE for IST DataRep	Jörg Meyer	Jörg Meyer	Yes	irods-ist.lsf.kit.edu - b2safe.irods	external
EUDAT COLLABORATIVE SERVICES	Heikki Helin	Heikki Helin Pietari Hyvärinen	Yes	svn.eudat.eu - eudat.codeRep.svn crowd.csc.fi - b2access.crowd jira.csc.fi - eudat.its.jira www.eudat.eu - eudat.webpresence confluence.csc.fi - eudat.cms.confluence	internal
B2ACCESS	Leon Charl du Toit	Sander Apweiler Benedikt von St. Vieth	Yes	b2access.eudat.eu - b2access.unity	internal
B2FIND	Heinrich Widmann	Heinrich Widmann	Yes	b2find.eudat.eu - b2find.ckan	internal
B2SAFE	Johannes Reetz	Giovanni Morelli Pavel Weber	Yes	irods1.storage.sara.nl - b2safe.irods eudat1.eudat.psc.pl - b2safe.irods opsevadat02.bsc.es - b2safe.irods eudat-icat.pdc.kth.se - b2safe.irods eudat-node1.esc.rl.ac.uk - b2safe.irods dtn01.hector.ac.uk - b2safe.irods ariane.cines.fr - b2safe.irods clarin.uni-tuebingen.de - b2safe.irods irods0.eudat.esc.rzg.mpg.de - b2safe.irods eud-rest02.csc.fi - b2safe.irods epic3.storage.surfsara.nl - b2handle.handle.api epic4.storage.surfsara.nl - b2handle.handle.api irods-3.eudat.lsf.kit.edu - b2safe.irods irods-ist.lsf.kit.edu - b2safe.irods irods1.mn.rug.nl - b2safe.irods epic.rzg.mpg.de - b2handle.handle.api data.repo.cineca.it - b2safe.irods irods1.mn.rug.nl - b2safe.irods lux115.mpi.nl - b2safe.irods	internal
B2SHARE	Lassi Lehtinen	Pietari Hyvärinen	Yes	eudat1.csc.fi - b2share.invenio epic3.storage.surfsara.nl - b2handle.handle.api b2access.eudat.eu - b2access.unity	internal

Figure 8: DPMT Registered Services overview panel

Figure 9 shows the Registered Service Component overview panel

Title	URL	Provider	Contact(s)	Monitored	Host name	State
ariane-test.cines.fr - b2safe.irods	http://ariane-test.cines.fr:1247	CINES	Stephane Coutin	No	ariane-test.cines.fr	internal
ariane-test.cines.fr - b2stage.gridftp	http://ariane-test.cines.fr:2811	CINES	Dugénie Pascal	No	ariane-test.cines.fr	internal
ariane.cines.fr - b2safe.irods	irods//ariane.cines.fr:1247	CINES	Dugénie Pascal	Yes	ariane.cines.fr	internal
b2access.eudat.eu - b2access.ca		JUELICH	JUELICH Contact	No	b2access.eudat.eu	internal
b2access.eudat.eu - b2access.oauth	https://b2access.eudat.eu:8445/oauth-demo/get_token.jsp	JUELICH	JUELICH Contact	Yes	b2access.eudat.eu	internal
b2access.eudat.eu - b2access.unity	https://b2access.eudat.eu:8445/home/home	JUELICH	JUELICH Contact	Yes	b2access.eudat.eu	internal
b2drop.eudat.eu - b2drop.owncloud	https://b2drop.eudat.eu	JUELICH	JUELICH Contact	Yes	b2drop.eudat.eu	internal
b2find.eudat.eu - b2find.ckan	http://b2find.eudat.eu	DKRZ	DKRZ Support Contact	Yes	b2find.eudat.eu	internal
b2gether.eudat.eu - b2gether.cms.drupal	https://b2gether.eudat.eu	Trust-IT	Trust-IT Support Contact	Yes	b2gether.eudat.eu	internal
b2safe.norstore.uio.no - b2safe.irods	irods//b2safe.norstore.uio.no:1247	UIO	UIO Support Contact	Yes	b2safe.norstore.uio.no	internal
bscirods.bsc.es - eudat.monitoring.ncsa-client	http://bscirods.bsc.es	BSC	Marcel Malet	Yes	bscirods.bsc.es	internal
clarin.uni-tuebingen.de - b2safe.irods	irods//clarin.uni-tuebingen.de:1247	EKUT	Emanuel Dima	Yes	clarin.uni-tuebingen.de	internal
cmon.eudat.eu - eudat.monitoring.ncsa-client	https://cmon.eudat.eu	GRNET	GRNET Support Contact	Yes	cmon.eudat.eu	internal
con01.rzg.mpg.de - b2host.gateway.ssh	ssh://con01.rzg.mpg.de	MPCDF	Florian Kaiser	Yes	con01.rzg.mpg.de	internal
confluence.csc.fi - eudat.cms.confluence	https://confluence.csc.fi/display/Eudat/Home	CSC	CSC Contact	Yes	confluence.csc.fi	internal
creg.eudat.eu - eudat.registry.gocdb	https://creg.eudat.eu/view_portal	MPCDF	John Kennedy	Yes	creg.eudat.eu	internal
crowd.csc.fi - b2access.crowd		CSC	CSC Contact	Yes	crowd.csc.fi	internal
data.repo.cineca.it - b2safe.dsi	gsiftp://data.repo.cineca.it:2811	CINECA	Giovanni Morelli	Yes	data.repo.cineca.it	internal
data.repo.cineca.it - b2safe.irods	irods//data.repo.cineca.it:1247	CINECA	Claudio Cacciani	Yes	data.repo.cineca.it	internal
data.repo.cineca.it - b2safe.irods	http://data.repo.cineca.it	CINECA	Claudio Cacciani	No	data.repo.cineca.it	internal

Figure 9: DPMT Registered Service Components overview panel

### Content Management Framework Plone

The DPMT is built on top of the open source web-based content management system *Plone* ([www.plone.org](http://www.plone.org)) and the DPMT profits from the Plone content type framework *Archetypes*, as well as other extensions. Because of the latter, the latest release of *Plone 4* is being used (version 4.3.7 at the time of writing) as Plone's recently released version 5 does not yet fully support *Archetypes* and *Archetypes-related* extensions (like ATExtensions which the DPMT also uses).



## Code Development

The DPMT application is developed as an add-on for Plone (a so-called *Product* in Plone's parlance). Its code base is managed on GitHub and is currently accessible at <https://github.com/EUDAT-DPMT> in the directory [pcp.contenttypes](#). In addition, there are Python packages available that realize the content migration and synchronization from the resource coordination tool (RCT) at [pcp.rctsync](#) and the EUDAT Central Site and Service Registry at [pcp.cregsync](#). The latter also includes the code base for the synchronization scripts with the forthcoming Service Portfolio Management Tool (SPMT). It is planned to factor this part out into its own package as it matures further. All the DPMT software that has been developed is available under the GNU Public License (GPL) as Plone itself is only available under GPL.

## Test and Pre-Production Instances

The deployment of the pre-production site is done at the MPCDF using a virtual machine on its VMWare cluster using SLES 12 as the operating system. Plone's universal installer has been used to install the recommended production setup for Plone, which includes a database replication server (ZEO) and multiple server clients to improve robustness and flexibility. User management and authentication is done using the EUDAT's B2ACCESS service through Apache via Shibboleth, i.e., authenticated users are passed by Apache to the application server. The site is accessible at <https://dp.eudat.eu>. Users need to be entitled.

## Creation of Content Types and Synchronization of Data

While content can, and has been, created using the web interface of the application, there is also a need to transfer content from already existing sites which the DPMT will replace or complement at some point. Most notably, these are the Resource Coordination Tool ([rct.eudat.eu](#)) and the Central Registry ([creg.eudat.eu](#)). So far, information about data projects, clients, contacts, providers and service hosting environments has been transferred using the scripts provided by [pcp.rctsync](#). The final synchronization will be done shortly before the DPMT goes into production. Additional information about providers, as well as registered services, their components and types, has been transferred from the Central Registry using the tools provided by [pcp.cregsync](#). The final synchronization with the GOCDB will also be done before the DPMT goes into production. Last, but not least, information about EUDAT's service offerings and the components available to realize them can be obtained from the Service Portfolio Management Tool using tailor-made scripts also available from [pcp.cregsync](#). The DPMT is periodically fetching new information from the SPMT.

## 3.2. Service Portfolio and Catalogue Management Tool

This phase of the EUDAT project has defined a Service Portfolio Management process (SPM) to formalize the use of various IT services along with their performance and support features reflected in SLAs. The SPM process is documented in the EUDAT2020 deliverable D2.1 *Service Portfolio Processes Definition and SLA Template Set*.

To facilitate the SPM process, task 6.1.3 is building the *Service Portfolio/Catalogue Management Tool* (SPM Tool). The SPMT makes it possible to manage definitions of the different classes of services and service components. Unlike the DPMT, the SPMT does not manage information about concrete service instances. The SPMT provides a programmatic interface and a web-user interface to the service portfolio and catalogue database (API and GUI). The tool facilitates the management of service definitions during the full lifecycle, from the service definition (which is created and progressively specified as part of the service portfolio) and later as an information object of the service catalogue.



### 3.2.1. SPMT Architecture

The SPMT is being developed as a Django<sup>39</sup> application with the following main components.

The *Data storage layer* is a persistent storage for the portfolio/catalogue information implemented by a MySQL<sup>40</sup> database. The SPMT is designed to support version control, such as roll back to any point in the portfolio/catalogue history, and safe deletion.

The *Application layer* is the SPMT's business logic using the Django framework with an object-relational mapper that implements the MVC pattern. The data model is represented by a set of Python classes ("Model"); a system for processing HTTP requests renders the data in combination with a browser ("View") and a regular-expression-based URL dispatcher for interacting with the tool ("Controller"). The RESTful API (with JSON) is based on the Django REST framework.

Clients like the DPMT or the EUDAT web site that displays the service catalogue for the public can use the SPMT API. The SPMT GUI will also use this API. This GUI will be built using a Javascript web framework (AngularJS<sup>41</sup>) and it allows the different roles within the EUDAT Service Portfolio Management process to manage the service portfolio and catalogue.

### 3.2.2. Model/Database Design

The SPMT implements the schema of the service portfolio and catalogue described in the EUDAT2020 D2.1 deliverable. Figure 10 displays the high-level model of the SPMT. As for the DPMT, the main information class is the *Service*. This class contains information, such as the name and description of the service, as well as associated attributes, such as *internal* (indicating belonging to the service portfolio only), or *external* (indicating belonging to the service catalogue as well). In order to support information about different *Versions* of the *Service*, the high-level model introduces the concept of *ServiceDetail*. The *ServiceOption* maintains information about the different options for each *Service* and the associated *SLA* that accompanies the option. Dependencies, components and contact people are defined by other information classes. Note that the high-level model in Figure 10 does not include the SPM-specific roles which have specific privileges to access the portfolio or catalogue.

---

<sup>39</sup> The Django framework, <https://www.djangoproject.com>

<sup>40</sup> MySQL, <http://www.mysql.com>

<sup>41</sup> AngularJS, <https://angularjs.org>

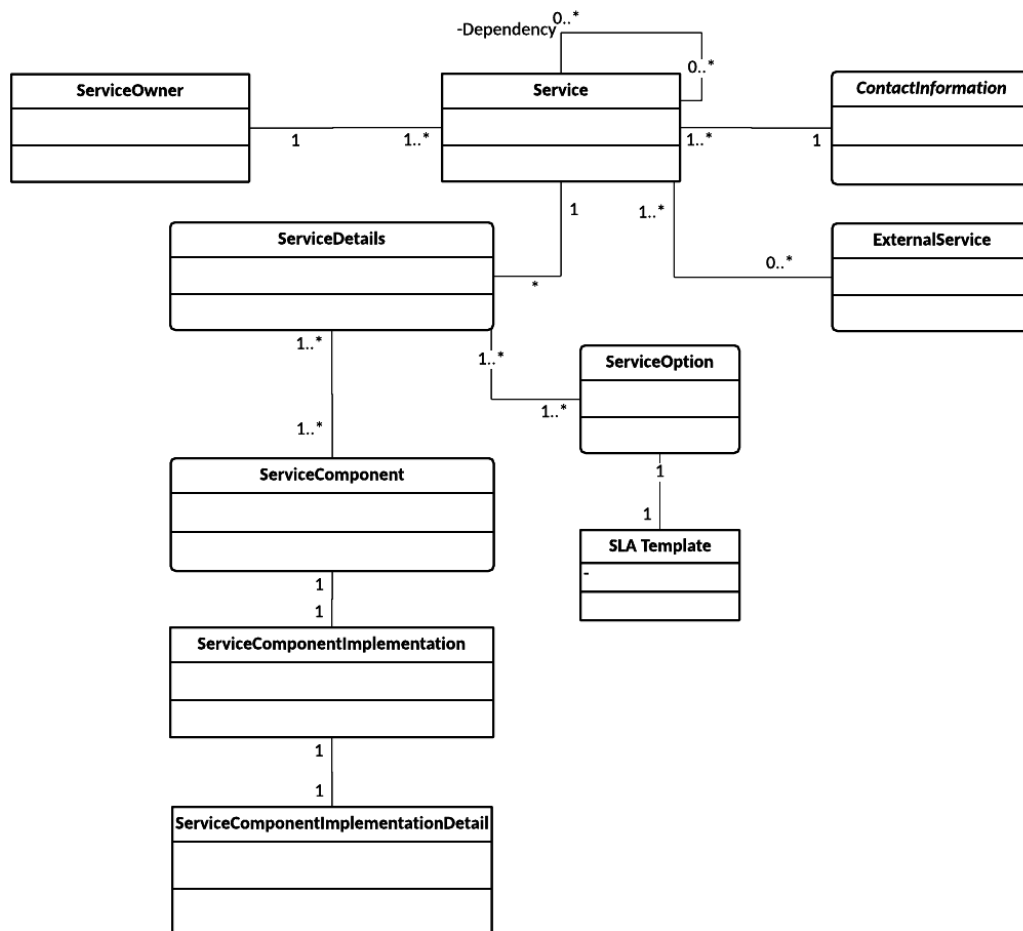


Figure 10: The SPM Tool – High-Level Model (UML)

### 3.2.3. Main Use Cases

The following main use cases and functional requirements have been identified for the SPMT.

1. A *non-registered user* can view only the service catalogue information.
2. A user registers to one of the SPM process roles defined in D2.1. This user is called a *registered user*.
3. A *registered user* views all the service portfolio information.
4. A *registered user* adds, removes or modifies service information from the database.
5. *Registered users* subscribe for (e-mail) notifications of changes in the service catalogue (i.e. introduction of new services, existing services entering production etc.).

### 3.2.4. API Basics

All information to and from the SPMT will be provided via a RESTfull API and be represented as JSON strings. The API is being designed according to the principles and best practices provided in JSON:API<sup>42</sup>. The following table provides some example API calls for reading information from the SPC.

<sup>42</sup> JSON:API, <http://jsonapi.org>

Table 5: Example of the SPMT API

METHOD	ROUTE	Description
GET	/portfolio/services	Lists all services
GET	/portfolio/services/SERVICE_UUID	Lists a specific service
GET	/portfolio/services/SERVICE_UUID/service_details	Lists all the service details of a service
GET	/portfolio/services/SERVICE_UUID/service_details/VERSION	Lists the service details a specific service version
GET	/portfolio/services/SERVICE_UUID/service_owner	Lists the service owner
GET	/portfolio/services/SERVICE_UUID/service_dependencies	Lists the service dependencies
GET	/portfolio/services/SERVICE_UUID/service_external_dependencies	Lists the external dependencies of a service
GET	/portfolio/services/SERVICE_UUID/contact_information	Lists a service's contact information

### 3.3. Monitoring

The Monitoring task provides status monitoring while the user communities can monitor the Availability and Reliability of the CDI services which they have booked in the context of specific data projects. The first year of the project was focused on two goals:

- Retaining the existing Nagios-based (CMON) status monitoring service (<https://cmon.eudat.eu>), the preliminary predecessor developed during the first phase of the EUDAT project, and migration from RZG to GRNET. CMON will be operated until the A&R monitoring service becomes available.
- Designing and starting the implementation of the new Availability and Reliability monitoring service (A&R Monitoring) making use of the ARGO<sup>43</sup> Monitoring Framework. This includes coordinating the development and deployment of the probes for the different B2 services and service suites.

#### 3.3.1. Migration and Operation of the Legacy CMON Service

During the first year of the project, GRNET took over the operation of the CMON Monitoring Service from RZG. The migrated service instance is basically a clone of the previous service instance that is now hosted and operated by GRNET.

#### 3.3.2. Design and Implementation

The new service for availability and reliability monitoring is based on the ARGO Monitoring Framework. This framework provides the building blocks for a scalable, customizable monitoring system. The goal of the first year was to provide a pilot instance of the new service. The two following principles are to be considered.

##### 1. Beyond the status monitoring of technical service components

Traditional status monitoring of services is useful for operations (from the service provider perspective). For customers and the end users, it does not add much value, particularly when the services are complex and if their technical components are distributed across sites. The customer of a service which is delivered in the context of a data project wants to understand how the service is performing from the user perspective and whether the availability and reliability of the service meets the requirements of the service level agreement. Each B2 service comprises several technical components typically operated by multiple service providers.

<sup>43</sup> <http://argoeu.github.io/>

ARGO can monitor any registered service when its properties are published via a configuration information service that offers the GOCDB API. EUDAT is currently using the GOCDB for describing the CDI and the topology of the different services which are provided in the context of data projects.. A topology is a set of service (component) endpoints in combination with the information about the relations between these service endpoints, whether they are front office or back office. The structure of the EUDAT Services is modelled this way and described using the GOCDB (see C.1 for a detailed discussion).

## 2. Monitoring deployment model and probe harmonization

Monitoring services from the user point of view means that all the services have to be monitored in the same way regardless of who the service providers are and where they are located. Apart from having location neutral service monitoring, each type of service component also has to be monitored in the same way. The same tests must apply to all the service components that deliver the same functionality.

In order to achieve these two objectives, we decided to adopt a centralized model for the availability and reliability monitoring in which all B2 services are monitored remotely: monitoring information must not be provided by the service providers themselves. This implies that the actual tests performed against the services must be atomic and at the same time mimic the actual end user behaviour and should not require special privileges or special configurations from the service provider side which are not typical for the user. The monitoring tests for each B2 service are being implemented by the EUDAT service development teams (WP5) according to the guidelines<sup>44</sup> for Nagios probes.



Figure 11: Centralized monitoring of B2 services from the user perspective

### 3.3.3. Status of the Availability and Reliability Monitoring Service

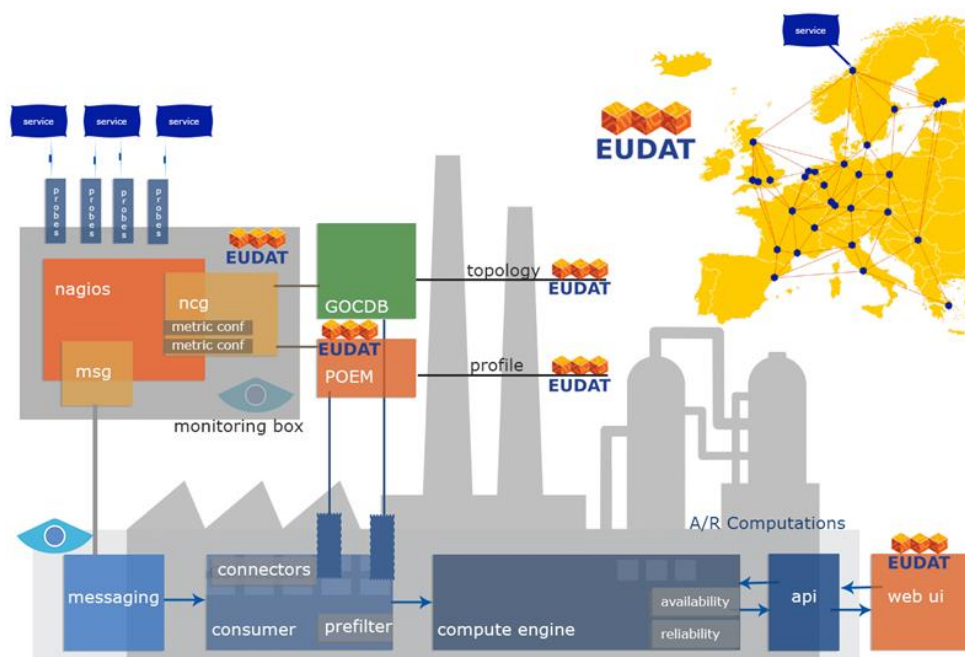
At the end of the first year of this phase of EUDAT, an initial instance of the new A&R monitoring system was provided by GRNET at <http://avail.eudat.eu>. The following steps have been completed:

- configuration and deployment of the new ARGO monitoring engine for EUDAT,
- modelling the topology of the B2 services in the EUDAT CDI using the EUDAT central registry,
- re-engineering the auto-configuration component (argo-ncg-eudat) so that it can automatically configure the monitoring engine using the B2 service schemas described by the EUDAT registry (The new auto-configuration is already used in the monitoring engine and it configures the monitoring engine for all B2 service groups that are registered in <https://creg.eudat.eu>),
- adapting the existing messaging component (argo-msg-nagios) for the EUDAT environment and configuring the ARGO Compute Engine on the pre-production test bed for the new EUDAT tenant,
- deploying and configuring a new ARGO POEM instance for EUDAT (This instance is using the GOCDB API from the EUDAT central registry in order to retrieve all the available service types and then it lets the metric profiles that specify which monitoring probes are used for which service types be configured.),
- creating new availability report configurations for all the B2 services,

<sup>44</sup> <https://nagios-plugins.org/doc/guidelines.html>

- testing all service types initially and by default using ping and a simple tcp\_check, when applicable (New probes for the B2 service will be added to the profile when they are available.),
- developing two probes for B2HANDLE: one that tests the create, read, update and delete capabilities of the API component of the service and the other that tests the resolution and mirroring capabilities of the resolver component (These probes serve as blueprints for the development of further probes.), and
- configuring and deploying a new ARGO Web UI instance for EUDAT on the pre-production server.

To summarise, in the first year of this phase of the EUDAT project the requirements and the current CDI architecture have been analysed and a schema has been defined that was implemented in CREG. It allows for the proper description of the EUDAT CDI. The ARGO Service Monitoring Framework was adopted to meet the requirements of the EUDAT CDI and, since December 2015, the pilot service has been operational on a pre-production environment monitoring the actual production B2 services. Guidelines have been delivered. The first probes for the B2Handle service have been developed and are operational on the pre-production environment.



**Figure 12: Architecture of the availability and reliability monitoring**

The schema used to describe the topology of each of the B2 services (and later also the B2 service suites) is evolving. The current Site and Service registry makes use of the GOCDB extension properties which will also be supported by the DPMT. The ARGO system can read from any site and service registry that supports the GOCDB API. For example, the B2Handle service is a multi-tenant service, meaning that one service instance can host Handle prefixes for many communities. The schema used with the GOCDB, as well as later with the DPMT, makes it possible to describe the project-wise delivered B2 services in detail. The auto-configuration component of the ARGO Monitoring engine will be enhanced in order to parse the new schema and generate the appropriate configuration.

Users from the scientific communities will be able to access the availability and reliability reports on the web UI after authenticating through the B2ACCESS service via the SAML protocol.

The current implementation of the Web UI already supports public and private access through SAML and certificate based authentication. In the case of private access, all the reports are marked as private and only authenticated users can access them. In the second year of this phase of EUDAT we will investigate the requirements for access controls on the report level, meaning that users from a specific scientific community

or project are entitled to see only the reports concerning their own data project. Figure 13 shows the current A&R report panel.

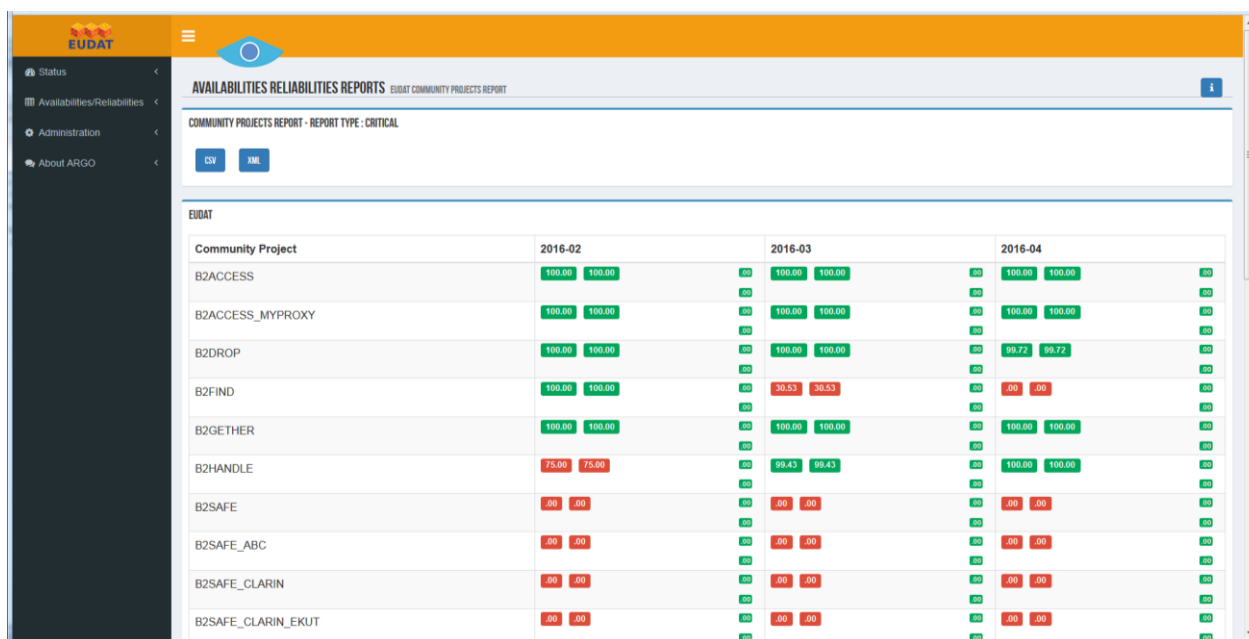


Figure 13: Availability & Reliability monitoring panel. The probes for the service instances are preliminary.

### 3.4. B2ACCESS – A Federated Identity and Access Management Solution for EUDAT

The B2ACCESS<sup>45</sup> service, <https://b2access.eudat.eu>, provides a unified authentication and authorization service which makes use of external (primary) Identity and Attribute providers. The service has been developed by WP5 using the Unity Identity Management framework. The service is presented and explained in the WP5 D5.1 deliverable. B2ACCESS simplifies both service management and user interaction with EUDAT services by offering multiple-credential single sign-on which can be integrated straightforwardly with any service.

B2ACCESS currently supports multiple technologies to integrate external IdPs: (1) SAML, (2) OAuth2 and OpenID Connect, (3) X.509 certificates (IGTF and DFN Global), and (4) username/password. From the service provider perspective B2ACCESS supports multiple technologies to integrate backend services: (1) SAML, (2) OAuth2 and (3) X.509. Users have the freedom to sign in with institutional, social or individual credentials; service managers can connect their services into B2ACCESS using authentication mechanisms supported naturally by the service software components.

The B2ACCESS service is arbitrating access to registered Service Providers (Downstream Service Providers) based on the user's chosen primary Identity. These Downstream Service Providers consume Attribute assertions provided by the B2ACCESS service when the end user accesses these services.

A specific level of assurance (LoA) is assigned to the user identity, depending on the authentication method that was used when the EUDAT ID was created. There are currently three levels of assurance supported. Firstly social IdPs and username/password are given a low LoA because these users can register and provide any information by themselves. Secondly eduGain IdPs are assigned a medium LoA because the IdPs check the authenticity of the identity holders more carefully. Identities that are based on personal x.509 certificates from IGTF or DFN Global are assigned a high LoA.

<sup>45</sup> EUDAT2020 WP5 deliverable D5.1 (chapter 4.6, p.41)



### 3.4.1. The Production Environment for B2ACCESS

The production service is provided by Jülich as described in the data privacy statement<sup>46</sup>. The service complies with the GEANT Code of Conduct and is registered with eduGain as the Service Provider. The integration with eduGain was very efficiently supported by the DFN and GEANT. Details for users are described in the user documentation<sup>47</sup>.

The integration of B2ACCESS with new services is described by the service integration document<sup>48</sup>. The service management document describes the tools and procedures for the administration of the B2ACCESS instance<sup>49</sup>.

The service is operated by two local service operators (having root access) and administrated by one local B2ACCESS manager in Jülich who shares this task with an administrator from the EUDAT partner Sigma/UiO. Figure 14 indicates the status of the integration with external IdPs and Service Providers.

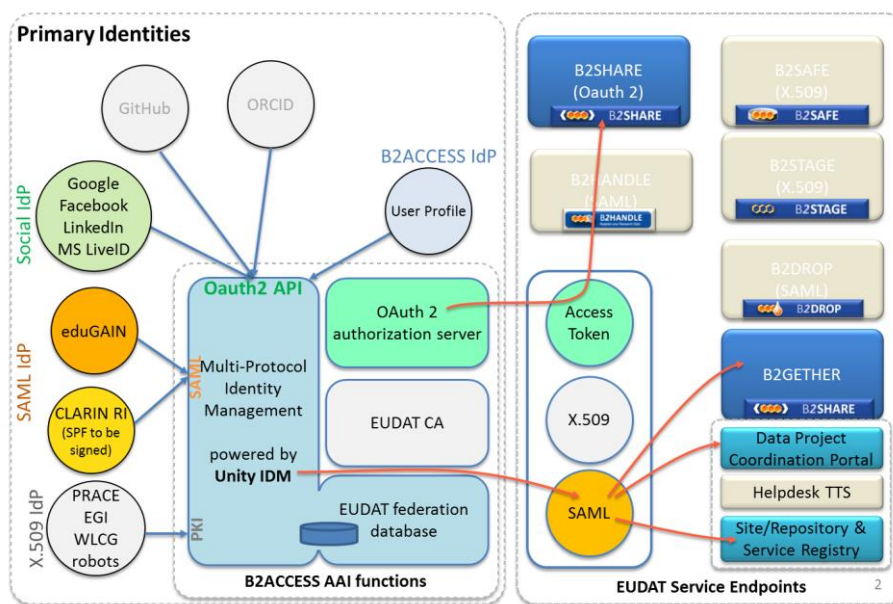


Figure 14 B2ACCESS schematic showing the integration status in Feb 2016

The following primary IdPs can be used via B2ACCESS:

- eduGain IdPs (as far as the attributes are released),
- Google IdP, Facebook IdP, MS Live IdP, and
- X.509 personal certificates.

### 3.4.2. eduGAIN SAML IdPs

Figure 15 shows the evolution of the uptake of the B2ACCESS service since it started in Nov 2015. The diagram differentiates between accounts that are linked to primary identities from eduGain IdPs, social IdPs and the B2ACCESS IdP (i.e. self-service registration, including users with x.509 certificates). Since the EUDAT user forum in Feb 2016, we have seen a constant growth of approximately 50 new accounts per month.

We notice that eduGain IdPs are not used as primary identity provider as much as was initially expected. There may be different reasons for this and a careful analysis is required because eduGain IdPs have a large research user base that should benefit from EUDAT services.

<sup>46</sup> <https://b2access.eudat.eu/files/data-privacy-statement.html>

<sup>47</sup> <https://eudat.eu/services/userdoc/b2access-usage>

<sup>48</sup> <https://eudat.eu/services/userdoc/b2access-service-integration>

<sup>49</sup> <https://eudat.eu/services/userdoc/b2access-management>



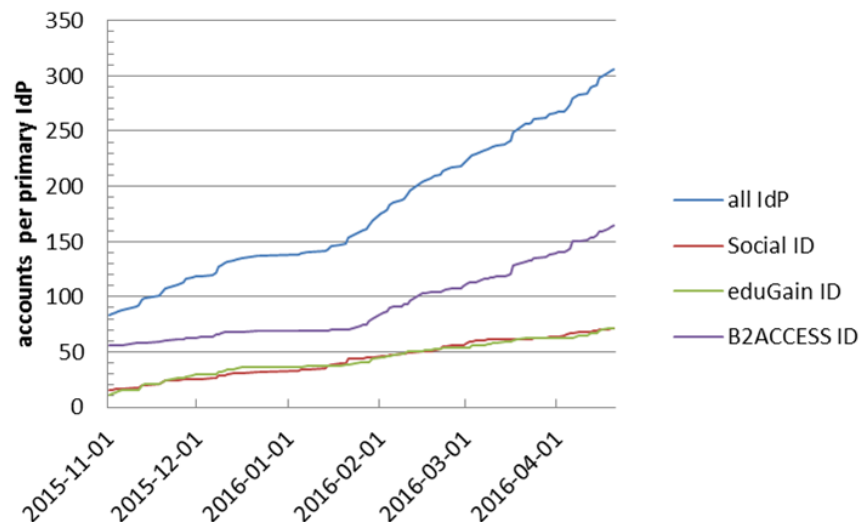


Figure 15 Evolution of the number of users signed in with B2ACCESS (snapshot from April 2016)

One of the reasons is the increasing amount of IdPs using Shibboleth v3, which is not supported by the current version of Unity version 1.6.1. The planned upgrade will solve this.

Another complication is the fact that many eduGain IdPs accept B2ACCESS (as SP) only if their users are explicitly demanding it and some IdPs want to have individual assertions (contracts) with the B2ACCESS provider (Jülich) before they release the few attributes (eMail and ePPN as mandatory, cn and schachHomeOrganization as optional) which are needed by B2ACCESS and its downstream SPs.

The fact that IdPs which are listed by the eduGain registry can individually reject attribute release requests from SPs which have signed the GEANT CoC indicates a well-known issue with the current eduGain federation approach: users cannot predict whether the “eduGain IdP” of their institutions accept the attribute release requests from the SPs that the users want to access, even when the users’ consent to attribute release has been explicitly documented. Another issue is that some IdPs do not follow the standard for a trusted information exchange and do not sign the SAML assertion elements in response messages.

### 3.5. B2HANDLE Setup and Operation

B2HANDLE is a joint service, currently provided by six<sup>50</sup> EUDAT partners (BSC, DKRZ, SURFsara, GRNET, MPCDF and PDC), for registering, storing, mirroring and resolving persistent identifiers based on the Handle system<sup>51</sup> and in collaboration with the ePIC<sup>52</sup> consortium. This service includes support in obtaining Handle prefixes that are bound to one of the B2HANDLE registries. The service is offered to the European research communities that are using EUDAT services and that want to create and manage their own persistent identifiers for repositories or other digital object registration services.

The key aspect of B2HANDLE is that Persistent Identifiers (PIDs) are supposed to provide long-term stable references to their target objects. In order to receive the added value of PIDs, users need to trust the providers of PID registries, as well as those instances that are responsible for managing and curating the PIDs for data sets in their domain. It is beneficial that many of the B2HANDLE service providers are also partners in the ePIC consortium. The B2HANDLE service development is described in the WP5 deliverable<sup>53</sup>.

The technology behind the B2HANDLE service is based on the *Handle.Net*<sup>54</sup> software with currently the *ePIC API* on top. This ePIC software fills a gap in the basic PID management functionalities of the Handle software.

<sup>50</sup> BSC and GRNET joined with their services in 2015 and so enhanced the B2HANDLE service with resources and competences.

<sup>51</sup> <https://www.handle.net>

<sup>52</sup> <http://www.pidconsortium.eu>

<sup>53</sup> EUDAT2020 WP5 deliverable D5.1, 2016, Chapter 5.2.1

<sup>54</sup> [http://handle.net/download\\_hnr.html](http://handle.net/download_hnr.html)

It offers a generic user/machine interface for creating and managing/curating PIDs per Handle prefix. It provides features such as an efficient reverse lookup mechanism per PID service provider. The API is generic to fulfil the needs of a broad range of usage scenarios from different scientific communities.

In addition to the software, the *Handle System* is necessary as global administrative framework. This System is an infrastructure that provides general purpose identifier resolution services. Using the open set of protocols that the Handle System provides, a distributed computer system can create and manage handles (identifiers) for any digital resources. The same set of protocols allows a client to locate the digital resource identified by a handle, query the metadata for that resource and access the data associated with the resource<sup>55</sup>.

The B2HANDLE service operations team consists of EUDAT experts operating and maintaining the service. The main activities of the team, in the past year, were setting up and enhancing the B2HANDLE service by including further service providers (BSC and GRNET), providing software update support, coordinating the assignments of prefixes<sup>56</sup> and the mirroring of PIDs between B2HANDLE service providers (which included the operation of the B2HANDLE service), preparing for the major service upgrades in 2016, harmonising the operational procedures with ePIC and contributing to the policy discussions that will impact the operational procedures and practices in EUDAT.

This first project year has been characterised by major changes that impact the operations of the distributed B2HANDLE service. CNRI<sup>57</sup> announced the major release of the handle software version 8.1. In conjunction with this release, changes in the management of the Global Handle Registry took place: CNRI is now a Multi-Primary Administrator (MPA) amongst others that are accredited by the DONA Foundation<sup>58</sup>, together with ePIC. This will enable the B2HANDLE providers that are also ePIC members to allot prefixes to users of the Handle System.

Via these partners and the B2HANDLE service providers, EUDAT is able to issue Handle prefixes according to a prefix management policy. The EUDAT prefix management policy should harmonize with the ePIC policies, and therefore the B2HANDLE team members collaborate closely with ePIC. Policy aspects are e.g. whether, when and how PIDs may be deleted, or the annual fee for keeping the prefix registered by the MPA. Both policies are currently in the process of being finalised and will become publicly available in 2016.

Handle version 8.1 is now available and includes many improvements, features and tools that were required but not available in previous versions. Some of these features will impact the usage of the ePIC API that will become obsolete after a necessary grace period.

The B2HANDLE operations team has planned for the upgrade to version 8.1 on the overall EUDAT Collaborative Data Infrastructure (CDI). During a first stage, the Handle.Net software, v7.3.1 will be upgraded with new version, v8.1. This process has already been started. And, in order to keep compatibility with all B2 services interfaces (B2SAFE, B2SHARE and B2STAGE), ePIC API v2.5 will replace v2.3. The second stage will then require the upgrade of the B2 Services that rely on newly developed B2HANDLE library developed by WP5. This will let us decommission the *ePIC API* and replace it with the B2HANDLE library.

All these changes, the new software stack and the new policies will impact service deployment, configuration, documentation, monitoring and a few other operational procedures and aspects, such as the specification of service agreements and the issuing of prefixes.

---

<sup>55</sup> EUDAT2020 WP5 deliverable D5.1, 2016, Fig. 27 (p.52): overview over the envisioned evolution of the B2HANDLE service.

<sup>56</sup> An inventory of all prefixes within EUDAT is maintained.

<sup>57</sup> <http://www.cnri.reston.va.us>

<sup>58</sup> <https://www.dona.net>

## 4. SERVICE AND RESOURCE PROVISIONING

This chapter provides an overview over the service management tasks in the area of the distributed *Service and Resource Provisioning* and *Operations Coordination*. The next section 4.1 describes the activity of the operations coordination that takes care<sup>59</sup> for the change management (PR12), capacity management (PR5) and other processes. The following section 4.2 presents an automatic solution for deploying B2 service software at the different provider sites via puppet. This can facilitate the installation and upgrade procedures at the provider sites. Section 4.3 discusses the change management process that has been further developed by the OPCT, and section 4.4 presents a solution that collects and publishes information about the actually installed product versions. This allows to compare the status of the installed software with the latest software releases and to create a trigger if a threshold is exceeded. Section 4.5 informs about the operational tools that have been employed in the first project year, section 4.6 is about EUDATs service hosting service and, finally, section 4.7 explains that the CDI providers are sharing information about their approaches to ensure the availability and continuity of their services.

### 4.1. Operations Coordination

The distributed CDI services are basically still operated following the rules that have been defined in 2012<sup>60</sup>. For instance, the service providers are in charge to keep the information about their sites, their responsible staff, and their services up-to-date in the GOCDB. These rules have been slightly adapted in the meanwhile. In addition the security policy, the change management policy and the user support guidelines are in place and have also been documented in the annexes of the previous WP6 deliverables of the first phase of EUDAT. While these rules are still valid, they are insufficient, need to be updated and consolidated, and therefore the operational procedures with roles and responsibilities have been started to be defined as service management processes which are discussed in section 2.5.

Capacities are to be managed as well as the continuation of the service provisioning, changes are to be managed, releases and configurations. The related local activities require coordination across the CDI. In order to facilitate this collaboration, WP6 organises the cooperation between *thematic* and *generic Service Providers* via an *Operations Coordination Team (OPCT)* that comprises senior site deputies (and staff members) that are representing basically the *integrated Service Providers*<sup>61</sup>. These integrated service providers play an important role for the knowledge exchange within the CDI: these partners are experienced in deploying, running and maintaining different core components of the CDI, and they often take responsibility, as general service providers, for provisioning of the distributed *Service Instances* in the context of data projects (e.g. B2SAFE topologies).

The information exchange between the providers is facilitated via the EUDAT Wiki, operational tools (section 2.3), the helpdesk system, and email lists. In addition, the OPCT has regular meetings once per month on invitation by the OPCT leader. On the agenda of these meetings are topics such as vulnerabilities, security incidents, operational disruptions and solutions, change, release and deployment management request - topics that are relevant to raise the awareness of all the providers. The OPCT is reviewing change management requests and deployment plans, and it decides about the adjustments of the overall operational approach.

The provision of the resource capacities, service components and service hosting facilities is coordinated to configure and deliver the service components in time that are required for the CDI services for the diverse data projects and pilots. The generic service providers have pledged resources (storage, service hosting) in three categories: “non refunded”, “eligible costs”, “pay-per use”. And they offer to run service components (B2 service components, operational tools, see Table 1). These offers are the basis of the capacity management, the provisioning of resources by the generic service providers, the deployment and

<sup>59</sup> actually the team leader is accountable for the process.

<sup>60</sup> <https://www.eudat.eu/d62-1st-year-report-status-and-progress-operations>, pp.49

<sup>61</sup> see page 13 for the definition of interoperable and integrated nodes.12

maintenance of software and the management of the configuration information via the site and service registry (currently RCT, GOCDB; in future replaced by the DPMT). The generic service providers made particular pledges for storage and service hosting capacity for the 1<sup>st</sup> call for data pilots in EUDAT2020.

In the focus of the OPCT within the reporting period were software upgrades, security related maintenances and the provisioning of the B2 services for the existing and new data projects.

## 4.2. Service Installation Support for Release and Deployment

The increasing number of EUDAT services requires a robust and structured approach to configuration management for the services and their components as for generic EUDAT data centers so for community repositories. In order to take control over the installation and configuration process currently for the most widely used B2SAFE services a unified configuration management and deployment procedure has been developed based on Open Source tool *Puppet*. The benefit of this approach is that it provides an automatic and uniform workflow for the installation and configuration of B2SAFE service. Fully documented with step-by-step instructions, the developed package is suitable also for the beginners without particular experience in B2SAFE administration. Current implementation of the package includes the following functionality:

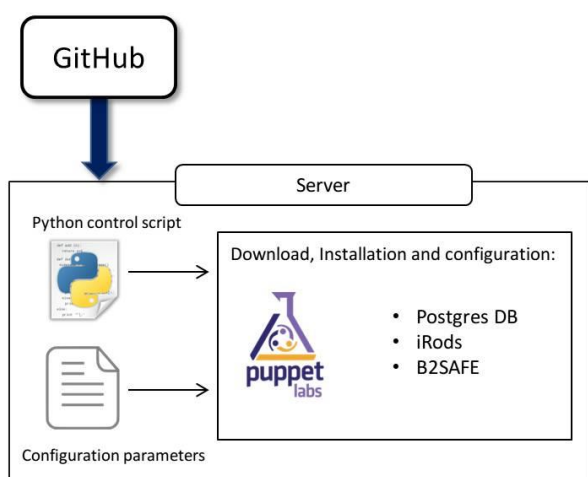


Figure 16: B2SAFE automatic installation via Puppet

- Installation and configuration of Postgresql 9.3
- Installation of iRods 4.1.5 or iRods 4.1.6
- Installation of B2SAFE package 3.0.0
- Support for CentOS7 and Scientific Linux operating systems.

The configuration parameters have to be defined for the installation procedure. These parameters include a set of configuration variables for database like the installation path, database admin password and user name, a set of specific configuration variables for iRODS and B2SAFE including parameters for the PID management. If the configuration parameters are not specified for most of them the default values will be used by the installation procedure.

The description of the package is located on GitHub at <https://github.com/EUDAT-B2SAFE/B2SAFE-puppet>. The typical application workflow of package as follows:

- Download the package from GitHub.
- Edit the configuration file and modify the default values of the configuration parameters.
- Run the python control script, which:
  - Installs the puppet repositories.
  - Downloads and installs puppet tool.
  - Executes the puppet manifest in the masterless mode to install and configure the given stack of software.

**Table 6: Summary of activities related to automatic iRODs/B2SAFE installation procedure.**

Site	Activity
KIT	Development of package, installation of production instances.
CINES	Testing of standalone automatic installation.
Juelich	Testing of the package in Puppet infrastructure. Testing and bug fixes.
PSNC	Application of the package for installation of production instance.
MPCDF	Testing of standalone automatic installation.

The package was initially released in November 2015. The development is currently focused on extending the support of additional operation systems and software versions. The summary of activities at sites related to development, testing and application of the package is given in Table 6. In the pilot phase the package has been used to deploy the B2SAFE instances at KIT and has been tested by CINES and MPCDF site deputies. The main application scenario is the application of the package in the standalone mode, i.e. without a Puppet infrastructure at the site. The application scenario and code which allow integrating B2SAFE instance in the established Puppet infrastructure is provided by branch “core” on GitHub and is currently being tested at the Juelich data center.

### 4.3. Change management

Change Management is one of the FITSM processes (see PR5 in 2.5, p.15). Changes on components in the distributed EUDAT infrastructure can have significant impact on the CDI. The impact of any change such as an upgrade of an existing service (component), the migration of the existing service to another location or the installation of a new service which could have a significant impact for other services, service providers and the users of production services must be classified. Significant changes need to be appropriately planned, documented and implemented. The operations team agreed on a standardized method change management in CDI infrastructure. The emergency or repair changes related to broken services are covered by the incident management and treated separately. The procedure of change management is divided in three phases:

- Preparation
- Implementation
- Evaluation

During the *preparation phase* the purpose of the change request is announced by a requester via the JIRA ticketing system. The ticket is being promoted to all the potentially affected sites. The change has initially to be described by the requester but the corresponding plan should take the comments and constraints of affected partners into account. As a result of the preparation phase the change can be scheduled. Especially changes that require peculiar efforts and take long time periods must be carefully planned. Upon the initial agreement on change the change requester follows the change management procedure and fills the change management request form:

- Brief explanation of the change, benefits and impact to sites and user communities
- Does the change affect EUDAT B2\* services, if yes, which services and which customers/user groups are potentially affected to which extend?
- Expected starting date and time
- Expected duration
- Short change plan
- Short testing plan and rollback plan if applicable
- Possible security risks
- Downtime required (yes/no)
- Follow-up actions
- Change management group, contact person

Currently we don't implement different types of change request forms and treat all requests in common way. If the change affects other EUDAT services the change management group with assistance of WP6 operations team should alert all sites explicitly on change and provide documentation.

The change request form is prepared as a template of JIRA ticket and after submission and approval also acts as the central communication and progress tracking unit during implementation phase. During implementation phase a scheduled downtime is declared if needed and the change is implemented and all major steps are documented in JIRA ticket.

At evaluation phase the change management group collects feedback and analyzes any problems reported which could be related to the change. The implemented change management procedure was successfully recently applied for update of TTS system at CINICA and ongoing upgrade of PID Handle System and B2SAFE instance at CINES. All open, pending and closed change management actions are described and tracked via the internal wiki page.

#### 4.4. Software Version Monitoring

The development teams deliver the Services as *Service Packages* which versioned bundles of software with descriptions. These packages do not necessarily contain the full stack of software components for every service component. There are prerequisites described in the release notes of each Service Package such as specific or minimum version levels of the service components and the properties of the system platform that is hosting the service components (Figure 9). The version numbers of the service packages as well as the version numbers of the required service components are specified in particular in the service descriptions of the SPMT (section 3.2). These descriptions provide important information for the deployment of the services and their components. From the operations point of view it is necessary to keep an overview over the over the actually installed software versions in the CDI. For this purpose the software version monitoring system has been developed. In combination with the information about the available from the to perform an efficient change management and automatically control the versions of installed software. As the integration of such system in the monitoring system would course several technical difficulties it was agreed to implement it as a standalone framework. The framework allows the monitoring of the following information:

- Operating system flavor and version of the endpoint.
- EUDAT service component software versions.
- Any service software version installed on the endpoint.

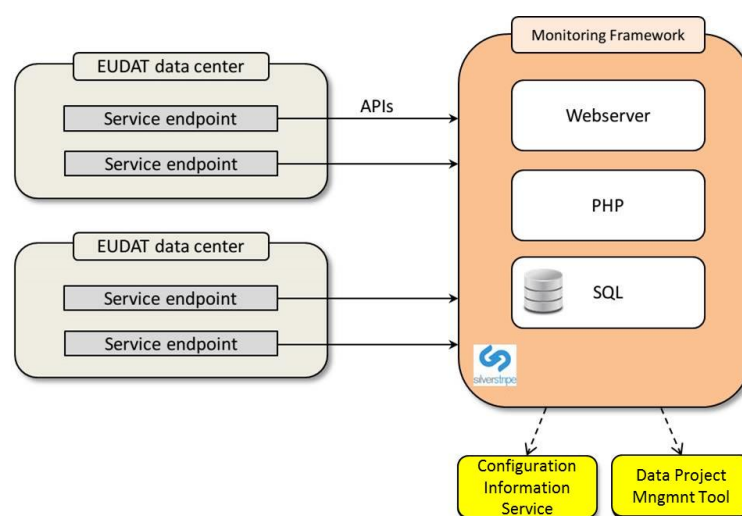


Figure 17: Product versions monitoring system



The software version monitoring framework consists of the web-based central portal which collects the information on software versions, stores it in the database and displays it in a compact overview table. The portal has been developed using the Silverstripe Content Management System which includes a standard webserver implementation, a PHP coding base and a SQL database backend as depicted in Figure 17. The information on software versions is collected on service endpoints at EUDAT data centers by the scripts running as cron jobs and transmitted to the remote server using a Restfull API.

By proposed design the monitoring framework should regularly update the Data Project Management Tool with the actual information about the software versions. In order to keep the list of all EUDAT endpoints and services up to date the monitoring framework contacts the Configuration Information Service (either the GOCDB or the DPMT) as an authoritative source of information, fetches the list of all endpoints via the REST API and updates its entries accordingly.

Currently the initial prototype of monitoring framework is ready and running at <http://eudat-tags.scc.kit.edu/eudat/>. A detailed description of the information that is currently collected by the software versions monitoring system is given in Annex E.

#### 4.5. Service Management and Operational Tools

The main tools used for the service management in the context of *Service and Resource Provisioning* and the *Operations Coordination* have been the *Site and Service Registry* based on the GOCDB (<https://creg.eudat.eu>), the *Resource Coordination Tool* (<https://rct.eudat.eu>), also used also for collecting and displaying accounting information, and the status monitoring system <https://cmon.eudat.eu>.

**Table 7: Statistics of currently registered service components**

Service Providers	Service Types	All Service Components	B2SAFE Service Components	Newly installed service components	Upgraded service components
22	29	55	19	11	14

The static configuration information about the *Service Providers* (Sites), *Service Components* (Service Endpoints) and distributed Services (Service Groups) have been stored and managed via the *Site and Service Registry*. This registry was also used to inform the CDI about scheduled downtimes for services.

The distributed services that have been enabled in the context of data projects are registered as *Service Groups*. Table 7 shows the current statistics about the number of *Service Components* in the CDI that are currently provided in the context of the different distributed *Services* and *Data Projects*. Each Service Component can be part of a distributed service, or, as it is often the case for the operational tools, the service has just one single component.

**Table 8: Services installed during the 2<sup>nd</sup> phase of EUDAT**

Site	Related Service	Service Endpoint
BSC	B2SAFE B2HANDLE	opseudat02.bsc.es eudat-epic.bsc.es
GRNET	CMON Status Monitoring ARGOS Availability & Reliability Monitoring B2HOST	cmon.eudat.eu avail.eudat.eu oceanos.grnet.gr
	B2HANDLE	epic.grnet.gr
Juelich	B2ACCESS	unity.eudat-aai.fz-juelich.de
KIT	B2SAFE B2SAFE B2HOST	irods-3-eudat.lsfdf.kit.edu irods-ist.lsfdf.kit.edu
MPCDF	DPMT (prototype) RCT	dp.eudat.eu rct.eudat.eu
Sigma/UiO	B2SAFE	b2safe.norstore.uio.no
CSC	WebSite (managed by Trust-IT) B2GETHER (managed by Trust-IT)	www.eudat.eu b2gether.eudat.eu

The third column of the Table 7 provides the number of all service components in production which are currently operated at the EUDAT sites. The last two columns display the numbers of newly installed and upgraded service components. Some detailed information about the newly installed services is presented in Table 8. The information concerning to sites and service components have provided by the site deputies and their staff (OPCT). The information concerning distributed services (service groups) are specified by the project enabling team. As soon as a distributed service is commissioned (in production) the OPCT takes over the responsibility for managing the information.

For planning and coordinating the provisioning of storage capacities and resource usage in the context of data projects the Resource Coordination Tool has been used. This tool provides information about the pledged resources (resources offers by the partner sites) and the allocated resources. This RCT was used beside the EUDAT wiki to register accepted data projects and to coordinate the requested for storage and compute resources. The RCT has also further been used to collect and display the storage usage accounting information as described in a previous deliverable<sup>62</sup>. The GOCDB and RCT are planned to be replaced by the DPMT, and at the same time an upgraded accounting information service will be introduced.

The quality of EUDAT operations relies also on the status monitoring system (see also section 3.3.1, p.32) that provides information about the status of the service components based on the configuration descriptions (topology) provided by the Site and Service Registry.

#### 4.6. Service Hosting Framework

Several data centers in EUDAT provide service hosting capacities (IAAS and PAAS) for research communities that need computing resources close to the data, for instance for data processing and analyses. Customers can order these service hosting resources in the context of a data project. Seven data centers are currently providing resources for this purpose based on various virtualization solutions and cloud infrastructures. A summary of the available service hosting resources, registration types and used technologies is presented in Table 9. GRNET and KIT joined in the second phase of EUDAT. The service hosting framework is available since the first phase of EUDAT<sup>63</sup>.

**Table 9: The Service Hosting Environment**

Provider	Cores	RAM	Registration	Description
CINECA	64	123 GB	E-Mail: <a href="mailto:eudat-support@cineca.it">eudat-support@cineca.it</a>	Physical Frontend Cluster
CSC	150	128GB	RCT request	OpenStack cloud
GRNET	64	64 GB	<a href="https://okeanos.grnet.gr">https://okeanos.grnet.gr</a>	IaaS service Okeanos based on Synnefo
MPCDF	32	128 GB	<a href="https://rct.eudat.eu/?path=/shes/">https://rct.eudat.eu/?path=/shes/</a>	VMware vSphere Cluster
Juelich	100	100 GB	RCT request	OpenStack cloud
SURFsara	64	256 GB	<a href="https://userinfo.surfsara.nl/systems/hpc-cloud/obtaining-account">https://userinfo.surfsara.nl/systems/hpc-cloud/obtaining-account</a>	OpenNebula HPC cloud
KIT	24	32 GB	E-Mail: <a href="mailto:eudat-support@lists.kit.edu">eudat-support@lists.kit.edu</a>	VMWare ESX Cluster

Service hosting (compute) resources can be requested by filling a form in RCT tool where the technical parameters such as the number of CPU cores, memory, and operating system are specified. The request for compute resources is part of the data project request that is until now registered via the RCT. The OPCT dispatches this request to the appropriate data center that confirms the possibility to provide the resources or rejects the request. If the overall data project request is accepted it will be possible to access the compute

<sup>62</sup> EUDAT1 D6.4, [https://b2share.eudat.eu/record/216/files/EUDAT-DEL-WP6-D6%204-Final\\_Report\\_on\\_Operations.pdf](https://b2share.eudat.eu/record/216/files/EUDAT-DEL-WP6-D6%204-Final_Report_on_Operations.pdf), p.16

<sup>63</sup> [https://www.eudat.eu/system/files\\_force/EUDAT-DEL-WP6-D6%203-2nd%20Year%20Report%20on%20Operations.pdf](https://www.eudat.eu/system/files_force/EUDAT-DEL-WP6-D6%203-2nd%20Year%20Report%20on%20Operations.pdf), pp.54

resources typically via a gateway at the provider site. Each data center has its own security rules and policies; the requester has to accept these rules.

An increasing number of data projects and pilots, in particular, are demanding for this kind of service hosting and data processing capacities.

#### **4.7. Ensuring Continuing Availability and Reliability**

The operation of the CDI infrastructure relies on the continuing availability of services and resources provided by the data centers. While the service providers are responsible to make sure that they can manage potential risks of unforeseen service disruptions or disasters it is a coordination task to mitigate risks on infrastructure level. In order to raise awareness and to facilitate the knowledge and solution sharing within the CDI, the OPCT asked the partners to update the information about the strategies for high availability and reliability applied by the service providers. An overview of these strategies and measures taken by the providers is maintained on the EUDAT wiki. For each service the site deputies have provided some detailed information about the

- server failover strategy and server backup solution,
- disaster recovery strategy,
- database backup solutions,
- storage backup solutions,
- internal monitoring solutions.

In case of alarm notifications received from EUDAT's status monitoring system (<https://cmon.eudat.eu>) the OPCT and the service owners can react and assign trouble tickets to the support line of the site that is hosting the service that is failing. After the problem has been fixed, the responsible site deputy provides a short report.

## 5. DATA PROJECT ENABLING AND SERVICE INTEGRATION

### 5.1. Introduction

As discussed in section 3.1, a *Data Project* is regarded a set of tasks for planning, configuring, providing, managing, monitoring and accounting CDI services and executing workflows for customers. The time span must be defined and can be prolonged subsequently. Important for conducting a data project in a sustainable manner is the *customer* who signs responsible for sponsoring the services and resources (capacities) that are provided to him in the context the project.

*Data Projects* are business cases for providing services. The science cases that rely on CDI data projects demonstrate the impact of the CDI. The evolution of the CDI, the evolution of the landscape of services and service providers (thematic and generic), is determined by the *Data Projects* from different science domains. WP4 is in charge to foster the evolution of the CDI by interacting with the science and research communities. While WP4 acts - partly - like a pre-sales activity that discusses and helps specifying *Data Projects* or *Uptake Plans* in close collaboration with the research communities (see EUDAT2020 WP4 deliverable D4.3), it is a task within WP6, the one presented in this chapter, to realize these data projects.

A *Data Project* has a life cycle that is initiated as a *customer order*, i.e. an *organisational person*, someone who “pays the bill” makes a request to enable CDI services, eventually gets some of them customized, and gets the necessary capacity (resources) provided for the associated science and research community. This order (project request) comes with a plan that describes in textual form the scope of the data to be managed, and it defines how, by which provider, when, how long, where (geographically) and under which further conditions the data shall be managed or stewarded. *Data Projects* describe the data management and data stewarding tasks as part of a *research data life cycle* which can include one or several of the following service categories:

- Data workspaces (for mainly mutable data)
- Sharing data (mutable or immutable)
- Versioning data
- Safeguarding data
- Archiving data (immutable)
- Transferring data
- Accessing data
- Discovering data

The project enabling and service integration task, one of the main tasks of WP6, provides the B2 service expertise needed to develop and realize data projects<sup>64</sup>. The project enabling team, a team of consultants with either overall expertise or specific knowledge in the domain of B2 services, has been established and the procedures for the communication between enablers, customers and service providers, and for the documentation of the enabling activities have been defined.

The purpose of the project enabling and service integration task is to

- plan, enable and integrate the services and service components of the CDI as it is necessary for each data project. The deployment and operation of the service components are in the responsibility of the service providers. The provisioning of a distributed service such as B2SAFE is in the responsibility of a specific *general service provider*. The project enabler facilitates the interaction between the different actors. He guides, drives and supports the enabling and the integration of the B2 services.
- support the research communities in connecting their repositories to the CDI (B2SAFE), to facilitate the deployment of the EUDAT services at the community sides, if necessary, and to support the use of the B2 services also via APIs.

<sup>64</sup> Pilots are data projects with the difference that pilots can comprise services or service components which are either experimental or need to be developed by WP5. Also a significant share of the service integration work is on the side of the pilot PI.

- support the customers of the existing data projects after these projects are in *(pre-)production*.

New data projects from different research areas are being planned and several sub-projects (B2 services) are being enabled (see Annex 0). The existing projects which started in the previous phase of EUDAT (see Annex F.1) have been reviewed and maintained, that is new releases of the B2 service packages had to be installed.

The project enablers are data services experts with a technical background. Preferentially they are in contact with research communities (eventually active in WP4) and they are aware of the typical data service configuration patterns. They are also experienced with service development aspects as well as with the aspects of the deployment and management of, at least, specific B2 services.

Each data project enabling task is a challenge for the individual experts, particularly if the overall background knowledge in data science is missing or not yet there. Fortunately, some project enablers could also be recruited from the community partners in EUDAT so that enablers from thematic and generic service providers are in the meanwhile forming a collaborating team that exchanges information and knowledge. This project enabling team of experienced consultants is complemented by data services experts from the community sides, the community service integrators. These teams are working together and they use the EUDAT wiki pages, mailing lists, TTS queues and periodic virtual meetings to communicate and to channel the flow of information, to keep the progress of the enabling projects traceable and documented. As the team of consultants and service integrators is growing, sub teams can be formed that share specific knowledge in the domain of specific B2 services. These teams share their experiences under the supervision of the project enabling team leader who ensures that the data project results are adequately documented by each of the project enabler and that the projects are realized in the given timeframe.

These teams of consultants are

1. acting as point of contact for qualified service-related requests and they provide enabling support and 2nd-level support for the specific B2 service. The project enablers together with the team leader decide who will be assigned to which data project or subproject as project manager and/or enabler.
2. monitoring the health status of accomplished data projects with their B2 services instances after they have been installed and configured, at least as long as these projects have the status “pre-production”. The team leader is supervising the operations while the daily operations duty is in the responsibility of the services and resources provisioning coordinators, the site deputies and the operations coordinator whose task is strongly supported by the operational tools. 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 according to the metrics defined by SLAs, and the service enabling team members need to take action only if issues affect the quality of service and which is not or cannot be solved timely by the resource and service provider site.

Currently<sup>65</sup>, this includes to

- check that the services (as service groups) and the associated service components of the relevant B2 services (as service endpoints) are properly registered in the Site and Services registry (<https://creg.eudat.eu>), that any necessary additional property is defined (e.g. extension properties), and that the information is up to-date (this is tested by automatic checks which are sending a message to the appointed enabler).
- check that the adequate (storage) resource capacity is made available and continued for the near future; otherwise further resources are to be requested using via RCT (<https://rct.eudat.eu>).
- validate that all the services (service groups in the Site and Service registry) and service components (service endpoints) which are relevant for the data projects are monitored (via <https://cmon.eudat.eu>, soon replaced by the A&R monitoring system) and that the quality of service is appropriate.

<sup>65</sup> There will be a change of tools when the DPMT becomes available in 2016. The DPMT will supersede other information systems such as the RCT, the CREG and the EUDAT wiki.

The team of project enablers meet monthly to articulate issues concerning the enabling processes which are difficult to document elsewhere and to share information on actual solutions to problems, to refer to new configuration pattern and enabling recipes, also in order to avoid redesigning existing solution patterns. Beside the meetings, each project enabler also reports on the shared wiki pages<sup>66</sup> to keep the whole project enabling team up to date.

Figure 18 depicts the current project enabling workflow which starts with the activity of WP4 that manages the *Customer Relationship* with the research communities (the pre-sales phase). Currently the EUDAT wiki is used to register new uptake plans, data projects and pilots. The shown data project coordination portal is currently represented by this tool but will be replaced in the coming months by the Data Project Management Tool (DPMT). In this context, also the GOCDB can be replaced.

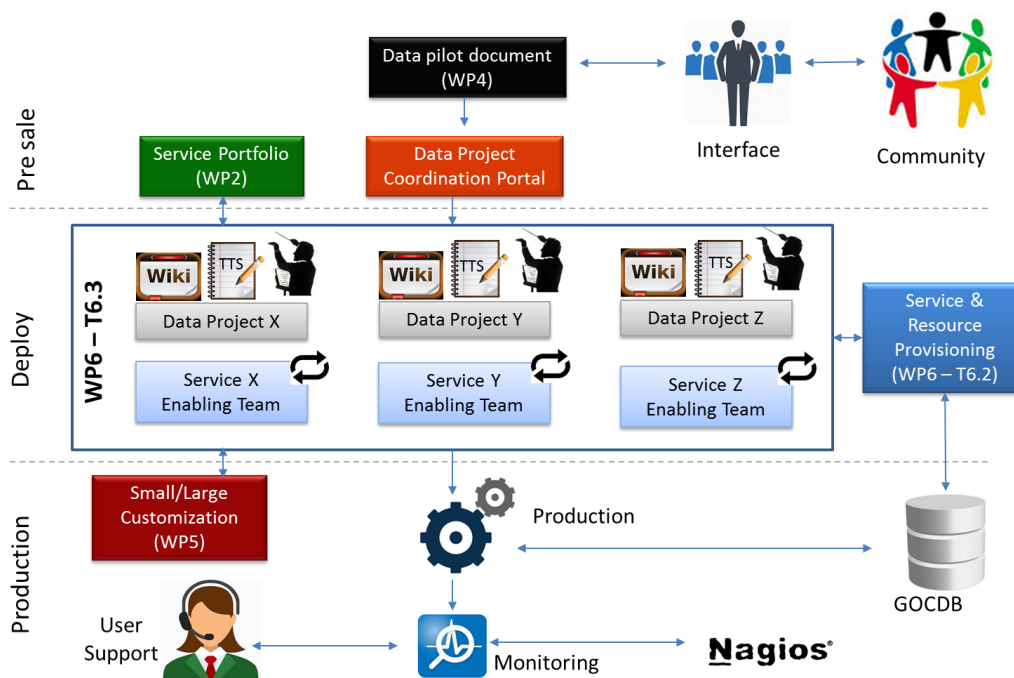


Figure 18: Project enabling workflow

## 5.2. Review of existing data projects<sup>67</sup>

During the previous EUDAT EC funded project, a number of data projects were already enabled in production or were under enabling to the use of B2 services at its end. EUDAT2020 enabling activity is thus taking over from those efforts to provide continuity to those communities still interested in making use of EUDAT CDI.

To this purpose, each community has been contacted via its Project enabler so to understand if the current status of the data project, and the community interested in continuing the EUDAT services engagement.

In case of a positive answer, more information are being collected on the further needs/requirements (i.e. no extra work needed for enabling but maintenance of availability of the service is required, or new requirements have been identified or new services are needed and new enabling work is needed by WP6).

In case the B2 services use is no more required by the community, the team tried to understand the reasons behind this decision (i.e. community changed opinion on data sharing or EUDAT services missing features etc.) and to define if roll-out plan is needed to remove data or information from the EUDAT services.

WP6 took over 37 data projects and of these only 3 were declared to be dismissed (“Terminated”) by the reference community, 4 need to be postponed/suspended to wait for service updates or the community to

<sup>66</sup> This approach will change when the DPMT becomes available.

<sup>67</sup> From the previous EUDAT project that ended in March 2015.



be ready. All other data projects will be followed up in EUDAT2020 with 4 of them replaced by new Uptake plans or data pilots.

**Table 10: Result of the review of the existing data projects from EUDAT 1**

Status	# of data projects
Terminated	3
Enabling	2
Postponed	2
Pre-Production	4
Production	20
Replaced	4
Suspended	2

A table on already existing data projects updated status is presented in Annex F.

### 5.3. Uptake Plans and Service Integration

In this section we report on the enabling activities carried out in relation to the uptake plans from the core communities. Relationships and uptake plans definition is part of WP4, and the activities are moved to WP6 when technical requirements are well enough defined and the projects can start to be planned in detail.

The project enabling and service integration in the first year concerned EPOS, LTER and CLARIN communities.

#### 5.3.1. EPOS

The European Plate Observing System (EPOS, <http://www.epos-eu.org>) is the integrated Solid Earth Sciences Research Infrastructure. EPOS has a long-term integration plan of existing national and international research infrastructures.

The mission of EPOS is to integrate the diverse and advanced European Research Infrastructures for Solid Earth Sciences to monitor and unravel the dynamic and complex Earth System. For more information on the specific Uptake plan please see D4.3.

In summary the EPOS use case suggests the requirements for a system able to:

- enable discovery based on multi-layered metadata: from high level discovery metadata to domain specific metadata (Req 1)
- identify, integrate and link data and products by means of PIDs (Req 2)
- enable data processing and visualization on distributed resources (Req 3)
- collect and manage provenance information (Req 4)
- bundle, publish and share results (Req 5)
- enable citation and reproducibility of science (Req 6)
- provide accountability and usage statistics (Req 7)

The following Figure 19 represents the mapping of EPOS requirements to the B2 services. The diagram and requirements catalogue may be useful as blueprint also for similar RI data management plans.

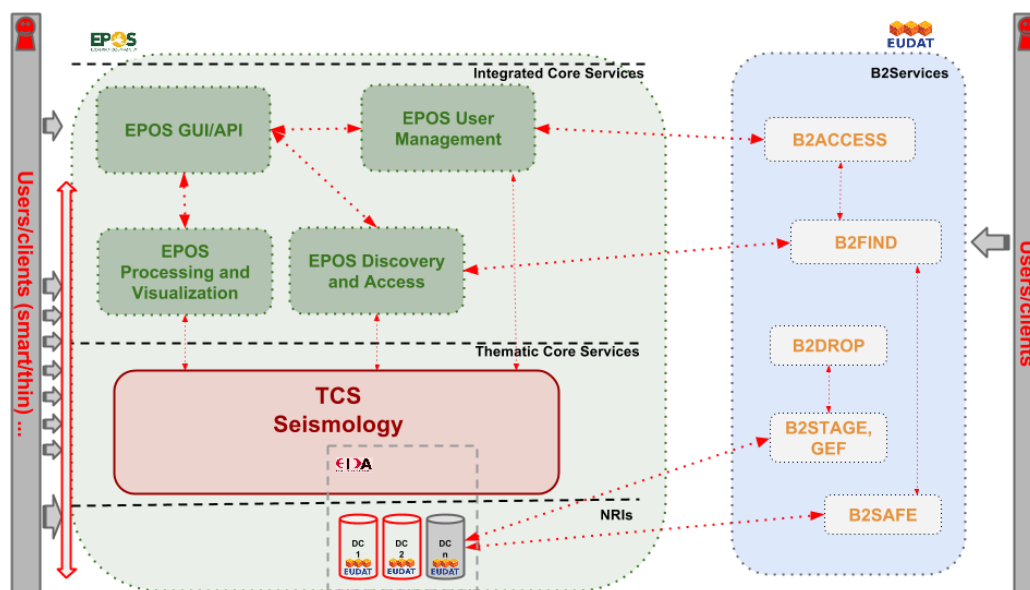


Figure 19: Schematic integration plan for EUDAT services in the EPOS-ORFEUS-EIDA pilot

### Project enabling and service integration

The EPOS community services enabling and integration started on April 2015 and will last until 2017. The enabling of B2SAFE, B2ACCESS, B2STAGE, B2DROP and B2FIND is on the agenda. The first period activities were focussed on the B2SAFE service with the three community nodes INGV, GFZ, and KNMI. Concerning B2SAFE, all three nodes are *joining* the CDI according to a comparable scenario. Therefore we explain the B2SAFE integration of the INGV node representatively for the other two nodes. Main difference is the first CDI ingest nodes: INGV works with CINECA, GFZ with KIT and KNMI with SURFsara.

#### Node: INGV

EPOS member INGV uses the EUDAT infrastructure to replicate its archive of raw waveform data (currently 50 TB, millions of files) to the EUDAT CDI. The corresponding CDI ingest node is provided by CINECA.

One important aspect of the use-case, a work in progress, is to link the metadata to the corresponding waveform files in order make data discovery easier. This assignment of metadata with data is planned to be realized by using iRODS.

Authentication and authorization is currently realized via iRODS: user data and ACLs are stored in the iCAT DB. The CINECA zone supports GSI authentication and OS authentication. Data inside the archive at INGV (in this particular case stored in a SAN) are registered via the “ireg” command in the iRODS instance that had been deployed at INGV. In this context feedback has been provided to the B2SAFE service developers which led to an improved and more efficient iRODS installation procedure. The data is synchronized daily via “irsync” with the iRODS instance at CINECA. Figure 20 shows, as an example also for the other EPOS nodes, the setup of the B2SAFE service components at the thematic and the generic service provider site.

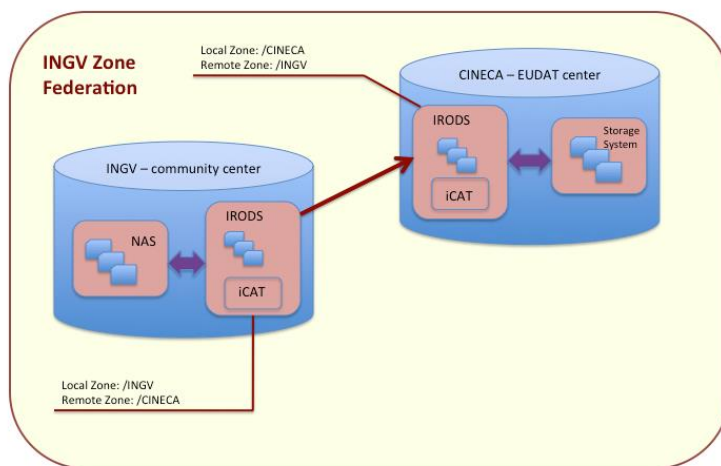


Figure 20: EPOS-INGV use case

#### Node: GFZ

EPOS member GFZ uses the EUDAT infrastructure to replicate its archive of raw waveform files (currently 70 TB, 27 million files) to the EUDAT CDI. The corresponding CDI ingest node is provided by KIT. The use case and configuration pattern is similar to the setup for INGV.

Different from the other EPOS nodes in EUDAT, GFZ has deployed their own ePIC Handle server and is minting their own PIDs. GFZ improved the performance for minting big amounts of PIDs in the context of B2SAFE. This knowledge has been fed back to the B2SAFE development team. THE GFZ PID service is intended to become part of the B2HANDLE. It will be used also for minting the PIDs for the data replicas stored at KIT.

#### Node: KNMI

EPOS member KNMI uses the EUDAT infrastructure to replicate its archive of raw waveform files (currently 30TB, millions of files) to the EUDAT CDI. The corresponding CDI ingest node as well as the PID service is provided by SURFsara. The use case and configuration pattern is similar to the setup for INGV.

#### 5.3.2. ICOS

The ICOS RI<sup>68</sup> has a long-term commitment to providing standardized, state-of-the-art observational data on greenhouse gases and ancillary parameters to all interested parties. ICOS relies on observation stations operated by its member nations. Thematic Centres (one each for Atmospheric, Ecosystem and Ocean themed data) perform quality control and aggregation of the sensor data from the stations, after which ICOS data products are sent to the Carbon Portal for curation, storage and dissemination to interested users. The ICOS ERIC is a partner in EUDAT2020. More information on the specific uptake plan of the ICOS RI community is provided in the WP4 deliverable EUDAT2020 D4.3.

Figure 21 shows the data flow within the ICOS RI. It is indicated which B2 services are intended to be use at which location in the diagram.

<sup>68</sup> <https://www.icos-ri.eu/>

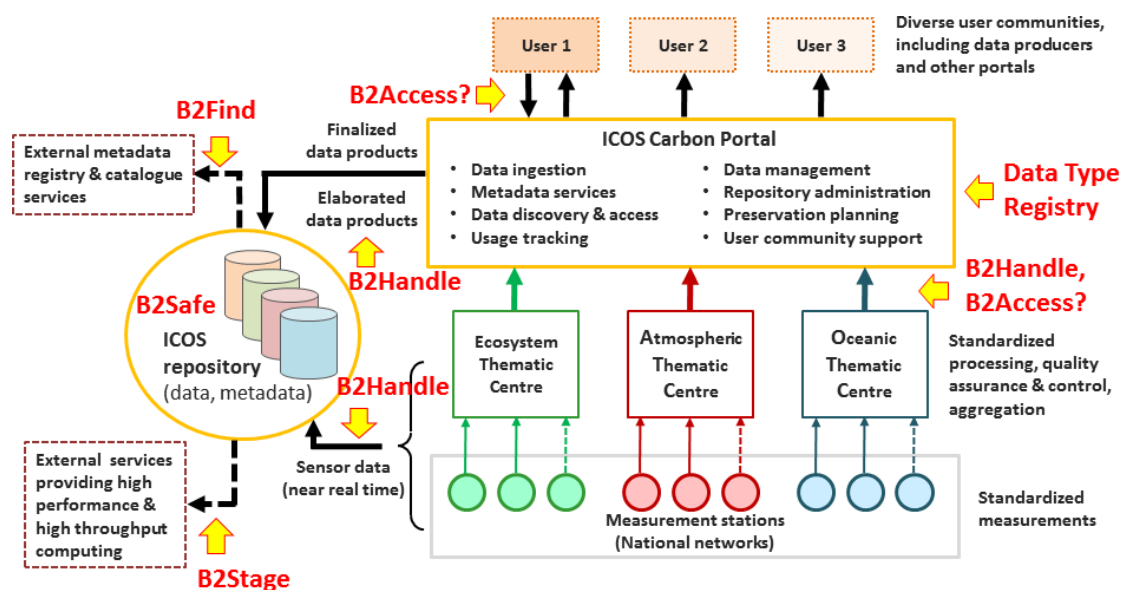


Figure 21: Data flow in ICOS together with the relevant B2 services that ICOS intends to use.

### Project enabling and Service integration

The project is still in the planning state. First interactions started the summer of 2015 between the Carbon Portal IT personnel (in Lund, Sweden) and experts at PDC-KTH (in Stockholm, Sweden) concerning the B2HANDLE service (PID service support provided by PDC) since PIDs will play a fundamental role for the data management in the ICOS RI. The (internal) data transfer mechanism to ICOS is being developed, including the backbone metadata store for all ICOS data objects. In these cases EUDAT facilitates valuable knowledge transfer while EUDAT profits from the domain specific data management expertise from the ICOS community.

Central to ICOS data management is the ICOS Repository - a long-term archive of data and associated metadata. ICOS will keep local copies of most (non-raw) data objects in a storage facility based on B2SAFE - including raw sensor data, finalized observational data, and elaborated data products. ICOS is building up its own metadata database for all ICOS data objects based on semantic technology. This cataloguing database will form the backbone of the Carbon Portal's own data discovery & visualization tools.

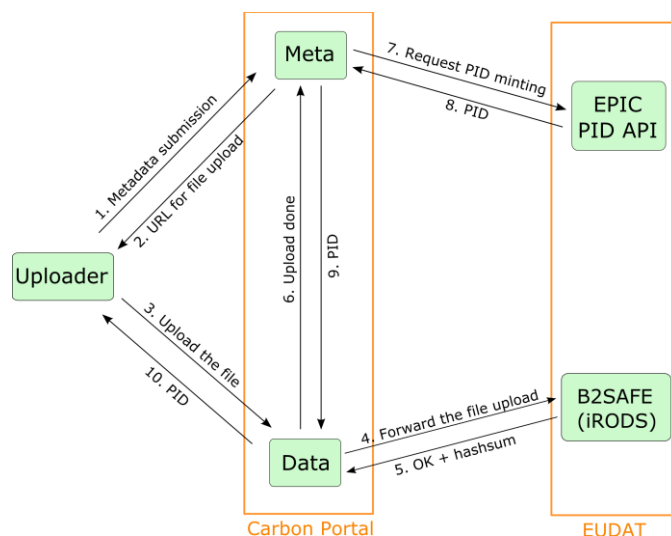


Figure 22: Schematic ICOS data flow (preliminary)

In this early implementation planning phase, it is envisaged that all data and metadata object transfers will be made via the Carbon Portal as the intermediary between the Uploader (ICOS-internal or external data

producer) and the EUDAT B2SAFE service. As the project matures it is expected that other ICOS nodes, specifically the Thematic Centers, will also be ingesting some types of data objects, such as raw sensor data or near-real time data into B2SAFE, following a similar procedure.

Figure 22 shows the schematic of the currently planned setup that involves in the backend of the Carbon Portal basically the B2HANDLE (PID minting) and the B2SAFE services.

### 5.3.3. LTER

The Long Term Ecological Research community Europe (LTER Europe, <http://www.lter-europe.net/>) is a network of around 455 long term research sites (420 LTER sites and 35 LTER Platforms) in 22 national LTER networks in Europe. LTER is an essential part of world-wide efforts to better understand ecosystems and it contributes to the knowledge base informing policy makers. More information on the specific Uptake plan of the LTER community is provided in the WP4 deliverable EUDAT2020 D4.3.

#### *Service integration*

In the first period activities have been focused on connecting the community service DEIMS with the B2SHARE service via the B2SHARE API. This requires a custom Drupal module to be used in DEIMS<sup>69</sup> which is developed by the associated community service integrator.

### 5.3.4. CLARIN

CLARIN is the research infrastructure for providing language resources (data) and language technology (LT, services) also to the wider Humanities, e.g. technologies for advanced and sophisticated text mining services.

The CLARIN ERIC is a partner in EUDAT2020 and the CLARIN RI comprises currently 33 CLARIN centers. Each one is required to have a CLARIN certified repository holding linguistic type of data and/or running LT services. More information on the specific Uptake plan of the CLARIN community is provided in the WP4 deliverable EUDAT2020 D4.3.

#### *Project enabling and Service integration*

Beside the still ongoing the B2SAFE-related data projects from EUDAT1 (see Annex E1), with the repositories of MPI-PL TLA, EKUT and CUNI (LINDAT), the enabling activities based on the CLARIN uptake plan started recently and first focuses on the enabling of the B2SAFE service for the SOAS<sup>70</sup> Endangered Language Archive.

## 5.4. Data Pilots and Project Enabling Activities

The 23 data pilots from the first call from WP4 have been passed over to WP6 to be enabled as data projects. Data pilots are different from data projects in that they typically require services to be newly or further developed by WP5.

To allow a more efficient allocation of the enabling team time and of the resources a progressive start of the enabling for the data pilots have been put in place. The selection was based on the readiness of the community, the distribution on different reference centres while trying to cover most of the scientific domains in every phase.

The progressive enabling roadmap has planned:

- 6 data projects were selected to start on Jan 2016
- 6 new data pilots are selected to start in Feb 2016
- 6 new data pilots to be started in March 2016
- 6 new data pilots to be started in April-May 2016

When a data project is planned to start:

- A Project Enabler (or enabling team) is assigned to interact with the community,

<sup>69</sup> [https://data.lter-europe.net/ilter\\_deims/](https://data.lter-europe.net/ilter_deims/)

<sup>70</sup> <http://www.clarin.ac.uk/endangered-language-archive-elar-soas>

- Availability of the reference centre is verified and recourses allocated,
- A timeline of intermediate achievements is defined together with a first milestone,
- The technical enabling starts.

Among all EUDAT services, B2-FIND and B2-SHARE are those most requested, but all services will require some level of enabling in the new data projects.

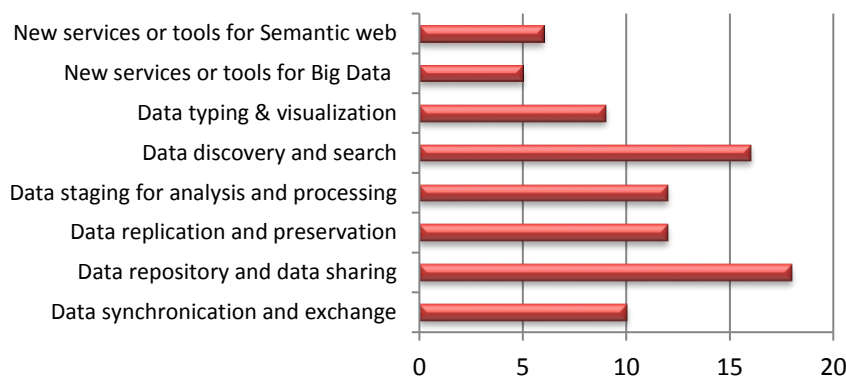


Figure 23: Distribution of service requests by the data pilots

All reference centers will be involved in the provision of storage resources for a total requested between 1220 and 4300 TB.

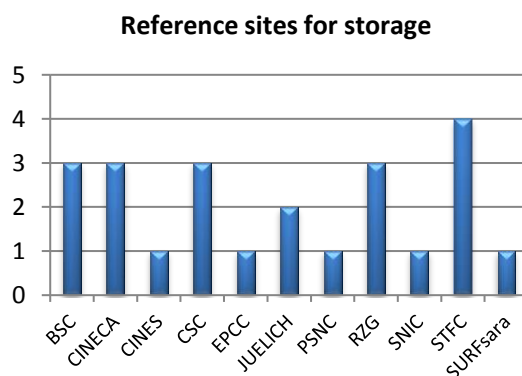


Figure 24: Storage Service Provider requested by the data pilots

A complete table with the new data projects originated from the data pilots is presented in the table in Annex B.

## 5.5. Templates and Recipes for the Project enablers

The result of an enabled data project is generally a customized suite of services with the requested resource capacity. During the first project year plenty of use cases have been documented in uptake plans and data pilots (see the EUDAT2020 WP4 deliverable D4.3). They indicate that different configuration options are required within a single B2 service domain as well as across the B2 services.

We have started to study these configuration requirement per B2 service and developed recipes for the (sub)project enabling. These will be translated into guidelines for the consultants. Figure 25 shows the decision graph which is used by the DPMT to select the adequate B2 service template.



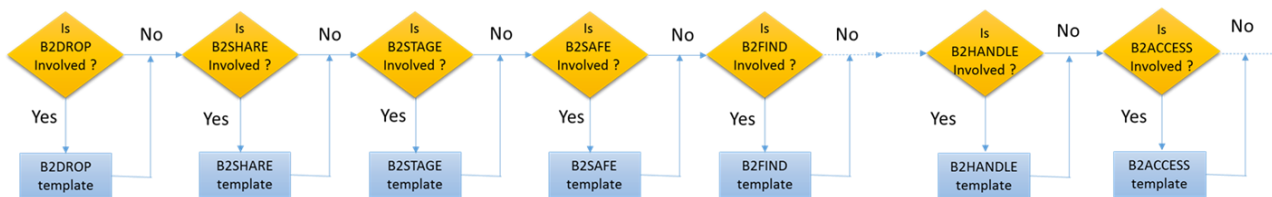


Figure 25: Decisions at the beginning of the B2 service enabling process

The B2 service specific template takes into account the service-specific configuration options. Such a template can be regarded as a service specific questionnaire or a dashboard that allows to specify different service-specific options. The classification of the configuration pattern is an ongoing activity. In order to explain the approach and preliminary results the example for the enabling of B2FIND is explained.

The B2FIND service allows to collect meta data from a variety of repositories. This is currently done on the basis of the B2SHARE metadata templates (i.e. harvesting the EUDAT B2SHARE instance) or the research community repository is harvested.

In order to publish the metadata two prerequisites must be fulfilled by the repository: (i) a service for providing and transferring the metadata, (ii) the definition of a community and domain specific mapping of the meta data using a normalised terminology.

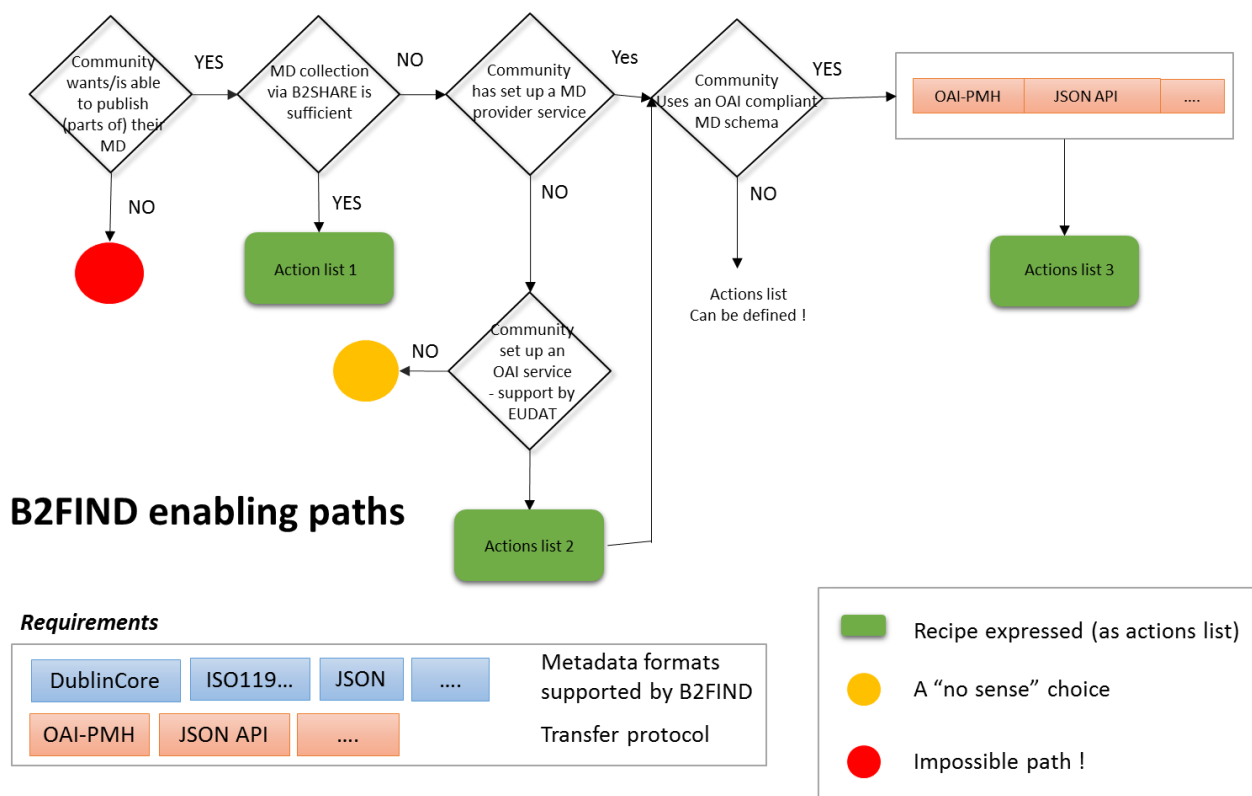


Figure 26: Decision graph for the enabling of B2FIND

Typical questions for B2FIND template are:

- Are metadata collected by B2SHARE or community repository? (B2SHARE/ Community repository)
- Which service or protocol is used to expose the metadata?(OAI-PMH, JSON API, others, none)
- Which metadata schema is used for the exposed metadata? (Existing metadata standard (as DublinCore, ISO11919, MarcXML, ...) / Own, community specific schema, None)

- Are there other community specific requirements w.r.t. metadata? (e.g. specific fields or attributes)
- Is metadata mapping necessary? (Yes/No)

The decisions often not independent and the final result can be a list of actions (the recipe for a specific constellation). Figure 26 shows the corresponding decision diagram.

### **B2FIND enabling actions list#**

1. Actions list
  - a. Check harvested MD records from B2SHARE
  - b. Discuss MD quality with community
2. Actions list
  - a. Discuss with community the easiest way to set up a data provider service (preferable OAI-PMH)
  - b. Give support during implementation and test phase  
(includes consulting concerning the appropriate MD schema etc.)
  - c. Go to Actionlist 3
3. Actions list
  - a. Perform initial harvesting of some samples (planning phase)
  - b. Fix issues concerning OAI-PMH (rsp. Other transfer protocol)
  - c. Define minimal, basic mapping and upload to a B2FIND test instance
  - d. Discuss 'mapping', visibility and data access etc. pp. with community (pre-production)
  - e. Send B2FIND mapping spreadsheets to community
  - f. Community representatives fill out 'General information' and 'MD specification'
  - g. Negotiate correctness of OAI endpoints and subsets (which part of available MD should be published)
  - h. Go into production
  - i. After 'ok' from community side, upload MD to the productive B2FIND catalogue
  - j. Stabilize ingestion and configure incremental and regular ingestion
  - k. Iterative enhancement of MD mapping, extent and quality ...

## 6. HELPDESK AND SUPPORT SERVICE

The purpose of the *Helpdesk and Support task* (T6.4) is (1) to coordinate the support infrastructure for the CDI, (2) to operate and maintain the EUDAT Helpdesk system, a ticketing system based on Request Tracker with interfaces to the support and contact form, and (3) to provide the 1<sup>st</sup> level support and dispatch incoming request to the adequate service expert teams and other *downstream support teams*. The task is designed to fulfil the helpdesk requirements according to OLAs and project-specific SLAs<sup>71</sup>. The EUDAT Helpdesk guidelines<sup>72</sup> apply as long as OLAs and SLAs are not in place. The different levels of support are managed by dedicated teams which are using the ticketing system [helpdesk.eudat.eu](https://helpdesk.eudat.eu) that is accessible by all EUDAT partners. The support task is an important interface for users, customers and service providers and it provides a centralized way for communicating with EUDAT on different levels, including the 1<sup>st</sup> and the 2<sup>nd</sup> level support. We discriminate between the provisioning of the *Helpdesk system* which includes the provisioning of the technical tools from the Support activity which deals with issue reports and requests on different levels. There are three support levels provided by EUDAT.

### 6.1. First level support

A dedicated First Level Support team is responsible for handling all incoming issue reports, support and contact requests which are normally received via the *EUDAT support and contact form*. The 1<sup>st</sup> level support is the initial point of contact for stakeholders and customers of the EUDAT CDI, i.e. customers of existing or planned data projects, community and data managers and the users.

From the answers provided by the 1<sup>st</sup> and 2<sup>nd</sup> level support frequently asked questions (FAQs) are created. The 1<sup>st</sup> level support provides basic information about the project, the services and how to use them. In addition, requests and issue reports are prioritized, classified, eventually clarified by contacting the requester and forwarded to the adequate 2<sup>nd</sup> level support teams. The first level support is answering with a short response time.

The 1<sup>st</sup> level support acts as a bridge to the 2<sup>nd</sup> level support by clarifying incoming requests which are too vague: standard questions are asked about more details if issue reports or requests are too unspecific, by that, requests are either filtered or enriched with further information before they are forwarded to the 2<sup>nd</sup> level support.

The 1<sup>st</sup> level support is responsible for managing the Helpdesk system and for monitoring the overall efficiency of the Helpdesk. In order to keep response times short, it is in charge to assign owners to tickets, remind ticket owners and escalate if tickets are neither closed nor touched for a longer period of time.

### 6.2. Second Level Support

Responsible for the Second Level Support are Data Project Enablers, Service Integrators, Service Owners, Service Area Managers, Site Support contacts, and Service specialists from the Provider of the service. The 2<sup>nd</sup> level support provides detailed responses to requests and issue reports concerning specific services, the service catalog, data projects, service design and feature requests. The 2<sup>nd</sup> level support is organised such that (1) data project enablers and dedicated B2 service integrators are responsible to response requests related to services or data projects, and that (2) service providers are answering requests related to services, service components and resources in their own domain. When a request ticket has been forwarded from the 1<sup>st</sup> level support, the 2<sup>nd</sup> level support responds and starts interacting with the requester in order to solve the issue. It is also acting as a bridge to the 3<sup>rd</sup> level support, in particular if bugs and issues are reported which can only be solved on the level of service and software development.

In order to be able to interact efficiently with the 3<sup>rd</sup> level supporters, the managers of the service-specific ticketing queues are in charge to translate the tickets into the specific format that is expected by the

<sup>71</sup> See EUDAT2020-DEL-WP2-D2.1, Annex A.4

<sup>72</sup> See <https://www.eudat.eu/d63-2nd-year-report-status-and-progress-operations>, pp.47

corresponding development team. This is important since each development team can use a specific software repository and issue tracking system such as GitHub. The 2<sup>nd</sup> level supporters are providing implicitly or explicitly the necessary details for the FAQ.

### 6.3. Third Level Support

Responsible for the Third Level Support are the Service Area Managers and Service developers (WP5).

The 3<sup>rd</sup> level support is handling bug and issue reports concerning the existing services, and they may register requirements for the developments of features or new EUDAT services. The 3<sup>rd</sup> level support is not necessarily responding to the original requester. Direct interactions or feedback are justified if the requests are made in the context of data pilots or community uptake plans that require new or modified services.

### 6.4. Helpdesk System and Statistics

The central point of contact for issue reports and requests from customers and users concerning all kind of services and data projects is the online contact form [www.eudat.eu/support-request](http://www.eudat.eu/support-request). It is part of the EUDAT website, [www.eudat.eu](http://www.eudat.eu), and links to this form are also provided on the front pages of different B2 services, e.g. [B2ACCESS](#), [B2DROP](#), [B2SHARE](#) and [B2FIND](#). The support web form can be used to submit any issue reports, support or contact requests. It has indeed become an important channel for communities and organisations to establish first contacts to EUDAT. Requests can be marked by the requester as being related to a specific community (science field) and B2 service. This facilitated the classification of incoming requests by the 1<sup>st</sup> level support. When a request is submitted via the online form, a ticket is created in the main support queue of the EUDAT helpdesk system ([helpdesk.eudat.eu](http://helpdesk.eudat.eu)) and the requester obtains an automatic confirmation. It is an internal issue tracking system using open source software named “RT-Request Tracker” by Best Practical Company.

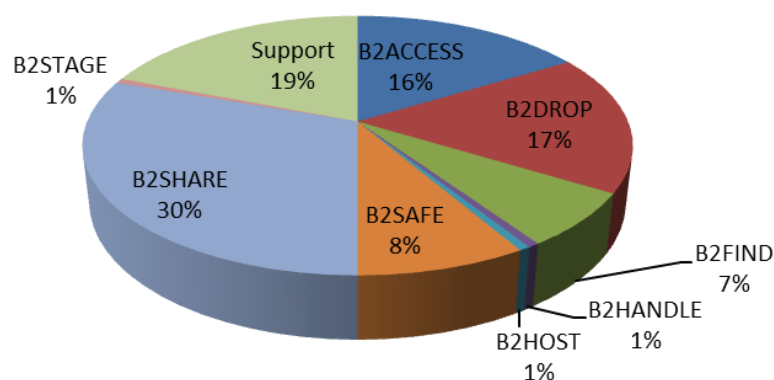


Figure 27 User tickets per service queue

During the first 12 months of the project 110 requests have been managed by the EUDAT support service. Figure 27 shows the distribution of the requests for the different services supported by the EUDAT infrastructure. In this case, the service with most requests is B2SHARE, followed by B2DROP, it is because they are the two services most used in the infrastructure, another important part of the requests are general or first level questions and they are managed directly by the user support team.

Figure 28 shows the incoming requests per service and month that have been managed by the 1<sup>st</sup> level support between Jan 2015 and Feb 2016. Recently, after the 1<sup>st</sup> User Forum in Feb 2016, we noticed an increase of requests concerning B2 services in the context of the data pilots.

The amount of requests increased during one year except the summer period, also it is important to notice the increase of tickets in the month when the call for Data Pilots was open (15<sup>th</sup> May) and around its deadline (9<sup>th</sup> October 2015) as the user support team was often contacted by interested parties.

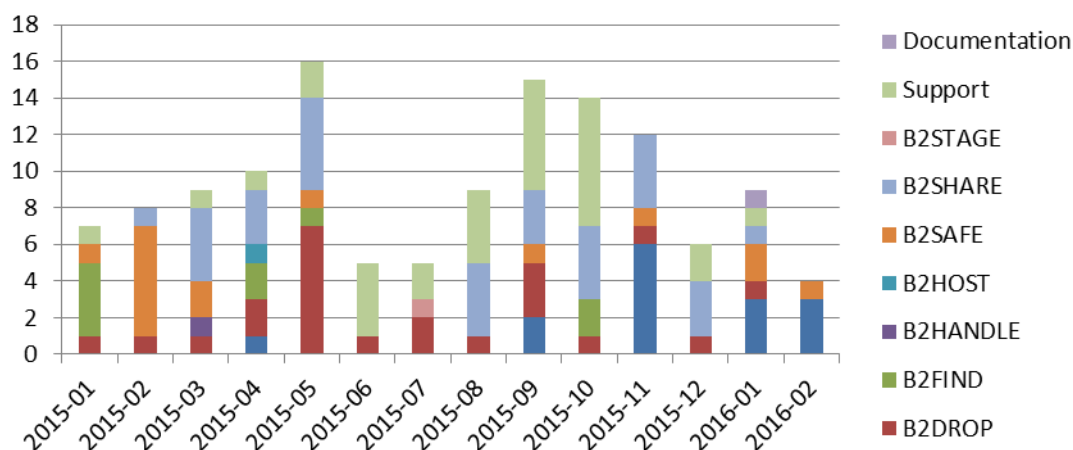


Figure 28: Support Requests managed per service per month (cleaned statistics)

## 6.5. Support activities

During the first 12 months of the EUDAT2020 project the User support team has become more important since an increasing number of users using the B2 services. The launch of the 1<sup>st</sup> EUDAT2020 call for collaboration projects was another reason for more requests from the users.

One task of the support team is managing the Helpdesk system for the EUDAT. An important activity was the upgrade of the RT Trouble Ticketing System from version 3.8.8 to version 4.2.

## 7. SECURITY

As the EUDAT2020 project has evolved from defining new concepts to produce operational services, the role of security within the project has also become more operational. The objective for the security function within the project has been to support and safeguard the project according to best security practices. Although several specific security controls, such as vulnerability and incident management have been applied, the foundation for security within the project relies also on good governance and mature IT service management. A reasonable amount of transparency and coordination of development projects and lifecycle management of operational services is also crucial for adequate security.

When EUDAT launches services that are described as *Secure*, it is important that the project can ensure that the services are really developed, tested, deployed and managed in reasonable secure manner according to well-known best practices. Security must be taken into account by design.

People are the key also in security, in addition to secure design and good governance. Therefore it is important that the project management supports security awareness building.

### 7.1. Vulnerability Management

Several information security vulnerabilities and weaknesses have been found during the project so far. Information about vulnerabilities has been disseminated and coordinated between sites, other research infrastructures and the open source communities. The premise for EUDAT operations is that the sites and communities already have adequate fundamental security practices in place and appropriate measures are taken by the sites by default.

A typical security vulnerability was CVE-2015-0235<sup>73</sup>. This vulnerability was a weakness detected in glibc, the GNU Project's implementation of the C standard library. By exploiting this vulnerability an attacker could execute arbitrary code via the standard function call `gethostbyname` bypassing authentication. Easy to use exploit tools were publicly available for this vulnerability.

CVE-2015-0235 was classified as a critical vulnerability for EUDAT and most providers delivered security patches in a timely manner. Sites responsible for providing EUDAT services and infrastructure had to ensure that the security patches were installed by risk that an attacker otherwise could exploit the vulnerability and endanger the confidentiality, integrity or availability of EUDAT services.

Vulnerabilities like CVE-2015-0235 are common. EUDAT Security Officers evaluate new vulnerabilities on weekly basis. Unfortunately it is not trivial to identify whether the vulnerability really threatens the project sites, as there is often uncertainty and outright disinformation about the risks of vulnerabilities.

Traditionally collaborating research infrastructures such as sites within EUDAT trust each other to be adequately managed also in terms of operational security. There is a trust that the sites by default allocate enough resources for system administration to avoid unnecessary risks caused by information security vulnerabilities or weak configurations. The same applies to software development processes, in academia there is a trust that the developers will consider risks related to weaknesses and harden the applications sufficiently by for example input sanitation and other well-known security controls. However, it might be that academic operational procedures in software development are not always as mature and comprehensive as in software industry.

Experience showed that sufficient security controls have not always been applied efficiently within EUDAT or within other sites or research infrastructures. A good example about this is the so-called POODLE vulnerability, CVE-2014-3566, and a few others which were detected in October 2014. The original POODLE vulnerability gives an attacker a possibility to reveal encrypted traffic. Sporadically we found the POODLE vulnerability still on several hosts in EUDAT and community sites and urged the responsible operations staff to take action.

---

<sup>73</sup> <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2015-0235>



EUDAT share most of the vulnerability risks with other research infrastructures, particular risks related to open source software, which are widely used for research. Therefore, EUDAT security officers are cooperating closely with peers from other infrastructures especially in Europe but also with international peers. Vulnerability and incident coordination is done in a close cooperation with EGI, PRACE, and GÉANT.

## 7.2. Security Scans

To proactively improve security and support the project, regular application vulnerability scans have been performed targeting EUDAT production services, services under development and operational services.

The scans are done both automatically and manually (penetration testing) with a set of commercial and open-sources tools. A scanner called Nessus has been used for the initial scans. The scans are currently done by a CSC subcontractor, by CSC or by JUELICH.

In the course of the scans several vulnerabilities have been detected and the scanning reports have been submitted to the concerned sites including suggestions on how the detected vulnerabilities should be fixed. Also, several efforts have been made to have security considerations included as security by design in earlier stages of development of new B2 services. This is addressed under the topic security in software development.

EUDAT services have been reviewed for security vulnerabilities including B2ACCESS, B2DROP, B2FIND, and B2SHARE.

Typical vulnerabilities found have been vulnerabilities in the underlying hosting platform, such as the POODLE vulnerability, weak passwords, cross site scripting vulnerabilities, cookies without the *secure flag* set, and unintended disclosures of configuration files.

Regular vulnerability scanning is a good and productive security measure within EUDAT. For the setup of the scans, the information about the interfaces has been used from the central site and service registry (CREG). One challenge is the systematic planning for these scans. For the next period we expect new B2 services being introduced and other instances of B2 service will become productive.

## 7.3. E-infrastructure Security

EUDAT support digital certificates from well-known trusted sources. A challenge are different certificate practices in different countries and among different user communities. The perhaps mostly widely used certificates used by EUDAT sites are certificates issued through GÉANT, which has an agreement with the US based DigiCert, Inc. to issue host and user certificates. With this new subcontractor the cryptographic quality of the certificates has been greatly improved, as certificates with SHA-1 or MD5 weak signature algorithms are not issued anymore. In addition for being insecure, certificates with weak algorithm gives a bad user experience as current browsers notify end users about the security issue.

Among the EUDAT Community members from Germany Certificates from DFN Global are used, and some sites also still use certificates from EUGRID PMA / IGTF. EUDAT supports all these three CA sources.

## 7.4. Security in Software Development

EUDAT is still in an early stage of development on security in software engineering and more effort has to be invested in this area to ensure systematic security checks during the service development lifecycle.

The security team helped with the review of the B2ACCESS design and assessed the configuration and policies for the production service. The software (Unity) itself had previously been reviewed by PL-GRID.

## 7.5. Building Operational Trust and Compliance - DSA

A possible way to support good governance and sustainable management of IT services are to follow well-known benchmarking frameworks for information security management and for governance in general.

Research infrastructures have developed the SCI security framework (SCI - A Trust Framework for Security Collaborating among infrastructures - <https://www.eugridpma.org/sci/>). The SCI framework has been found to be a reasonable good and feasible checklist for developing security in EUDAT. EUDAT is not at least yet ready to comply with industrial level security standards such as ISO/ IEC 27001. An effort to build compliance according to SCI principles was discussed with the EUDAT task 2.2 on *Trust and Certification*. This task decided to focus first on the methodology defined by the *Data Seal of Approval* (<http://datasealofapproval.org/>) which supports security through good governance by the requirements for documented processes, explicit work flows, and due diligence with legal requirements and agreements. However, the EUDAT Security Officer facilitated the information and knowledge exchange between *Trust and Certification Task* and operations. This interaction is work in progress.

## 7.6. WISE - Liaison with other Research Infrastructures

EUDAT shares many information security risks and related security controls with other related research infrastructures. Joint efforts have been made to share best information security practices, deliver security training and develop a security framework. GÉANT provided resources to facilitate these joint activities (<https://www.terena.org/activities/ism/wise-ws/>). Information security management practices have been further developed under the name of 'WISE' (<https://wise-community.org>).

Within WISE, infrastructures established working groups for updating the SCI framework, for Security Training and Awareness, for Security Review and Audit, for Risk Assessment and Security in the domains of Big and Open Data.

## 7.7. Security Policy and AUP

EUDAT has a short but comprehensive security policy according to best information security practices. The policy defines the generic security requirements and the organisation of security related roles.

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.

An acceptable use policy, conforming to the AUPs of other scientific infrastructures, has been included in the EUDAT Terms of Use<sup>74</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 to be accepted by the users of EUDAT services. Acceptance of the ToU is anticipated if the EUDAT services are used, in particular when users create new EUDAT accounts via B2ACCESS they are explicitly asked to agree with the ToU.

## 7.8. Joint Security Training

Security training has been delivered jointly with other European e-infrastructure on October 2015<sup>75</sup>. EUDAT has contributed by providing presentations about how to implement security<sup>76</sup> and on how to share best security practices among sites<sup>77</sup>.

## 7.9. Risk Management

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

<sup>74</sup> See EUDAT D6.3 Annex B or as specifically shortened and adapted versions: <https://b2access.eudat.eu/files/terms-of-use.html> (for B2ACCESS), <https://b2drop.eudat.eu/themes/b2drop/terms-of-use.htm> or <https://b2share.eudat.eu/docs/b2share-tou>

<sup>75</sup> <https://www.terena.org/activities/ism/wise-ws/agenda.html>

<sup>76</sup> [https://www.terena.org/activities/ism/wise-ws/slides/EUDAT\\_Security\\_kaila\\_WISE-2015-10-20.pdf](https://www.terena.org/activities/ism/wise-ws/slides/EUDAT_Security_kaila_WISE-2015-10-20.pdf)

<sup>77</sup> [https://www.terena.org/activities/ism/wise-ws/slides/Sharing\\_Best\\_Practices-kaila\\_WISE-2015-10-22.pdf](https://www.terena.org/activities/ism/wise-ws/slides/Sharing_Best_Practices-kaila_WISE-2015-10-22.pdf)

## 8. FUTURE WORK

The previous sections outlined work that has been done during the last EUDAT2020 project year. The following summarizes the future work which is planned for the 2<sup>nd</sup> project year.

### 8.1. Central Services and Operational Tools

The authentication via B2ACCESS will be enabled also for EUDAT'S collaborative tools such as the EUDAT confluence wiki, JIRA and SVN.

The DPMT will be in production in the second year and it will be used for the service management as well as for the data project enabling. After a short transition phase, the RCT (storage and compute resources registry) and the GOCDB (site and service registry) will be replaced by the DPMT. The DPMT supports the workflows and project stages as described in Annex C. In production the tool fetches periodically the service descriptions from the SPMT, gets notified about the latest changes and will compare the version numbers of the latest B2 service software releases with the versions actually installed by the CDI sites. Service owners and managers will be notified about differences above a given threshold. The DPMT will mimic the GOCDB API (based on GOCDB version 5.4) as far as required by the A&R monitoring system. Appropriate schema changes and extensions of the DPMT will be implemented as far as necessary. The support of recipes for the project and service enabling will be advanced. The DPMT is based on Plone using Python. The developed software will be released on Python's Package Index<sup>78</sup> for ease of reuse.

The SPMT will be in production and used by service owners and service developers to describe the services and for maintaining the service descriptions. The API allows the embedding of information from the service catalogue into other webpages. The administration web interface will be available for the service portfolio manager. The SPMT will be integrated with the B2ACCESS.

The Site & Service Registry (<https://creg.eudat.eu>) as configuration information service will be realized in first instance by the DPMT which replaces the GOCDB. For interoperability reasons the DPMT offers an implementation of the GOCDB API. By default the DPMT provides information via the JSON format.

The A&R monitoring system will be enhanced to support definitions at sub-service endpoint level and the development of monitoring probes for the B2 services will be coordinated. The Web UI will be integrated with B2ACCESS so that customers and community managers are able to read specific reports. The corresponding authorization controls will be implemented on report level.

An advanced accounting information aggregation service with an API will be available. Programmatic examples will demonstrate the use of the client scripts for uploading the accounting records. The accounting service will be integrated with the DPMT, and authorized users will be able to obtain accounting information related to a specific provider, a specific data project or they can get information about the overall resources usage.

Beside the operation, maintenance and support of the B2ACCESS service, the integration with further B2 services and operational tools will be supported as well as the inclusion of further IdPs as far as just configuration and policy aspects matter.

Beside the operation, maintenance and support of the B2HANDLE service, further B2HANDLE service providers will be included. The assignment of handle prefixes will be better balanced. The PID structure (support of the new EUDAT PID record format) will be upgraded.

### 8.2. Service and Resource Provisioning

The daily operational processes will be continued under the coordinating guidance of the operations coordination team which includes at site-level the installation and maintenance of software packages, the monitoring of the services and their components, the provisioning of storage and compute resources

---

<sup>78</sup> <https://pypi.python.org/pypi>

ensuring that all critical information is backed up and that adequate disaster recovery plans exists. The operations coordination team will take care for the service management for projects (that is the project-specific suite of services) which are in production.

Apart from that the following activities are in the focus of the next 12 months: The service management policies including those for configuration management, capacity management, continuation management, change management, and release and deployment management will be further developed and the management processes will be consolidated by using the Data Project Management Tool as tool for the project-oriented coordination of the provisioning of service and resources. The DPMT will be used to notify service owners and sites about new software releases and any critical discrepancy between actually installed and the available software versions. Storage usage accounting will be supported by the sites. The A&R monitoring system (data project level monitoring) will be utilized to improve the quality of the service delivery. The automation of the deployment of the B2 services will be extended to further platforms and service types. A directory for frequently asked questions and answers as well as known issues and solutions in the domain of service management and resource provisioning will be introduced and maintained.

### **8.3. Project enabling and Service Integration**

New and already accepted data projects will be further coordinated, planned, enabled and supervised during the pre-production phase. The planning of the enabling of the data pilots will be further supported and based on these plans the enabling will be done. Implicitly or explicitly made service development and feature requests will be extracted, generalized and forwarded to the service developers. The results of the project enabling will be documented using the DPMT.

The different data project cases will be further analyzed and classified in order to consolidate the inventory of solution patterns. Generic and B2 service-specific recipes will be developed for service integrators and data project enablers. This will help to consolidate the knowledge and solution sharing platform for enablers and integrators. Second level support will be provided by project enablers and service integrators during the planning, enabling and pre-production phases of the data projects, and contributions will be made to the corresponding FAQs.

### **8.4. Helpdesk and Support**

The Helpdesk system and the first level support with response times of less than 12h will be continued, and the responsiveness of the 2<sup>nd</sup> and 3<sup>rd</sup> level support will be supervised as far as the Helpdesk system permits. New services (e.g. B2GETHER, B2NOTE and others) will be included into the helpdesk system as well as new site queues if new service providers join the CDI. The support team will provide monthly statistics that include the frequency of tickets in the different queues, response and solution times. The EUDAT support guideline will be updated and the FAQ with typical questions and frequent answers from the first and second level support will be further assembled.

### **8.5. Security**

EUDAT security is planning for the following 12 months to focus on following activities: The EUDAT basic security related documentation will be updated and consolidated, this includes the harmonization of the Terms of Use across the B2 services and service providers (in collaboration with WP2), an updated security policy and guideline and the security risk assessment. The e-Infrastructure security-related aspects will be investigated, for instance the setup of an EUDAT CA with an appropriate certification policy and the management of bundles of certificates of trusted CAs. The collaboration with the other European e-Infrastructures will be continued. The risks concerning the integration of services from external cloud providers (e.g. Amazon WS) will be discussed and an operational data privacy guideline (EUDAT privacy policy) will be drafted in collaboration with WP2. The regular vulnerability scans will be continued and further developed, and service security assessments will be performed as far as necessary. The EUDAT CSIRT function and the vulnerability management will be continued.

## 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.
B2ACCESS	Brand of the EUDAT service for federated authentication and authorisation
B2DROP	Brand of the EUDAT trusted cloud storage service
B2FIND	Brand of the EUDAT central metadata catalogue
B2HANDLE	Brand of the EUDAT persistent identifier service
B2HOST	Brand of the EUDAT service to deploy community applications close to the data storage location
B2NOTE	Brand of the EUDAT service to manage semantic annotations
B2SAFE	Brand of the EUDAT service via which the data management policies are implemented within the CDI network
B2SAFE DPM	Brand of the EUDAT B2SAFE data policy manager via which community data manager are able to manage policies which the CDI network from a central portal
B2SHARE	Brand of the EUDAT easy-to-use data repository service
B2STAGE	Brand of the EUDAT comprehensive set of API's and tools to access data managed within the CDI network
B2 service suite	Aggregation name of the EUDAT B2 services
CERT	Computer Emergency Response Team

CIS	CDI Configuration Information Service: central or distributed service that provides information about the configuration of the CDI. Entitled clients can look-up information, in particular, about service instances, service component instances.
CI type	Configuration Item type, denotes a class of configurable entities
CDI node	Generic of Thematic service provider who has signed the CDI collaboration agreement either as interoperable or integrated partner
CDI	EUDAT Collaborative Data Infrastructure
CDI Gateway	Service or service endpoint (e.g. API or WUI) which is part of the Access Layer of the CDI layered architecture.
CLARIN	Common Language Resources and technology Initiative. An ESFRI project in the Social Sciences and Humanities domain.
CMON	Central nagios-based monitoring of all the service endpoints within the CDI
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.
CREG	EUDATs Central Site and Service Registry, <a href="https://creg.eudat.eu">https://creg.eudat.eu</a>
CRM	Customer Relationship Manager
CROWD	Centralized Identity Management solution from Atlassian provided by CSC
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	The <i>Data Project</i> is a planned set of related tasks in the domain of data management and stewardship to be executed over a defined period of time and under cost and other constraints. A <i>Data Project</i> can be broken down into manageable units (subprojects). Specifically, it is regarded as the set of tasks for planning, configuring, providing, managing, monitoring and accounting CDI services and executing workflows for customers. The time span must be defined but can be prolonged subsequently, so that a data project can last as



	long as it is sponsored or funded by a customer. A data project can be taken over from one customer to another.
DoW	Description of work – specifically the EUDAT DoW
DPCP	Data Project Coordination Portal is the same as the Data Project Management Tool (DPMT)
DPMT	Data Project Management Tool
DSA	Data Seal of Approval – <a href="http://www.datasealofapproval.org">http://www.datasealofapproval.org</a>
DSI	GridFTP Data Storage Interface
EB	EUDAT Executive Board
EGI	European Grid Initiative
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.
GÉANT	European computer network for research and education purposes.
General Service Provider	The General Service Provider is a Service Provider within the CDI who is responsible to ensure that SLA with the Customer is fulfilled. Prime Contractor for delivering a specific (distributed) service instance in context of a Data Project. The General Service Provider is acting as Prime Contractor for delivering a distributed service instance in context of a Data Project. A General Service Provider can also be responsible for the quality of suite of services that are provided to a customer in the context of a Data Project (the project Principal).
Generic Service Provider	Generic Service Providers have regional, organisational or national mandates to support scientific research, usually from different disciplines.
GOCDDB	Technology to maintain general information about sites and service endpoints, is currently used as site and service registry within the CDI network
HPC	High Performance Computing
IaaS	Infrastructure as a Service
IdP	Identity Provider
iRODS	Integrated Rule-Oriented Data System <a href="http://www.irods.org">http://www.irods.org</a>
ISO	International Organisation for Standardization

Local SP Coordinator	Service Provisioning Coordinator (or Site Deputy): person in charge of coordinating the provisioning of requested services, service components and resources in time when requested.
MoU	Memorandum of Understanding
NREN	National Research and Education Network
OAIS	Open Archival Information System
OASIS	Organisation for the Advancement of Structured Information Standards
OPCT	Operations Coordination Team; site deputies of the service providers supported by their staff that are sharing information and coordinate actions within the CDI.
OSCT	Operational Security Coordination Team
PaaS	Platform as a Service
PID	Persistent Identifier
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.
RCT	Resource Coordination Tool for managing information necessary for coordinating the provisioning of resources.
Resource Provisioning	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. Important is the fact that a Resource Provider is responsible to provide the resources with a contractually asserted quality, reliability and availability.
RI	Research Infrastructure
RP, Resource Provider	An organisational instance that provides maintains and operates compute, storage and network resources at a site.
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 Portfolio	Internal list that details all the services offered by the service provider (those in preparation, live and discontinued). Note: The service portfolio includes meta-information about services such as their value proposition, target customer base, cost and price, risk to the provider, service level agreements offered and operational level agreement supporting them.
Service Provisioning	Make services available which are deployed on top of resources. Services can be deployed and operated on top of resources provided at another site.
Service Provider	Organisation or federation or part of an organisation or federation that manages and delivers a service or services to customers
Site	A legal entity that is able to sign contracts such as Service or Operational Level Agreements. A site can provide different resources and services. A site can be a Resource Provider, a Generic or a Thematic Service Provider. At the same time it can act as General Service provider (being contractually responsible for the provisioning of a distributed service) and a Data or Content provider
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
SPMT	Service Portfolio Management Tool, a tool that facilitates the Service Portfolio Management process.
SVN	Subversion – a version control system <a href="http://subversion.apache.org">http://subversion.apache.org</a>
Thematic Service Provider	Thematic Service Providers are discipline-specific organisations mandated to support a defined scientific community or group of customers and users.
ToU	Terms of Use
TTS	Trouble Ticket System

## Annex B. CDI SERVICE MANAGEMENT PROCESSES AND ROLES

The following summarizes the service management processes and roles that are mentioned in section 2.5.

The overall EUDAT CDI must fill the following roles

Role	Responsible for	Accountable to
<b>Secretariat</b>		
CDI Head of Secretariat	Head of the EUDAT CDI Secretariat. Service management system Senior Responsible Owner. Manages the <i>Partnership of CDI Service Providers</i> . Acts as primary interface for <i>organising involvement of pan-European communities, potential customers and/or new service providers</i> .	CDI Council
CDI Technical Architect	CDI service design, development and release	CDI Head of Secretariat
CDI Operations Coordinator	CDI service enabling, operations & support	CDI Head of Secretariat
CDI Outreach Coordinator	Organise communication towards and involvement of users, communities and service providers beyond the CDI. Promote & market the CDI and its services.	CDI Head of Secretariat
<b>Service Owners</b>		
B2STAGE	Service design & development. Operational Service instances.	CDI Steering Group
B2SAFE	Service design & development. Operational Service instances.	CDI Steering Group
B2DROP	Service design & development. Operational Service instances.	CDI Steering Group
B2FIND	Service design & development. Operational Service instances.	CDI Steering Group
B2HANDLE	Service design & development. Operational Service instances.	CDI Steering Group
B2ACCESS	Service design & development. Operational Service instances.	CDI Steering Group
CDI Gateway API	Service design & development. Operational Service instances.	CDI Steering Group
Metadata Store	Service design & development. Operational Service instances.	CDI Steering Group
Data Type Registry	Service design & development. Operational Service instances.	CDI Steering Group
Service Hosting	Service design & development. Operational Service instances.	CDI Steering Group
Consultancy	Project enabling, Service enabling, Professional Services.	CDI Operations Coordinator
Training	Service-specific & general training.	CDI Operations Coordinator
<b>Process Managers</b>		

Service Portfolio Manager	Maintain the EUDAT CDI Service Portfolio.	CDI Steering Group
Service Level Manager	Define and maintain the CDI Service Catalogue and agreed SLAs. Define and maintain service report templates and reporting mechanisms.	CDI Operations Coordinator
Availability & Capacity Manager	Manage provisioning and configuration of all the resources and services. Ensure that the use of the Service Components and Resources are accounted on the projects. Manage Customer requests for the delivery of services.	CDI Operations Coordinator
Security Manager	Manage information security risks & policies for the CDI.	CDI Operations Coordinator
Customer Relationship Manager	Manage the relationship between Service Providers and Customers (both Users and other Service Providers).	CDI Steering Group
Supplier Relationship Manager	Manage or coordinate the relationship between CDI Service Providers and key software and hardware providers.	CDI Steering Group
Incident & Service Request Manager	Manage service desk and 1 <sup>st</sup> line support.	CDI Operations Coordinator
Problem Manager	Manage in-depth (2 <sup>nd</sup> line) support and problem resolution.	CDI Operations Coordinator
Configuration Manager	Provide and maintain a logical model of all configuration items and their relationships and dependencies.	CDI Operations Coordinator
Change Manager	Ensure changes to configuration items are planned, approved, implemented and reviewed in a controlled manner.	CDI Operations Coordinator
Release Manager	Manage process of bundling changes of one or more configuration items into releases.	CDI Technical Architect
CSI Manager	Identify opportunities for improvement, evaluate and approve proposals for improving services.	CDI Steering Group

Each Service Provider in the EUDAT CDI must fill some or all of the following roles, depending on their level of integration (interoperable or integrated).

Role	Responsible for	Accountable to
CDI Site Manager	Local service availability & reliability monitoring and reporting	Service Level Manager
	Local resource & data usage accounting and reporting	Service Level Manager
	Registration and management of local service instances	Availability & Continuity Manager
	Registration and capacity management of local resources	Capacity Manager
	Local information security management	Security Manager
	Local incident and service request management	Incident and Service Request Manager
	Local problem management	Problem Manager

## Annex C. DATA PROJECT MANAGEMENT TOOL - USE CASES AND WORKFLOWS

### C.1. Use Cases and Workflow Requirements

As tool for the Data Project and Service Management the DPMT has been designed to support the following use-cases and workflows (Table 11). These use cases comprise specific and general requirements which resulted in the data model which is presented in section 3.1.3.

Table 11: DPMT Use Cases and Workflows

#	Use case (and requirements)	Workflow / Activity	Roles involved
1	Information about (Service) <b>Provider</b> <sup>79</sup>	Registration Request, Registration, Maintenance of Information, characterization of the provider <sup>80</sup> , suspend collaboration, leaving the collaboration; CDI manager accepts the registration and supervises modifications; the <i>CDI manager</i> entitles a <i>Provider</i> , the <i>Provider</i> entitles <i>Site Admins</i> ; <i>Site Admins</i> entitles and <i>Service</i> , <i>Service Component</i> and <i>Resource Managers</i> . Per default the <i>Provider</i> has all these roles.	CDI Manager, Provider (role delegation), Site Admins
2	Information about of <b>Service Offers</b> <sup>81</sup> and the conditions under which they can be realized (e.g. B2SAFE for CLARIN repositories)	<i>Registration of a selection of Services</i> from the <i>Services Catalogue</i> as <i>Service Offers</i> , specifying the conditions for providing the <i>Service</i> ; modifying or retracting registered service offers. A <i>Provider</i> who is offering a distribute <i>Service</i> are potential <i>General Service Providers</i> .	Provider <sup>82</sup> , CDI Manager (Supervisor)
3	Information about <b>Service Component Offers</b> (as parts of distributed services, e.g. CKAN, DSPACE, iRODS v4, OAI-PMH). <i>Service Component Offers</i> indicate the capability of the providers.	Select a <i>Service</i> from the <i>Services Catalogue</i> and register a selection of necessary service components as <i>Service Component Offers</i> ; specifying the constraints for providing the offered service components; modifying or retracting registered <i>Service Component Offers</i> ;	Provider, CDI Manager (Supervisor)
4	Information about <b>Resource Offers</b> (storage, network, service hosting, computing). <i>Resource Offers</i> indicate the pledged capacity per resource class.	<i>Resource Provider</i> specifies and registers his <i>Resource Offers</i> ; specifies constraints for providing the resources; modifies or retracts registered <i>Resource Offers</i> .	Provider, CDI Manager (Supervisor)
5	Information about the <b>Customer</b> as the principal of one or more Data Projects for a community.	Registration of the Customer. Customers of Data Projects can register themselves or are registered by the CRM; Customers of active Data Projects can be deactivated but not deleted. A customer must be a legal entity	CRM, Customer (Data Project Principal), CDI Manager (supervisor), Customer Admins

<sup>79</sup> The characterization of a provider as general, resource, service, and thematic or generic service provider does not matter at registration request time; basically also external (non-CDI) providers can register, if they provide and maintain sufficient information.

<sup>80</sup> Based on the Role that a provider has as associated partner or member of the CDI, the provider can be characterized as Resource Provider, Service (Component) Provider, General Service Provider, and at the same time the provider can be associated (non-CDI) partner or CDI member (EUDAT Partner). In addition, providers can be classified as Generic Service Provider or Thematic Service provider; latter often provide domain repository services (data provider) and/or are acting as front-office to specific user communities.

<sup>81</sup> Indicate the ability to offer specific Service Instances under certain constraints

<sup>82</sup> By offering a (distributed) full-service the provider indicated interest to become a general provider and be responsible for at least one instance of a distributed service.



#	Use case (and requirements)	Workflow / Activity	Roles involved
		(a person representing a legal entity). The Information about customers of active projects must be kept up-to-date. A Customer can delegate the Customer Admin as a role to represent the Customer.	
6	Information about the <b>Data Project Requests</b> (e.g. a data management plan submitted via a <i>Data Management Planning Tool</i> )	Register <i>Data Project Requests</i> by providing an initial Data Project Description at least in narrative form that characterizes the service categories, describes the data flow and indicate preferential CDI service providers.  The registration binds a (globally) unique persistent identifier to the data project.	DMP tool <sup>83</sup> , Customer / Customer Admins, CDI Manager, CRM
7	Manage the Information about the <b>Data Project</b> (project structure, project description, work plan and status) during the <b>Project Life Cycle</b> . The <b>Project Structure</b> (creating phase-specific or permanent subprojects) and the work plan for the (sub)project can change from stage to stage and also the description of the overall project as well as of the subprojects is advancing as the project is progressing from planning, implementation / enabling, depending also on the timeline of the enabling of subprojects. If the project is decided to be structured into several subproject (e.g. one per service instance to be enabled), there must be a project manager responsible for the overall project.  As far as required for accounting, each subproject should obtain an own (globally) unique persistent identifier.	The implementation of a project starts after the project description has been approved by a review process and project enabler / manager has been appointed. This (overall) project enabler decides in agreement with the customer, who has to enabling from the customer side, about the project structure and if subprojects are defined, he ensures that each subproject has an enabler assigned when the subproject leaves the planning phase and get enables..  The project description / information for each project phase (see section 3.1.4). Main actor in this phase is the project manager. The Customer shall be able to track the project implementation progress. The customer has to make sure that the necessary service integration is done at his site.  The principal Project manager can entitle other Project Managers called <i>Project Enablers</i> . These enablers can be assigned to subprojects. A Project Manager is per default a Project Enabler.	Reviewer,  Project Manager, Project Enabler,  CRM, Customer, Customer Admins (PI:observer, technical contact: information co-manager),  CDI Manager (observer)
8	Registration of a <b>Service Instance</b> and management of information about the service instance.  The registration provides a (globally) unique persistent identifier for the <i>Service Instance</i> .	Service Instances (e.g. B2SHARE for project A, B2SAFE for project B) are initially registered as a request that needs to be confirmed by a General Service Provider that offers to be responsible for providing the requested distributed service instance for a specific Project. The General Provider acts as (SLA) contractor concerning the service instance or instantiated service suite. Once the General Service Provider has been identified (by checking the Service Offers and conditions, defined by the	Project Manager, Project Enabler (makes request),  Provider, Provider Admins (confirm and maintain information),  CDI manager (supervisor)

<sup>83</sup> E.g. the Data Management Planning Tool from DCC: <https://dmponline.dcc.ac.uk/>

#	Use case (and requirements)	Workflow / Activity	Roles involved
		<i>Project Enabler</i> ) a service (instance) manager needs to be identified who take care about the correctness of the information about the service instance. This service instance owner can view the information about the related service components. The role of a Service Instance Manager can be taken by a Provider Admin.	
9	<p>Registration of a <b>Service Component Instance</b> and maintenance of the information about this <i>Service Component Instance</i>.</p> <p>The registration provides a (globally) unique persistent identifier for the <i>Service Component Instance</i>.</p> <p>The identifier can be used for accounting the provisioning and usage of the registered service component.</p>	<p>A <i>Service Component Instance</i> (e.g. iRODS v.4 instance for B2SAFE for projects A,B, C; or e.g. Invenio v3 for B2SHARE for projects X,Y) is registered on basis of a request that needs to be confirmed by the <i>Service Provider</i> that is involved. The <i>Service Provider</i> is responsible for providing the service instance; the provider is an OLA contractor concerning the provided service component instances. Once the <i>Service (Component) Provider</i> has been identified (by checking the directory of <i>Service Component Offers</i> and delivery constraints) a service component (instance) manager is appointed to take care for the completeness and correctness of the information about the <i>Service Component Instance</i>. This <i>Service Component Instance Manager</i> has to make sure that the proper Resources with the necessary capacity are provided, and he supervises that the information about the Resources is adequate (the resource manager is responsible to keep the information up-to-date). The role of a Service Component Instance Manager can be taken by a Provider Admin.</p>	<p>Project Manager, Project Enabler (makes request),</p> <p>[General Service] Provider (supervision in case of a distributed service),</p> <p>[Service Component] Provider, Provider Admins (confirms and maintains information),</p> <p>CDI manager (supervisor)</p>
10	<p>Registration and Management of information about a <b>Resource</b> which is linked to a specific (sub)project and provider. Resources are grouped into resource classes (e.g. online storage with one copy, nearline storage with 2 copies) and can be further characterized including the specification of the (minimum and) maximum capacity that is allocated in the context of the corresponding data project.</p> <p>The registration provides a (globally) unique persistent identifier for the <i>Resource</i>.</p>	<p>A <i>Resource</i> is registered on basis of a request that needs to be confirmed by the <i>Resource Provider</i> that is involved (a <i>Provider</i> is often <i>Service and Resource Provider</i> at the same time). The <i>Resource Provider</i> is responsible for providing the resources as specified by the resource request; in cases of a mismatch of the profile of the requests and the profile of the available resources (regarding capacity, class, ...) requester and resource provider shall be able to negotiate; the resource provider is an OLA contractor concerning the provided registered resource. Once the <i>Resource Provider</i> has been identified (by checking and selecting a provider from the directory of <i>Resource Offers</i>) a local <i>Resource (Instance) Manager</i> must take care about the correctness and completeness of the information about the</p>	<p>Project Manager, Project Enabler (requests resource capacity, checks that resource is accounted),</p> <p>[Service Component] Provider, Provider Admins (can also make resource request)</p> <p>[Resource] Provider, Provider Admins (registers the requested resource)</p> <p>CDI manager (supervisor)</p>

#	Use case (and requirements)	Workflow / Activity	Roles involved
		<i>Resource Instance</i> . The <i>Resource Manager</i> has to make sure that the committed capacity is continuously available until the capacity limit is reached. He informs in advance about scheduled downtimes. He makes sure that there is no disruption and that the availability and quality of the provided resource complies with the OLA. He makes sure that disasters can be recovered with a remaining risk that is accepted by the OLA. The Site Deputy is normally acting as a proxy to the local Resource Manager. The Resource Manager makes sure that the usage of the <i>registered Resource</i> is regularly reported to the accounting system.	
11	<p><b>Storage Capacity</b> management of <i>Storage Resources</i> that belong to specific projects. Storage Capacities per project should be adaptable to a certain extend.</p> <p>The unique <i>Resource Identifier</i> is used for accounting the usage of the registered resource.</p>	Customer (can <i>view</i> the current storage space that currently allocated under a project. He can <i>request</i> for additional storage capacity) as well as the Project enabler (requests resource capacity) while the Site deputy can register or modify the information about additional storage capacity.	<p>Customer, Customer Admin (<i>view, request</i> storage capacity),</p> <p>Project Enabler (request),</p> <p>[Service Component] Provider, Provider Admins (request)</p> <p>[Resource] Provider, Provider Admins (register the capacity change)</p>
12	Entitled clients (applications, services) have <b>API</b> access to information of the DPMT.	<p>The DPMT information can be fetched using an API. Depending on the scope of information (public, non-public) clients have to authenticate and, in case of access restrictions, they have to be authorized.</p> <p>In order to support the interoperability with tools used by EGI (e.g. the A&amp;R monitoring service), the GOCD v5 API should be supported as far as necessary.</p>	Entitled clients; requirements for authentication and authorization depend on the scope of the information.
13	The DPMT integrates with <b>B2ACCESS</b> .	All actors can authenticate against the DPMT using their B2ACCESS credential. At the first time, the credential has to be registered at the DPMT. Roles and access rights are assigned depending on the LoA. Fundamental roles (e.g. CDI Manager) are managed by the DPMT site manager. Some roles should be managed by other roles.	All Actors, all Roles
14	<b>Delegation of Roles</b> and Access management	Role and Access management is delegated by some fundamental roles: CDI Manager can assign <i>Provider, Customer, and Customer Relationship Manager</i> . A <i>Provider</i> can appoint <i>Provider Admins</i> (acting in behalf of	All Actors, all Roles

#	Use case (and requirements)	Workflow / Activity	Roles involved
		the provider). A <i>Customer</i> can appoint <i>Customer Admins</i> .	
15	<b>Fetch</b> regularly the descriptions <b>from the Service Portfolio</b> (including the Service Catalogue with details about supported software packages and service components). <b>Display</b> the <b>Service Portfolio</b> .	Interface to the SPMT and import detailed descriptions from there; regular sync of the DPMT information objects and templates with the detailed descriptions of <i>all services</i> on at least hourly bases; create new DPMT information object versions only in case of changes of the SPMT descriptions; use the latest version but indicate recent changes also on the overview panel.	Project Enabler, Providers, Provider Admins, Service Portfolio Manager, CDI Manager, CRM, Service Developers (view)
16	<b>View</b> the information from the <b>Service Catalogue</b> (including supported package and component versions).	Use and filter information recently imported from the SPMT.	Project Enabler, Providers, Provider Admins, Service Portfolio Manager, CDI Manager, CRM, Service Developers (view)
17	<b>View and compare</b> the <b>version numbers</b> of actual installed service instances and service component instances and supported service Instances; <b>notify about new software versions</b> ;	Import information from the (service and service component) software version information system; notify the service providers that are running service components about changes component versions.	Service Provider, Service Portfolio Manager (supervision), CDI Manager (supervision)
18	<b>Incorporate service component implementation details</b> (attribute names) into the service component information classes; notify about mismatches if attributes have been added of existing attribute names are modified;	Use and analyse the information recently imported from the SPMT. Check for potential mismatches between the existing schemas of already registered service components and the new service component implementation details. Apply the schema as defined by the recently important service component implementation details and notify in case of mismatches or modifications of attribute names.	Service Provider, Service CDI manager (supervision)

## C.2. Data Project Stages and Transitions

The Data Project life cycle begins with a project plan (data project request) and ends with the termination of the project if it has not been retracted or replaced by another project before. Data projects can comprise multiple services and service instances.<sup>84</sup> This may require a breakdown of a data project into several subprojects – ideally one per service class. Each subproject can be enabled by a different consultant for a dedicated life cycle, work plan and time line. The following project phases may apply also to subprojects.

<sup>84</sup> E.g. a customer orders multiple B2DROP instances to be provided for his community: one B2SHARE instance and several repositories are to be connected with B2SAFE. In addition the metadata from each of the community repositories shall be harvested via OAI-PMH connectors. These different services can most efficiently be enabled by concurrent processes managed by different consultants.

Table 12: Project Phases

Phase	Description	Phase transition
<b>Submitted</b>	The Data Project is preregistered by the DPMT as a project request. It can be a data management plan that refers to the kinds of services that EUDAT provides (see service classes listed in section 3.1). This request can be described in textual form. The Data Project request needs to be reviewed and accepted before any enabling efforts and resources can be spent and accounted on this project.	previous: pre-sales, WP4 next: reviewing
<b>Reviewing</b>	The Data Project request is being reviewed by an assigned project reviewer or a review team. During the review phase a preliminary project implementation plan becomes available that maps the high-level service request tentatively to a suite of services (from the SPMT) that EUDAT is able to provide and enable in the requested timeframe. While having been preliminarily planned the project description still needs to be accepted before any enabling efforts and resources can be accounted on this project. The project request can be retracted by the customer at that time.  The review requires a reviewer or board of reviewer that is able to recommend and negotiate improvements and precisions on the Data Project Request.	previous: submitted next: accepted   retracted
<b>Accepted</b>	The refined <i>Data Project</i> request, the preliminary project implementation plan, is either accepted or recommended to be retracted. The accepted data project is registered by assigning a persistent identifier which should be used for accounting any dedicated resources used for implementing and running the project. The data project can be accepted under the premise that a kind of service is requested that needs to be newly or further developed; in this case the service development team (WP5) need to be informed about the service requirements. The accepted and registered project can and should be suspended for a limited period of time if the human, financial or system resources are not available at the time when the project is accepted.  Data Projects need to be accepted by a committee that can make decisions in behalf of the EUDAT CDI consortium. This holds also for data pilots which have a strong development component and which are going to be implemented via the same process as Data Projects.	previous: reviewing next: planning   retracted developing, suspended
<b>Developing</b>	If service development is required the appropriate service development team (WP5) is informed about the service requirements. This is communicated via the service development request channel provided by WP5. This also holds if new features are required from existing services. New services required should be negotiated and requested via CRM. New features of existing services should be requested by the Project Enabler.  A project can be suspended or retracted (latter by the customer) when it is in the phase <i>Developing</i> .	previous: accepted, next: planning, suspended, retracted

	<p>The project remains in status <i>Accepted</i> until resources are assigned (e.g. a project enabler) that allows the project to be planned.</p> <p>Currently there are development requests regarding: dev:BigD=Big Data Analytics, dev:DataT=Data Type Registry, dev:Sem= Semantic Annotation, dev:Viz=Visualisation (see Table 4)</p>	
<b>Planning</b>	Data Project/Service enabling is still under discussion; customer interest has been re-confirmed if interaction is set on standby, planning with customer continuing	previous: accepted next: enabling by community, enabling, suspended, developing, retracted
<b>Enabling (community)</b>	service deployment/integration/enabling at customer (community) side (repository) only, CDI providers are selected and B2-service enabling can be done in parallel, but storage service is not needed to be provided as long as the integration at the customer side is going on.	previous: planning next: enabling , suspended, retracted
<b>Enabling</b>	Project, subproject or service enabling at community and EUDAT side, the selected service provider are contacted, enabling work concerns the community (repository) and EUDAT nodes	previous: planning next: pre-production, (developing features), suspended, retracted
<b>Pre-production</b>	The service or service suite is operational and passed predefined tests. Limited extend of the distributed service (not all nodes are involved). Limitation regarding disaster recovery. Service (suite) has not yet proven long-term availability or performance need to be improved or availability is not ensured. Accounting and/or monitoring not necessarily enabled.	previous: planning, enabling next: production, suspended, terminated
<b>Production</b>	The service / service suite is fully enabled (installed, configured, tested) and has been used by the customer community since 1 one month in compliance with the SLA. Accounting and monitoring are enabled.	previous: pre-production next: suspended, terminated, replaced
<b>Suspended</b>	A project can be set on hold for a limited period of time. It shall be continued after this period of time or a decision is to be made how handle the project further.	previous: accepted, planning, enabling (communities), enabling, next: production, suspended, terminated
<b>Terminated</b>	A project on status <i>Production</i> or Pre-Production can be terminated. The implications and actions in the context of project termination has to be defined beforehand.	previous: production, pre-production next: none
<b>Replaced</b>	A project can be replaced by a successor project.	previous: pre-production, production next: none
<b>Retracted</b>	A Data Project can be retracted by the customer on several stages of the project.	previous: accepted, planning, enabling next: none



## Annex D. A & R MONITORING AND THE TOPOLOGY OF B2 SERVICES

This annex refers to section 3.3.2. The text is the version of a document from the 1<sup>st</sup> project year that considers the topologies of the distributed services and service components<sup>85</sup> and how these topologies can be described on the basis of the GOCDB v5.4 data model. Such a dedicated instance of the GOCDB is currently used as Site and Service Registry for EUDAT.

### 1. Service Registry Overview

The GOCDB offers a flexible way to inform about distributed services provided by an infrastructure as collections of service endpoints. The data model of the GOCDB has the notion of *Service Types*, *Services* (*Service Endpoints*), *Service Groups* and *Sites*<sup>86</sup>.

According to the GOCDB terminology, a *Service* is a single entity which is formed by a *hostname*, one *hosted service* (*Service Endpoint* with a *Service Type*) and a *URL*. A *compute resource* can host many services and there can be many *service endpoints* per *compute resource*.

A *Service Type* represents the functional properties of a service, particularly of its endpoint (port, socket). Each *Service Endpoint* in GOCDB is associated with a *Service Type*. *Service Types* are classes of services while a *Service Endpoint* is a specific instance of a service that is running in a certain context.

A *Service Group* is a collection of existing *Service Endpoints* that can be distributed across different sites. Each service that appears in a *Service Group* must already exist and be hosted by a physical *Site*. *Service Groups* are typically used for monitoring a particular collection of *Service Endpoints* which may form a distributed service.

A *Site* is a grouping of resources collating multiple *Service Endpoints*.

### 2. EUDAT Services

EUDAT has the following product offerings: B2DROP, B2FIND, B2SAFE, B2SHARE and B2STAGE (and B2ACCESS & B2HANDLE as support services). Table 13 shows that each of these products exposes one or more *Service Types* to the EUDAT customers.

#### 2.1 B2DROP

The B2DROP service offering is based on the OwnCloud technology and there is one *Service Type* associated with this product. A community using the EUDAT B2DROP service will be using one OwnCloud instance hosted at a specific site. B2DROP depends on B2ACCESS when the services are integrated.

#### 2.2 B2FIND

The B2FIND offering is based on CKAN and there is one “Service Type” associated with this product. A community using the EUDAT B2FIND service will be using one CKAN instance hosted at a specific site. For metadata harvesting and in order to keep the CKAN repository up-to-date, the B2FIND service depends on OAI-PMH service endpoints which are provided by the different repositories which are harvested.

#### 2.3 B2SHARE

B2SHARE has two *Service Types* associated. The service uses Invenio for the digital repository and the B2HANDLE service for assigning and managing the PIDs to the digital objects that are uploaded and registered by Invenio. In addition, B2SHARE depends on B2ACCESS. Therefore, a community that is using the EUDAT B2SHARE service will be using

- 1 Invenio instance hosted at a specific site

<sup>85</sup> The model of the distributed Services as Compositions of Service Components is also shown by Figure 5, p.26.

<sup>86</sup> For compatibility reasons, the DPMT uses the same properties, also extension properties, for *Site*, *ServiceGroup* and *ServiceEndpoint*. GOCDB:*Site* maps to DPMT:*Provider*; GOCDB:*ServiceGroup* maps approximately to DPMT:*Service* and GOCDB:*ServiceEndpoint* maps to DPMT:*ServiceComponent*.

- 1 B2HANDLE instance hosted at a specific site

NOTE: B2HANDLE is a distributed service that comprises a handle resolver, a handle API service and one or more handle mirrors. The b2handle.handle.api type corresponds to the B2HANDLE instance mentioned above (EPIC API service). Since every service depends on B2ACCESS, the AAI service will not be considered separately.

## 2.4 B2STAGE

B2STAGE is a product that has currently two *Service Types* associated. B2STAGE uses b2safe.dsi (the data staging interface which is GridFTP in front of an iRODS service) and the b2stage.gridftp service types. A community using the B2STAGE service will be using:

- Multiple DSI interfaces hosted at a different sites and connected to B2SAFE or
- Multiple GridFTP service hosted at a different sites

## 2.5 B2SAFE

B2SAFE is more complex product that has two “Service Types” associated to it. B2SAFE uses multiple iRODS services and B2HANDLE service instances for assigning and managing the PIDs of the digital objects that are stored and managed by B2SAFE. A community that is using the EUDAT B2SAFE service will be using:

- Multiple iRODS instances hosted at different sites
- Multiple B2HANDLE instances hosted at different sites

**Table 13: EUDAT Products for monitoring and Service Type**

EUDAT Product	Service Type	Comment
B2DROP	b2drop.owncloud	The url registered in the GOCDB corresponds to the http interface of the service
B2FIND	b2find.ckan	The url registered in the GOCDB corresponds to the http interface of the service
B2SHARE	B2SHARE.invenio	The url registered in the GOCDB corresponds to the http interface of the service
	b2handle.handle.api	The url registered in the GOCDB corresponds to the http interface of the EPIC API
B2STAGE	b2safe.dsi	The url registered in the GOCDB corresponds to the data staging interface (i.e GridFTP on top of iRODS)
	b2stage.gridftp	The url registered in the GOCDB corresponds to the gridftp interface of the service
B2SAFE	b2handle.handle.api	The url registered in the GOCDB corresponds to the http interface of the EPIC API
	b2safe.irods	The url registered in the GOCDB corresponds to the gridftp interface of the service
B2ACCESS (Internal)	b2access.unity	The url registered in the GOCDB corresponds to the http interface of the Unity service
B2HANDLE (Internal)	b2handle.handle.api	The url registered in the GOCDB corresponds to the http interface of the EPIC API
	b2handle.handle.master	There are no entries in the GOCDB
	b2handle.handle.mirror	There are no entries in the GOCDB

## 2.6 B2ACCESS (support service)

The B2ACCESS service is based on Unity and there is currently one *Service Type* associated with this product. B2ACCESS is used by all communities and there is no specific service instantiation for each scientific community. B2ACCESS is provided as a central service in the CDI.

## 2.7 B2HANDLE (support service)

The B2HANDLE service is based on Unity and there is one “Service Type” associated with this product. B2HANDLE is used by all the scientific communities that are using either B2SHARE or B2SAFE.

## 3. Customers (Scientific Communities), EUDAT Services and the GOCDB

The EUDAT services are hosted by EUDAT infrastructure partners. When a EUDAT service is provided to a customer, namely a scientific community, it is not necessarily provided by all the infrastructure partners who are hosting the corresponding service endpoints, but a subset of them. An infrastructure partner for a specific EUDAT Service can provide one or more instances of that service according the nature of the service and the scientific communities that it supports.

GOCDB is representing this as follows:

- Each Service Provider (CDI node) is registered as a *Site* in the GOCDB.
- Each Service Provider is responsible for registering the *Service Endpoints* that are hosted on-site. The registration is a manual process.
- *Service Types* in the GOCDB do not correspond to the EUDAT Services which are generally understood as distributed services, but they correspond to the technical components of an EUDAT Service. The naming of the *Service Types* follows the scheme *eudat\_service.technical\_service* e.g. *b2safe.irods*
- Each EUDAT Service is described as a GOCDB *Service Group* which has all the *Service Endpoints* for the *Service Types* which are included. Adding *Service Endpoints* to a GOCDB *Service Group* is a manual process.

For example there is a B2SAFE GOCDB Service Group, which has all the “Service Endpoints” with “Service Types” *b2handle.handle.api* and *b2safe.irods*

- Each EUDAT Service that is provided on the basis of a customer order (data project) is hosted by one or more service providers (*sites*). This is modelled in the GOCDB as a GOCDB *Service Group*, which has all the *Service Endpoints* included which *Service Groups* are provided for a specific data project. The registration of the *Service Endpoints* in the GOCDB is a manual process. The name of these GOCDB *Service Groups* follows currently the scheme “EUDAT SERVICE”\_“SCIENTIFIC COMMUNITY” e.g. *B2SAFE\_CLARIN*

## 4. Clustered services and roles

### 4.1 Clustered set services with the same role

With the term *Cluster Services*, we mean a set of *Service Endpoints* of the same *Service Type* that are part of a logical cluster in order to provide a service with higher quality characteristics than it would have if the service was delivered by one service endpoint. Usually these Service Endpoints are hosted at different *Sites*, but it can be the case that some or even all of them are hosted at the same *Site*. This set up is very common and can be easily modelled in the GOCDB.

### 4.2 Clustered services with roles, independent of the grouping

Things become more difficult when in such a cluster of *Service Endpoints* of the same *Service Type*, one or more of the *Service Endpoints* have different roles. For example, the Handle service: A *Handle Service Endpoint* can be either a *Master* or a *Replica* instance. This can be modelled in the GOCDB in multiple ways:

1. It is possible to have two different *Service Types*, one for the *Master* (e.g. *handle.master*) and one for the *Replicas* (e.g. *handle.mirror*). This is possible for the Handle service as the *Master* and *Replica*

roles can apply to the *Service Endpoint* which can have only one of the two regardless of the GOCDB Service Groups that the Service Endpoint is registered in.

2. There could be one single *Service Type* (e.g. *handle.server*) and in the extensions of each *Service Endpoint*, we could add a role key.

Option 2 is demonstrated in the configuration example:

```
<SERVICE_ENDPOINT PRIMARY_KEY="5G0">
<HOSTNAME>epic.rzg.mpg.de</HOSTNAME>
<GOCDB_PORTAL_URL>
https://creg.eudat.eu/view_portal/index.php?Page_Type=Service&id=5
</GOCDB_PORTAL_URL>
<SERVICE_TYPE>handle.server</SERVICE_TYPE>
<HOST_IP>X.Y.Z.W</HOST_IP>
<HOST_IPV6/>
<HOSTDN/>
<IN_PRODUCTION>Y</IN_PRODUCTION>
<NODE_MONITORED>Y</NODE_MONITORED>
<ENDPOINTS/>
<EXTENSIONS>
<EXTENSION>
<LOCAL_ID>282</LOCAL_ID>
<KEY>handle_version</KEY>
<VALUE>7.0</VALUE>
</EXTENSION>
<EXTENSION>
<LOCAL_ID>281</LOCAL_ID>
<KEY>handle_version</KEY>
<VALUE>7.2.3</VALUE>
</EXTENSION>
<EXTENSION>
<LOCAL_ID>282</LOCAL_ID>
<KEY>role</KEY>
<VALUE>master</VALUE>
</EXTENSION>
</EXTENSIONS>
</SERVICE_ENDPOINT>
```

The first solution, using specifically *typed Service Endpoints* already supported by the ARGO monitoring framework. The only drawback can be a moderate inflation of *Service Types*, but the number will be low.

The only advantage of the 2nd solution is that any additional *Service Type* is needed. The drawback of this solution is that efforts have to be spent for standardizing the use of the extensions in the *Service Endpoints* and to update the tooling that we have available in order to interpret these extensions.

#### 4.3 Clustered services with roles, dependent of the grouping

In this scenario a *Service Endpoint* can have multiple roles. Let's take for example the scenario of a MySQL database that is provided as a service. Sites A, B and C run MySQL instances. These are registered as three different *Service Endpoints* of the *Service Type* *mysql.server*. Community X uses this service from Sites A and B with the configuration that Site A is hosting the *master* database, while Site B hosts a *slave* database. Community Y uses also this service, but in this set up, Site A and Site B host slave database servers, while the master database server is in Site C. In this scenario, the *Service Endpoint* at Site A is a *Master* in one configuration and a *Slave* in the other configuration.

This problem appears because the MySQL database server is multi-tenant, meaning that one *Service Endpoint* can provide access to multiple databases for different customers with different configurations. Although the resource registered in the GOCDB is *siteA.xyz:3306*, actually the customers are using *siteA.xyz:3306/community\_a* and *siteA.xyz:3306/community\_b*. With this finer grained view of the service, *siteA.xyz:3306/community\_a* can be either a *master* database or a *slave* database, but NOT both, while *siteA.xyz:3306* can be both a *master* and a *slave* database server at the same time.

This scenario can be modelled by using option 2

```
<results>
<SERVICE_GROUP PRIMARY_KEY="30G0">
<NAME>MYSQL_COMMUNITY_A</NAME>
<DESCRIPTION />
<MONITORED>Y</MONITORED>
<CONTACT_EMAIL>...</CONTACT_EMAIL>
<GOCDB_PORTAL_URL>
https://creg.eudat.eu/view_portal/index.php?Page_Type=Service_Group&id=30
</GOCDB_PORTAL_URL>
<SERVICE_ENDPOINT PRIMARY_KEY="10G0">
...
</SERVICE_ENDPOINT>
<SERVICE_ENDPOINT PRIMARY_KEY="121G0">
<HOSTNAME>db.sitea.xyz</HOSTNAME>
<GOCDB_PORTAL_URL>
https://creg.eudat.eu/view_portal/index.php?Page_Type=Service&id=121
</GOCDB_PORTAL_URL>
<SERVICE_TYPE>mysql.server</SERVICE_TYPE>
<HOST_IP>X.Y.Z.W</HOST_IP>
<HOST_IPV6/>
<HOSTDN/>
<IN_PRODUCTION>Y</IN_PRODUCTION>
<NODE_MONITORED>Y</NODE_MONITORED>
<ENDPOINTS/>
  <!-- SERVICE_ENDPOINT_EXTENSIONS -->
<EXTENSIONS>
<EXTENSION>
<LOCAL_ID>273</LOCAL_ID>
<KEY>mysql_version</KEY>
<VALUE>7.3.1</VALUE>
</EXTENSION>
</EXTENSIONS>
</SERVICE_ENDPOINT>
<!-- SERVICE_GROUP_EXTENSIONS -->
<EXTENSIONS>
  <EXTENSION>
    <LOCAL_ID>XXX</LOCAL_ID>
    <KEY>service_endpoint:121G0:db</KEY>
    <VALUE>community_a</VALUE>
  </EXTENSION>
  <EXTENSION>
    <LOCAL_ID>YYY</LOCAL_ID>
    <KEY>service_endpoint:121G0:XXX:role</KEY>
    <VALUE>master</VALUE>
  </EXTENSION>
</EXTENSIONS>
</SERVICE_GROUP>
</results>
```

## Annex E. SOFTWARE VERSIONS MONITORING FRAMEWORK

### E.1. Software products and components table.

Figure 29 illustrates a snapshot of the front-end of the software versions monitoring framework. It shows the status achieved the first 12 months of the project and provides the version numbers of the service components (*service endpoints* according to the GOCDB nomenclature). For each site the table lists the endpoints, operating systems, software products, components and their versions. It also gives *last time* when the agent running at the endpoint has contacted the server and transmitted information on installed versions and operating system. The system is being under development.

### Software products and components

	Endpoint	Op. System	Component	SubComponent	Tag at site	Prod. Tag	Changed On
BCS	opesudat02.bsc.es	SUSE 11	b2safe.irods	irods	4.0.3	NN	2016-03-26 14:05:09
				b2safe	NN	NN	2016-03-26 14:05:09
	bscirods.bsc.es		eudat.monitoring.ncsa-client	monitoring.ncsa-client		NN	2016-02-26 11:44:21
CINECA	Endpoint	Op. System	Component	SubComponent	Tag at site	Prod. Tag	Changed On
	data.repo.cineca.it		b2safe.irods	irods	4.1.7	NN	2016-03-26 07:10:10
				b2safe	3.1-0	NN	2016-03-26 07:10:10
			b2safe.dsi	dsi	1.7.3	NN	2016-02-26 11:48:06
	helpdesk.eudat.eu		eudat.helpdesk.rt	helpdesk.rt	4.2.12	NN	2016-02-26 11:44:49
	grid.hpc.cineca.it		b2access.myproxy	myproxy		NN	2016-02-26 11:45:13

Figure 29: View on the panel that informs about installed software versions.

With a few exceptions almost all B2SAFE component instances (iRODS and the B2SAFE rule set) are connected to the framework and report their actual version numbers. The reporting period is 6 hours.

Each type of service component requires an individual approach to extract the version number. If necessary the agent script which is locally testing the service component of the endpoint can be adapted for each type of service. As a next step the product versions of all other B2 services will be collected.



## Annex F. DATA PROJECTS

### F.1. Current status of projects originating from EUDAT1

The following table gives an update of the existing projects that started during the first EUDAT1 project (compare with [EUDAT-DEL-WP6-D6 4-Final Report on Operations.pdf](#), p.37).

Table 14: Status Update existing Data Projects from EUDAT1

Start	Project	Status	Enabler & Contact	Community	Discipline/Topic
Dec 12	B2STAGE+B2SAFE & VPH	replaced	E. Erastova, RZG	VPH	Lifescience, Bio Medicine
Dec 12	B2SAFE & EPOS PP WG7	replaced	C. Cacciari, CINECA	EPOS	Earth System Science
Dec 12	B2SAFE & CMIP5	terminated	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	replaced	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	L. Lehtinen, CSC	EUDAT	multiple science domains
Jan 14	B2SHARE & CLARIN	prod	L. Lehtinen, CSC	CLARIN	Linguistics, Humanities
Jan 14	B2FIND & ENES	prod	H. Widmann, DKRZ	ENES	Climatology
Jan 14	B2FIND & generic B2SHARE	prod	H. Widmann, DKRZ	EUDAT	any research data
Jan 14	B2FIND & PanData	prod	H. Widmann, DKRZ	PanData	Nuclear Physics
Jan 14	B2SAFE & MfN Berlin	prod	E. Erastova, RZG	Naturkunde Berlin	Biodiversity, Museum
Feb 14	B2FIND & CLARIN	prod	H. Widmann, DKRZ	CLARIN	Linguistics
Mar 14	B2SHARE & DRIHM	prod	P. Hyvarinen, CSC	DRIHM	Earth System Science
Mar 14	B2SHARE & BBMRI	prod	P. Hyvarinen, CSC	BBMRI, ELIXIR NL	Biobanking, Medical Sc.
Mar 14	B2SAFE & diXa	suspended	J. Heikkinen, CSC	diXa	Lifescience, Toxicology
Mar 14	B2SHARE & EUON	prod	L. Lehtinen, CSC	EUON	Ontology
Mar 14	B2SAFE & iMarine	suspended	G. Morelli, CINECA	iMarine	Biodiversity
Apr 14	B2FIND & EISCAT	replaced	H. Widmann, DKRZ	EISCAT	Earth System Sciences
Apr 14	B2FIND & ICOS	replaced	H. Widmann, DKRZ	ICOS	Earth System Sciences
Apr 14	B2SAFE & IST	replaced	P. Dugenie, CINES	University IST	University
Apr 14	B2SAFE & IPSL	prod	P. Dugenie, CINES	ENES	Climatology,
Apr 14	B2STAGE+B2SAFE & VPH (2)	prod	E. Erastova, RZG	VPH	Lifescience, Bio Medicine
Apr 14	B2FIND & SDL	prod	H. Widmann, DKRZ	SDL	Agriculture
Apr 14	B2FIND & European Lib.	prod	H. Widmann, DKRZ	European Library	Cultural Heritage
Jun 14	B2SAFE & ELIXIR	replaced	J. Heikkinen, CSC	ELIXIR	Lifescience, Biobanking
Jun 14	B2FIND & DataCite	prod	H. Widmann, DKRZ	DataCite	multiple science domains
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	prod	H. Widmann, DKRZ G. Morelli, CINECA	ALEPH/INFN/ DPHEP	High Energy Physics
Aug 14	B2SHARE & RDA	prod	P. Hyvarinen, CSC	RDA	RDA documents
Sep 14	B2FIND & BBMRI	suspended	H. Widmann, DKRZ	BBMRI	Biobanking, Medical Sc.
Oct 14	B2FIND & WDC-Terra/EP	prod	H. Widmann, DKRZ	WDC-Terra, EPOS	Earth System Sciences
Oct 14	B2SHARE & NRM	prod	P. Hyvarinen, CSC	Swedish Museum of Natural History	Biodiversity, Museum
Oct 14	B2SHARE & GBIF	prod	P. Hyvarinen, CSC	Finnish Museum of Natural History	Biodiversity, Museum
Dec 14	B2SAFE & ABC	prod	J. Heikkinen, CSC	Ascona B.DNA Consortium	Lifescience, Molecular Dynamics
Jan 15	B2FIND & IVOA	prod	H. Widmann, DKRZ	IVOA	Astronomy
Jan 15	B2FIND & CESSDA	prod	H. Widmann, DKRZ	CESSDA	Social Sciences

### F.2. New data projects

The following Table 15 provides an overview over the status of new data projects in the second phase of EUDAT. The projects are broken down into subprojects. The start dates of some of the data pilots from the 1<sup>st</sup> call are tentative.

Table 15: New Data Projects and Pilots in EUDAT2020

Project / Pilot Name	Customer Community / Inst.	(Sub)Project Start Date	Project ID	Services	Status	Enabler	Generic SP	Plan, Proposal
INGV	INGV	2015-04	47.1	B2SAFE	pre-prod	M. Fares	CINECA	1UPTP#1
KNMI	KNMI	2015-04	47.2	B2SAFE	pre-prod	M. Fares	SURFsara	1UPTP#2
GFZ	GFZ	2015-04	47.3	B2SAFE	pre-prod	M. Fares	KIT	1UPTP#3
LTER	LTER	2015-04	48.1	B2SAFE	enabling	C.J.Håkansson	Juelich	2UPTP#1
LTER	LTER	2015-04	48.2	B2SHARE	enabling	C.J.Håkansson	Juelich	2UPTP#2
LTER	LTER	2015-04	48.3	B2FIND	enabling	H. Widmann	DKRZ	2UPTP#3
DEIMS	LTER	2015-10	49.1	B2FIND	enabling	H. Widmann	DKRZ	
SOAS	SOAS / CLARIN	2015-10	52.1	B2SAFE	enabling	W. Elbers	MPCDF	4UPTP#1
ICOS	ICOS	2016-01	50.1	B2SHARE, B2AC	planning	M. Hellstrom	CSC, Juelich	5UPTP#1
ENES	DKRZ / ENES	2016-01	51.1	B2FIND B2SHAR	planning	H. Thiemann	DKRZ	6UPTP#1
IST	IST Research Cen	2016-01	1001.1	B2SAFE	pre-prod	E. Erastova	KIT	2C4C#66
		2016-01	1001.2	B2FIND	pre-prod	H. Widmann	DKRZ	2C4C#66
Herbadrop	Digitarium of the	2016-01	1002.1	B2DROP	planning	P. Dugenie	CINES	2C4C#111
		2016-01	1002.2	B2SHARE	planning	P. Dugenie	CINES	2C4C#111
		2016-01	1002.3	B2SAFE	pre-prod	P. Dugenie	CINES	2C4C#111
		2016-01	1002.4	B2STAGE	planning	P. Dugenie	CINES	2C4C#111
		2016-01	1002.5	B2FIND	planning	H. Widmann	DKRZ	2C4C#111
		2016-01	1002.6	dev:BigD	dev	P. Dugenie	CINES	2C4C#111
		2016-01	1002.7	dev:DataT	dev	P. Dugenie	CINES	2C4C#111
		2016-01	1002.8	dev:Sem	dev	P. Dugenie	CINES	2C4C#111
Aalto	Aalto University,	2016-01	1003.1	B2DROP	enabling	J. Heikinen	CSC	2C4C#125
		2016-01	1003.2	B2SHARE	enabling	J. Heikinen	CSC	2C4C#125
		2016-01	1003.3	B2SAFE	planning	J. Heikinen	CSC	2C4C#125
		2016-01	1003.4	B2FIND	planning	H. Widmann	DKRZ	2C4C#125
		2016-01	1003.5	dev:BigD	dev	J. Heikinen	CSC	2C4C#125
		2016-01	1003.6	dev:DataT	dev	J. Heikinen	CSC	2C4C#125
WebCoverage	FZ Juelich	2016-01	1005.1	B2SHARE	enabling	S. Apweiler	JUELICH	2C4C#131
		2016-01	1005.2	B2FIND	planning	H. Widmann	DKRZ	2C4C#131
SIMCODE	Alma Mata Univ	2016-01	1006.1	B2STAGE	prod	G. Morelli	CINECA	2C4C#132
		2016-01	1006.2	B2SHARE	planning	G. Morelli	CINECA	2C4C#132
		2016-01	1006.3	B2FIND	planning	G. Morelli	CINECA	2C4C#132
NFFA	CNR-IOM	2016-02	1004.1	B2SHARE, B2FIN	planning	H. v. Piggelen	STFC	2C4C#127
Westlife	STFC	2016-02	1007.1	B2DROP, B2SAFE	planning	H. v. Piggelen	STFC	2C4C#59
FAIR	Dutch Techcentre	2016-02	1022.1	B2SHARE, B2SAF	planning	Chr. Staiger	SURFsara	2C4C#134
Sphinx	ISAC-CNR	2016-03	1015.1	B2SHARE	planning	G. Morelli	CINECA	2C4C#118
		2016-03	1015.2	B2STAGE	planning	G. Morelli	CINECA	2C4C#118
TurboBase	University of Ro	2016-03	1016.1	B2SHARE	planning	G. Morelli	CINECA	2C4C#120
		2016-03	1016.2	B2STAGE	planning	G. Morelli	CINECA	2C4C#120
		2016-03	1016.3	B2FIND	planning	H. Widmann	DKRZ	2C4C#120
ClinicalTrial	Heinrich- Heine U	2016-03	1017.1	B2SHARE	planning	S. Apweiler	JUELICH	2C4C#123
		2016-03	1017.2	B2SAFE	planning	S. Apweiler	JUELICH	2C4C#123
		2016-03	1017.3	B2FIND	planning	H. Widmann	JUELICH	2C4C#123
		2016-03	1017.4	dev:DataT	dev	S. Apweiler	JUELICH	2C4C#123
TurbulentFlow	School of Aeron	2016-03	1018.1	B2SHARE	planning	M. Fares	BSC	2C4C#128
		2016-03	1018.2	B2SAFE	planning	M. Fares	BSC	2C4C#128
		2016-03	1018.3	B2FIND	planning	H. Widmann	DKRZ	2C4C#128
ClimateModel	Institut Català de	2016-04	1008.1	B2DROP	planning	M. Fares	BSC	2C4C#73
		2016-04	1008.2	B2SHARE	planning	M. Fares	BSC	2C4C#73
		2016-04	1008.3	B2SAFE	planning	M. Fares	BSC	2C4C#73
		2016-04	1008.4	B2STAGE	planning	M. Fares	BSC	2C4C#73
		2016-04	1008.5	dev:DataT	dev	M. Fares	BSC	2C4C#73
		2016-04	1008.6	dev:BigD	dev	M. Fares	BSC	2C4C#73
ResearchStude	University of Hel	2016-04	1009.1	B2SHARE	planning	J. Heikinen	CSC	2C4C#89
Europeana	Europeana Found	2016-04	1010.1	B2STAGE	planning	M. Brzezniak	PSNC	2C4C#94
		2016-04	1010.2	B2FIND	planning	H. Widmann	DKRZ	2C4C#94
CloudyCulture	National library c	2016-04	1011.1	B2DROP	planning	P. Dugenie	EPCC	2C4C#106
		2016-04	1011.2	B2SAFE	planning	P. Dugenie	EPCC	2C4C#106
		2016-04	1011.3	B2STAGE	planning	P. Dugenie	EPCC	2C4C#106
EISCAT	EISCAT scientific	2016-04	1013.1	B2SHARE	planning	J. Heikinen	CSC	2C4C#114
		2016-04	1013.2	B2SAFE	planning	J. Heikinen	CSC	2C4C#114
		2016-04	1013.3	B2FIND	planning	H. Widmann	DKRZ	2C4C#114
		2016-04	1013.4	dev:DataT	dev	J. Heikinen	CSC	2C4C#114
		2016-04	1013.5	dev:Viz	dev	J. Heikinen	CSC	2C4C#114
CervixCancer	Uppsala Universi	2016-04	1021.1	B2SHARE, B2DR	planning	C. J. Håkansson	SNIC	2C4C#133
DataPublicatio	University of Por	2016-05	1012.1	B2SHARE, B2DR	requested	E. Erastova	MPCDF	2C4C#110
Tokamak	Culham Centre fo	2016-05	1014.1	B2DROP B2SAFE	requested	(tbd)	STFC	2C4C#117
B2ANNO	Natural Environm	2016-05	1019.1	B2SHARE, B2FIN	planning	H. Widmann	STFC	2C4C#129
AirQuality	LIFE+RESPIRA	2016-05	1020.1	B2SHARE, B2SAF	planning	(tbd)	BSC	2C4C#130
OCR	University of Leip	2016-06	1023.1	B2SHARE, B2DR	requested	(tbd)	MPCDF	2C4C#135