H2020-INFRAEOSC-2018-3

NI4OS-Europe

National Initiatives for Open Science in Europe

Deliverable D5.1

Provider landscape analysis and provider categorization

Lead beneficiary(s):	UKIM
Author(s):	Boro Jakimovski (editor), All partners
Status –Version:	Final – k
Date:	May 13, 2020
Dissemination Level:	Public

Abstract: Deliverable D5.1 – Provider landscape analysis and provider categorization – presents the collection of detailed information about resource providers – generic computing services, generic data storage services and generic data management services, thematic (discipline-specific) services, and repositories (publication repositories, datasets repositories, software repositories). The information collected contains technical, operational and support information about the candidate services.

The NI4OS-Europe project is funded by the European Commission under the Horizon 2020 European research infrastructures grant agreement no. 857645.

The NI4OS-Europe Consortium consists of:

GRNET SA	Coordinator	Greece
ATHENA RC	Beneficiary	Greece
CYI	Beneficiary	Cyprus
UCY	Beneficiary	Cyprus
IICT	Beneficiary	Bulgaria
SRCE	Beneficiary	Croatia
RBI	Beneficiary	Croatia
KIFU	Beneficiary	Hungary
DE	Beneficiary	Hungary
ICI BUCURESTI	Beneficiary	Romania
UEFISCDI	Beneficiary	Romania
ARNES	Beneficiary	Slovenia
UMUKM	Beneficiary	Slovenia
IPB	Beneficiary	Serbia
UOB	Beneficiary	Serbia
RASH	Beneficiary	Albania
UNI BL	Beneficiary	Bosnia-Herzegovina
UKIM	Beneficiary	North Macedonia
UOM	Beneficiary	Montenegro
RENAM	Beneficiary	Moldova (Republic of)
IIAP NAS RA	Beneficiary	Armenia
GRENA	Beneficiary	Georgia

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
31.01.2020	a	Boro Jakimovski	Initial ToC defined
06.03.2020	b	Dushan Vudragovic	Updated ToC. Section 2
20.03.2020	c	Anastas Mishev	Section 1
25.03.2020	d	Boro Jakimovski	Section 3
03.04.2020	e	Brina Klemenčič	Section 5
14.04.2020	f	Andreas Athenodorou	Section 4
20.04.2020	g	Anastas Mishev	Conclusion
25.04.2020	h	Boro Jakimovski	Final editing
28.04.2020	i	Milutin Radonjić	Internal quality check
29.04.2020	j	Boro Jakimovski	Addressing quality check remarks
13.05.2020	k	Kostas Koumantaros, Ognjen Prnjat	Final Review

Table of contents

References

[1] Project NI4OS-Europe-857645 - Annex I - Description of the Action

[2] AGORA - Service Portfolio Management Tool (AGORA/SPMT) https://grnet.github.io/agora-sp/

[3] FitSM – Standards for lightweight IT service management - https://www.fitsm.eu/

[4] Briefing Paper - EOSC Federating Core Governance and Sustainability, https://www.eosc-hub.eu/publications/briefing-paper-eosc-federating-core-governance-and-sustainability

[5] OpenAIRE Guidelines - https://guidelines.openaire.eu/

List of Figures

List of Tables

List of Acronyms

AAI	Authentication and Authorization Infrastructure
API	Application Program Interface
DANS	Data Archiving and Networked Services
DEI	Days of E-Infrastructure 2020
DPO	Data Protection Officer
EOSC	European Open Science Cloud
FAIR	Findability, Accessibility, Interoperability, Reusability
FitSM	Federated IT Service Management
IPR	Intellectual Property Rights
ITSM	IT Service Management
LMS	Learning Management System
MOOC	Massive Open Online Course
OLA	Operational Level Agreement
ORDM	Open Research Data Management
OS	Open Science
REST	Representational State Transfer
SCORM	Sharable Content Object Reference Model
SLA	Service Level Agreement
TRL	Technology Readiness Level

Executive summary

What is the focus of this Deliverable?

The main aim of this deliverable is to analyse the landscape of the services and repository providers from the NI4OS-Europe participating countries and to provide the categorization of the resources that will be candidates for on-boarding by the NI4OS-Europe project. Several different categorization perspectives will be used to categorize the resources, based on their levels of technology readiness, EOSC integration levels and management integration levels.

The deliverable builds on the initial WP2 landscape analysis, with the collection of detailed information about resource providers – generic computing services, generic data storage services and generic data management services, thematic (discipline-specific) services, and repositories (publication repositories, datasets repositories, software repositories). All this information about the different resources will be collected in the AGORA service catalogue and portfolio management system. Service operational details for each service is also be collected, including user support channels, security channels, monitoring, accounting, access policies, conditions of access, SLA approach, etc. The checklist of actions per level of integration towards harmonisation with operational requirements defined within the WP3 is then generated per service.

Based on all collected information, a timeline for resources on-boarding is defined, representing the roadmap throughout the project. The roadmap is subject to change depending on the evolution of on-boarding rules and processes within EOSC – something which is not under NI4OS-Europe control.

What is next in the process to deliver the NI4OS-Europe results?

This deliverable holds the timeline for the resources on-boarding within the timeline of the project. It will be followed by the D5.2 “First report on provider and repository integration” in M16, as well as the D5.4 “Second report on provider and repository integration and horizontal service delivery” in M34. The update of the service catalogue will be described in D5.3 in M19. In parallel, the lessons learnt from the on-boarding process will be used to update the D3.4 “Best practices for on-boarding and related policies 2nd version”, in M19.

What are the deliverable contents?

The deliverable starts with the categorization of the resource providers in the NI4OS-Europe partnering countries and the resources they are providing. A set of relevant attributes is identified, depending on the type of the resource. Next, three types of classifications of the resources are suggested. They are based on the Technology Readiness Level (TRL), EOSC Integration Level (EIL), and Management Integration Level (MIL). They are used to produce the cumulative readiness matrix. Based on those levels and the described methodology, the deliverable proposes the categorization of all NI4OS-Europe resources to be on-boarded to the EOSC. The on-boarding activities will be coordinated by the on-boarding team, with representatives per each participating country and resource type. A categorization of different types of resources, such as generic services, thematic services and repositories. Finally, a detailed integration timeline for the on-boarding of the identified resources to EOSC is presented.

Conclusions and recommendations

Using the classification and categorization of the resources provided in this document, the on-boarding of the resources from the NI4OS-Europe partner countries will take place. This document represents the main guide for the NI4OS-Europe on-boarding team to successfully collect relevant metadata and following the proposed timeline, enable the on-boarding of those resources to EOSC. The document also provides metrics to evaluate the maturity of resources and their providers, as well as a detailed timeline of the expected activities to be undertaken by the resource providers to make their resources EOSC ready.

1 Introduction

The ultimate goal of all EOSC related initiatives is the establishment of a single point of reference for scientific services and data for the pan-European research communities. Toward the fulfilment of this goal, an important role has been assigned to the EOSC implementation projects, with NI4OS-Europe being one of them. Aside from the policy related actions that include the establishment of the open science initiatives in the partnering countries, as well as the promotion of the FAIR data principles, these projects have an important technical goal to achieve. This technical goal is focused on identifying, integrating and on-boarding of the national and regional resources, including services, repositories and data, into the common EOSC pool of resources.

In the preparatory phase of the project, as well as in its initial months, lots of effort was concentrated on the identification of such resources. A comprehensive regional landscaping study was performed by the project partners in their home countries to identify the national resources that could be of pan-European interest and potentially be on-boarded to EOSC. As a result, more than 90 resources were identified as candidates for such on-boarding.

Additionally, effort was made, in cooperation with other relevant projects, task forces and working groups, to identify the relevant policies and rules of participation for these resources providers and their resources to be on-boarded to EOSC. These policies and rules, documented in the D3.1 “Best practices for on-boarding and related policies”, represent the basics for the additional detailed categorization defined later in this document.

To better coordinate the effort of the partners in the on-boarding process, a need to establish an on-boarding team, with persons responsible for different types of resources per country was identified. The content of the team and their roles are presented in this document. They will be guided by the on-boarding timeline, also part of this document, to successfully integrate, verify and on-board national resources to EOSC.

The deliverable is comprised of the following sections. Section 2 defines the categorization of the resource providers, based on the three different categorizations. The categorization criteria include Technology Readiness Level (TRL), EOSC Integration Level (EIL) and Management Integration Level (MIL). The section finished with the on-boarding timeline production and the team responsible for its implementation. Section 3 describes in detail the generic services, including HPC, cloud and storage services. Section 4 gives the details regarding the thematic services, while section 5 is focused on the repositories. Finally, in Section 6, a detailed integration timeline is defined for the first set of services.

2 Categorization of resource providers

During the first six months of the project, WP5 developed a methodology based on which a plan for resource providers on-boarding to the EOSC is created. The development started by collecting detailed information about resource providers in our region. As it was defined in the project's DoA, our focus was on generic computing services, generic data storage services, and generic data management services, thematic (discipline-specific) services, and repositories (publication repositories, datasets repositories, software repositories). Although we have identified an initial set of prospective regional resources at the proposal stage and extended this list based on the WP2 landscape survey information, the proper analysis that led to the EOSC integration timeline plan required an additional, more technical information. For example, for Cloud computing resources, supplementary data incorporate CPU specification, the total number of nodes, cores per node, RAM per core, etc.; for storage resources: filesystem type, read/write performance, total storage, supported interfaces, bandwidth to the Internet, etc.; for high-end computing resources: type and number of accelerators, type of interconnectivity and bandwidth, CPU peak performance, accelerator peak performance, etc.; for repositories: metadata schemas, access methods, etc. All these parameters, criteria for data collection, are stated clearly in the corresponding sections of this deliverable for different types of resources.

Once the information about the regional resource was collected, we have tried to identify a minimal set of metadata that is common to all types of resources taken into consideration, and based on these, to produce a general resource description template regardless of the type of resources. Such a template is given in Appendix I, and it accumulates:

- basic information, such as the name of the resource and its endpoint;
- marketing information, such as resource tagline, description, logo, website, target communities;
- classification information, scientific domain, and category;
- location information, in terms of geographical availability and language;
- resource provider-related information.

Although the template was deducted from the limited number of resource types, on which the project is mainly focused, we did not notice any major problem in the usage of the same template for any type of resources, ranging from experimental devices and e-Infrastructures, sharing and discovery, processing and analysis, security and operations, training and support, aggregators and integrators, etc. During the template construction, we have used the resource description template previously developed by the EOSC-Hub project[1] and eInfraCentral[2], and more recently we have also taken into account the EOSC Resource & Provider Description Specification and Classifications developed by the EOSC-Enhance[3] project.

After this, we ended up with the list of resource-type specific metadata. We have organized these into dedicated templates for each resource type. It happens to be that these are

[1] https://www.eosc-hub.eu

[2] https://einfracentral.eu

[3] https://www.eosc-portal.eu/enhance

more technical properties, practically, technical specifications of the resources. For generic computing services, generic data storage services, and generic data management services, thematic services, and repositories, the content of the templates are described in Sections 3, 4 and 5, respectively.

Both generic and technical specifications of the resources are non-EOSC-related aspects of the resource description. More interesting for us were EOSC-related aspects that drove the NI4OS-Europe on-boarding timeline production:

- Technology Readiness Level (TRL),
- EOSC Integration Level (EIL),
- Management Integration Level (MIL).

Figure 1 summarizes different aspects of a resource description that will be gathered by the NI4OS-Europe resource portfolio system.

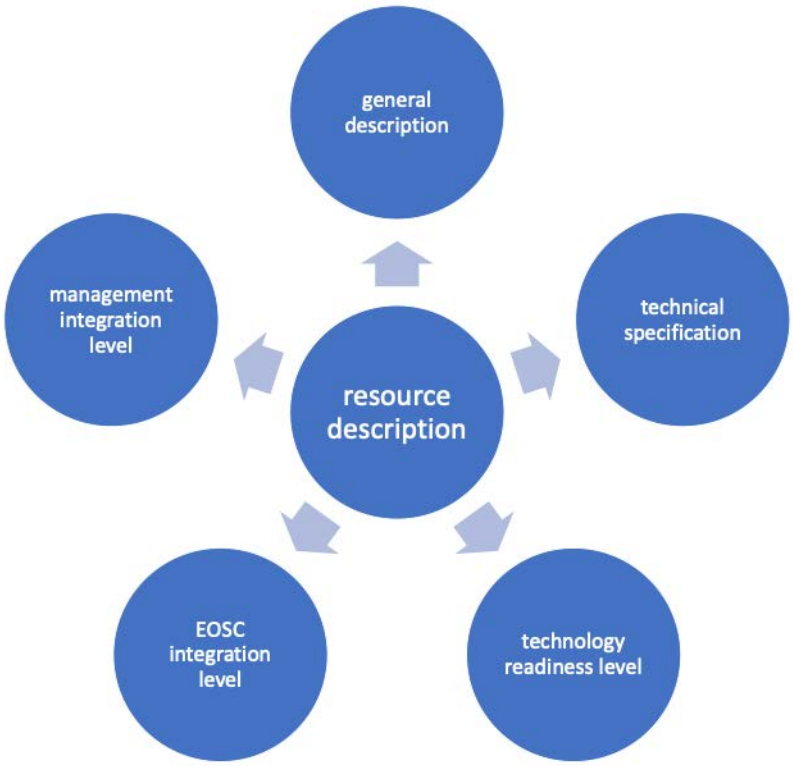


Figure 1. Different aspects of a resource description

Aspects of the resource description presented in this section do not include the scientific impact of a particular resource, which will be equally considered during the on-boarding process. The scientific impact of the particular service will be assessed within the WP6, which will directly affect the selection of the services to be on-boarded to the EOSC.

2.1 Types of categorization

Resources can be categorized based of different aspects. These aspects are described in this chapter.

2.1.1 Technology readiness levels

We will use the Technology Readiness Level (TRL) to assess a resource development stage. Figure 2 illustrates nine different stages that TRL oversees in the process of development. From the on-boarding perspective, only high-level TRLs are of interest, i.e., we will initiate the on-boarding process only for resources with TRL > 7. However, in our portfolio system, we will also collect and describe resources that are currently under development, with TRL < 8. We strongly believe that EOSC could influence the course of the development, and therefore we would like to be able to track the progress of a particular resource and inform the developers about EOSC features and functionalities that could be integrated and reused in the early resource development stage. We expect that the integration with a component from the pre-production environment could be easier accomplished during the process of development of a resource than later, when the resource is mature enough and when low-level functionality (provided by the pre-production environment) needs to be redesigned. At the early stage, the developers will be able to design the resource in a way suitable for integration with the project's pre-production environment.

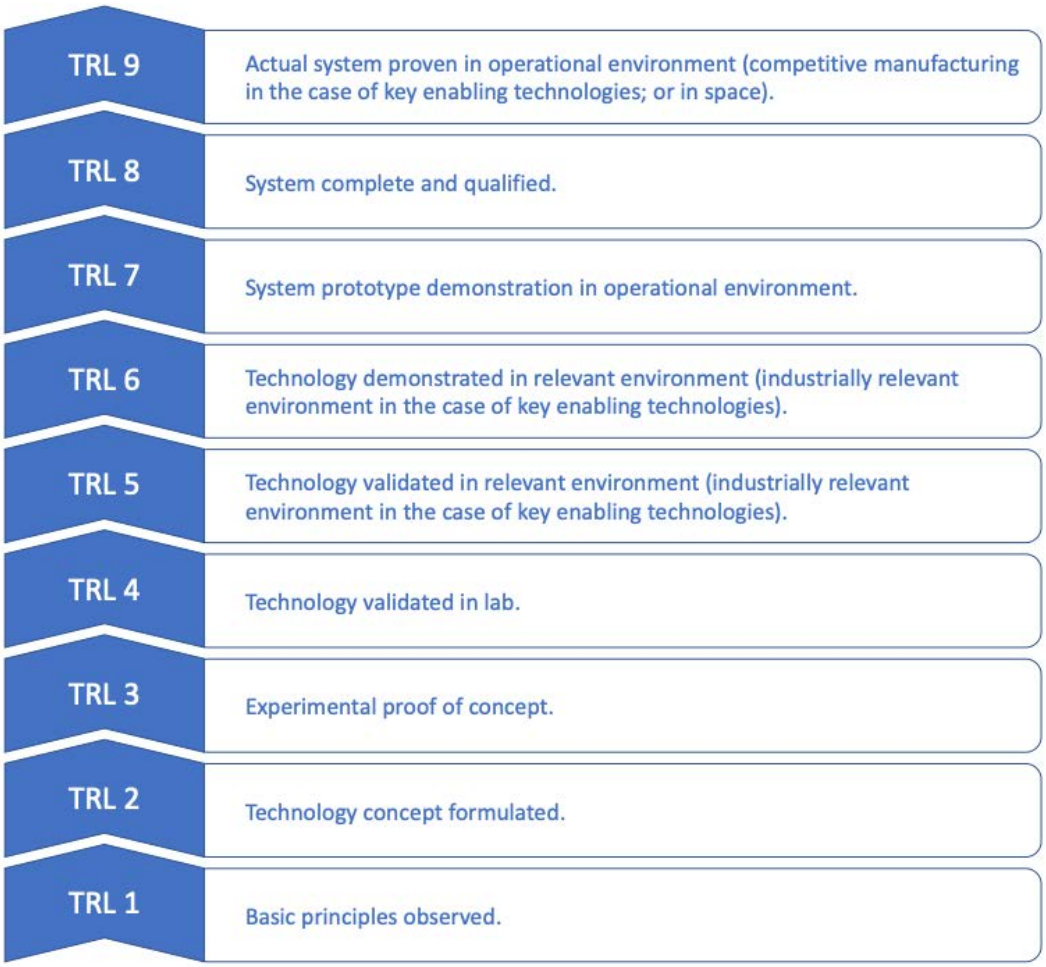


Figure 2. Different technology readiness levels in the process of resource development

2.1.2 EOSC integration levels

Inspired by the TRLs, to be able to measure the EOSC integration levels, we have put on the same equal footing various levels of integration that could be achieved. We have tried to organize gradually these levels based mainly on integration complexity and the necessity of a particular integration. Figure 3 illustrates the NI4OS-Europe EOSC Integration Levels (EILs) to be accomplished by a resource.

At the first level, all resources (both services and repositories) have to be described within the project's portfolio system using the generic and technical resource description templates. At the next stage, EIL 2, the resource user manual has to be prepared, as well as resource's terms of use. For the user manual production, resource owners will be granted access to the project's Wiki system, while creation of the resource's terms of use will be performed by the NI4OS-Europe on-boarding team. At the third level (EIL 3), a privacy policy is expected, as well as integration with the project's helpdesk system, which will be used as the main resource support channel. As can be noticed from Figure 3, the first three levels require the very same aspects of integration for all three types of resources under the project's consideration.

Figure 3. NI4OS-Europe EOSC Integration Levels for generic services (blue), thematic services (green), and repositories (orange)

To achieve EIL4, resource providers will be encouraged to perform integration with the AAI, if there is benefit from such integration. All resources will be equipped with the unique ID through the PID system integrated within the catalogue. The Resource Provider will have to accomplish MIL1 or higher in order to ensure that the appropriate management

structures are set-up behind the resource. Wherever it technically feasible we strongly recommend that services are integrated with NI4OS Federated AAI and Monitoring Services in order to evaluate their readiness for the integration with the equivalent EOSC Core Services. In addition, for thematic services, the source code has to be uploaded to the project's repository. EIL 5 requires an integration with the monitoring system, and for thematic services an additional integration with an extra generic service. It is expected that, by default, each thematic service will depend on generic service. For example, the majority of our thematic services depend on HPC/HTC generic infrastructure, and in order to produce results, utilize generic services. Thus, these will be very early integrated with the generic services. In such a case, EIL 5 expects integration with an additional generic resource that will increase the capacity of the thematic service, but also the software stack of the generic service resource. So, both resources will benefit from the integration, which is the main criterion of EIL 5. At level EIL 6, the usage of the resources will be measured by the accounting system. Also, services will deliver training materials, while thematic services will produce datasets registered and fully searchable within a dataset repository.

The following EIL levels (EIL > 6), for all types of resources, are primarily oriented to the resource management aspects, which are explained in the next section. However, in the case of thematic services, production of the training materials and API interfaces is expected at EIL 7. The creation of APIs here corresponds to the deconstruction of a service into independent building blocks, which could be utilized separately by an external service or as a part of the workflow process. Such a workflow process that uses separate EOSC building blocks, provided by generic or thematic services and repositories, is foreseen as an ideally integrated EOSC resource at EIL 8 and 9.

The levels of integration presented in this section might not be of interest to a particular resource in the case when direct benefit from integration is not apparent. In such cases, we will consider that a resource accomplished a particular level of integration, and it will be able to proceed with higher-level integrations. In particular, this is expected at EIL 4 - EIL 6 levels. For example, at the moment, AAI integration with HPC resources is not supported by design, so we will not be able to perform this kind of integration. However, in such case, we will continue with integration activities allowing resources to achieve a higher level of integration, although some lower-level requirement is not fully established.

2.1.3 Management integration levels

The resource management integration level (MIL) procedures and policies will ensure the practical implementation of various Rules of Participation produced by WP2, WP3, and WP4, and indirectly implement recommendations presented by EOSC WGs. Technically, although resource management reflects the achieved level of integration with the EOSC, it imposes some concrete obligations on the resource providers related to the resource maintenance and operations, and therefore it is considered independently here. Using the same approach as for the TRL and EIL categorization, we have introduced nine different levels that the resource could reach in the integration with the project's resource management procedures and policies. Figure 4 illustrates classification of resource management types of procedures introduced in the deliverable D3.1. Here we have tried to align the management integration levels with the EOSC integration levels by identification of tools relevant for the performance measurement, i.e., essential to practical procedure implementation.

At the lowest level, MIL 1 offers a set of procedures and policies for interaction with the resource portfolio/catalogue system. So, in this case, the Service or Resource Portfolio Management Tool (AGORA) is a tool essential for practical implementation of a procedure. Realistically, all resources start communication with EOSC by the provision of resource information within the resource portfolio system. Therefore, from the project's perspective, it is crucial to produce MIL 1 related documentation very early, i.e., from a resource provider's perspective, to reach MIL 1 level of integration.

At EIL 3, the resources will be integrated with the helpdesk and equipped with the Terms of Use and privacy policies. From the MIL perspective, this is followed by procedures that ensure consistent and coordinated security operations across the resources provided in the portfolio (MIL 2), as well as procedures for incident and resource request management (MIL 3) that are more oriented towards categorization, prioritization, and escalation of the requests/incidents in the different support units of the helpdesk.

MIL 4 further analyses support provided within the helpdesk and proposes procedures for the periodic incident trend analysis required to detect recurrent problems. MIL 5 directs improvements in the relationship between customers and resource suppliers, defining resource order management, management of the stakeholder information, etc. At level MIL 6, the resource is fully integrated with the operational tools, and the project is able to measure its performance in terms of utilization, provided support, availability, and reliability. At this stage, a discussion can be opened regarding an agreement between a resource provider and an entity that will specify conditions under which the services will be provided to end-users - Service Level Management (SLM). Such an arrangement can also cover aspects of service availability and continuity management (SACM).

Further MIL development incorporates capacity management (CAPM) at level 7 in the sense of dedicated resources for the realization of a call for a production use of services. At the same level, we have added service reporting management (SRM), which defines processes related to identification, definition, production, and distribution of various reports. The integration with the project's configuration management system (CONFM), together with the release and deployment management (RDM), is required at MIL 8. The goal of these is to plan and oversee the implementation of approved changes into production. At the highest level, MIL 9, we grouped procedures for change management (CHM) and continual service improvement (CSI). The CHM-related procedures define how a request for a change is handled, identify standard, non-standard and emergency changes, calculate the risk level, etc., while the CSI-related procedures describe the steps that need to be taken for managing of the audits. In particular, how to identify, record, prioritize, evaluate, and approve an opportunity and suggestions for improvement, how to manage and review the status and progress of improvements, etc.

Figure 4. NI4OS-Europe resource management integration levels

2.2 Some technical aspects of implementation

All five aspects of a resource description that will be gathered by the NI4OS-Europe resource portfolio/catalogue system (Figure 5) could be organized as a relational database. In such an approach, the common resource description could be established within a single table, which will contain a basic description of all resources. Fields of such a table are described in Appendix I.

The resource type field will reference an additional table with the technical specification of the resource. Due to the heterogeneity of our resources, we expect that there will be an independent table for each resource type. In the case of NI4OS-Europe, these could be generic computing, generic data, generic data management, thematic, publication repository, dataset repository, and software repository tables illustrated on the right side of Figure 5. Fields in these tables are described in Sections 3, 4 and 5.

In order to avoid duplication of the resource provider information, an independent table could be used for this purpose and linked to a particular resource. A minimal set of fields

that describe resource providers is given in Appendix I, and this will be collected for all NI4OS-Europe resources.

The integration-related information could be, as well, organized into independent tables and linked to a resource from the common description table. Here we want to accumulate desired and achieved levels of integration per resource. The first one is used for on-boarding timeline production presented in Section 6, while with the second one we will be able to track integration progress per resource, and, if it becomes necessary, to modify the initial on-boarding timeline. The integration-related tables are illustrated in the left side of Figure 5. Since their content is similar, for illustration purposes only the fields at the desired level of integration of a thematic service table are described in Appendix I.

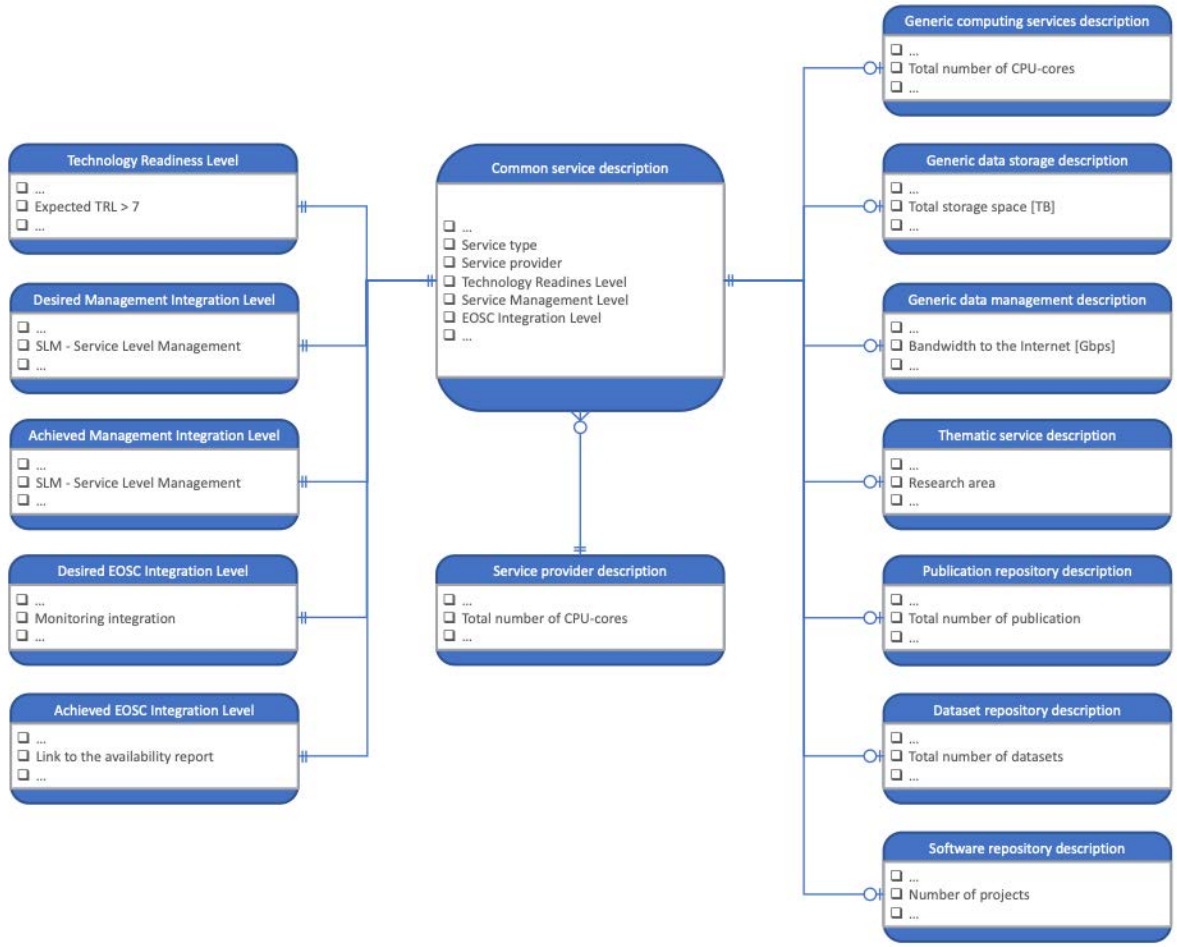


Figure 5. Relation-oriented view of resource description information

The technical aspects of implementation that we have shown here are already explored by the AGORA Resource Portfolio Management Tool [2]. This key element of the NI4OS-Europe pre-production environment facilitates all aspects of the resource description, as well as service management in IT service provision. AGORA allows maintaining and managing of the descriptions of resource in the portfolio and allows managing the transition from the portfolio to the catalogue. The service management system has been designed to be compatible with the requirements for service portfolio management

according to FitSM IT [3]. All these features set this service apart from other available IT management solutions, which has oriented us toward this tool.

2.3 On-boarding timeline production

In the deliverable D3.1 “Best practices for on-boarding and related policies”, the WP3 team has identified three different stages that a resource can reach during the integration with the EOSC: low, medium, and high. In this section, we layout the previously introduced TRL, EIL, and MIL categorizations onto this more descriptive classification. The primary purpose of this is to identify a minimal integration level that a resource has to reach to be exposed through the NI4OS-Europe catalogue, as well as minimal integration level that a resource has to reach to be distinguished as an EOSC resource, i.e., to be listed within the EOSC catalogue. The requirements summarized in Figure 6 are created based not only on the NI4OS-Europe internal considerations, but also based on numerous discussions within the EOSC working groups (primarily the Architecture WG), as well as On-boarding Task Force group established by EOSCsecretariat.eu. It is important to note that these requirements are subject to change depending on the evaluation of EOSC rules and requirements, and are not under NI4OS-Europe control.

The conditions shown in Figure 6 should not be considered as a static set of rules. They reflect the current EOSC development stage and will evolve through time following the EOSC expansion. We believe that the EOSC-core resources that provide the means to discover, share, access, and re-use data and resources have to evolve dynamically with the EOSC-Exchange layer [4], which holds all on-boarded resources.

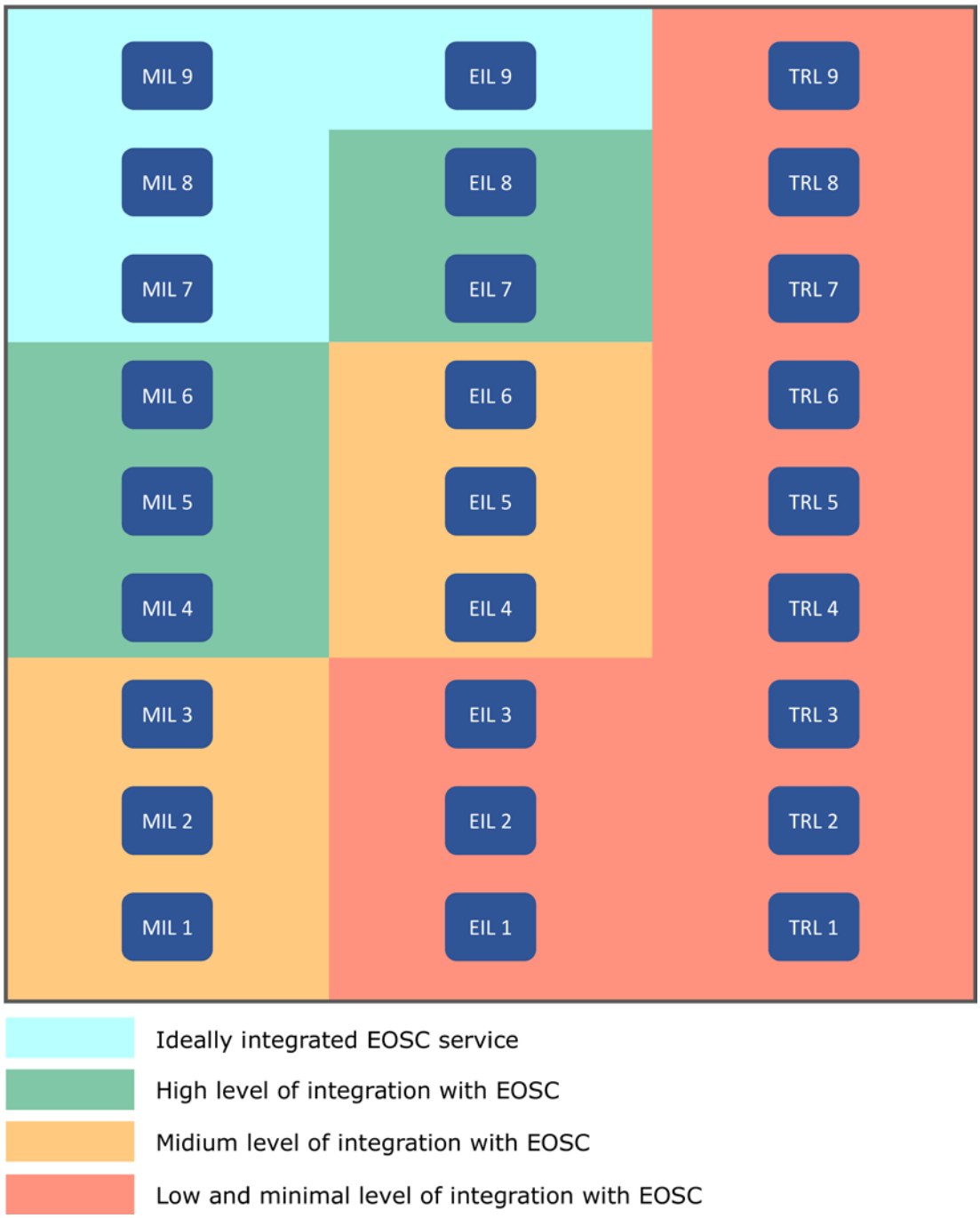


Figure 6. Different cumulative levels of integration with EOSC from the NI4OS-Europe perspective

Therefore, NI4OS-Europe identifies the following cumulative levels of integration:

- Low and minimal level of integration with EOSC, requires TRL 9[4] and EIL 3. No MIL requirements at this stage.

[4] Exception for HPC systems that are TRL 8 and above but have minimal or low integration levels.

- Medium level of integration with EOSC, requires TRL 9, EIL 6, and MIL 3.
- High level of integration with EOSC, needs TRL 9, EIL 8, and MIL 6.
- In addition, we have introduced the Ideally integrated EOSC resource, which requires level 9 of all three different aspects of integration.

Based on the methodology described in this section, we have categorized all NI4OS-Europe resources to be on-boarded to the EOSC. The process of a resource on-boarding into the EOSC will comprise all practical activities taken to incorporate a research resource into the EOSC federation. These activities represent a wide range of support actions to be directly offered to the resource provider in the on-boarding process and achieving the desired levels of different aspects of integration. For example, for a particular resource, this could be the establishment of the support channel, integration with the existing EOSC resources, data FAIRification, integration with monitoring, accounting, or authentication/authorization frameworks, preparation of end-user tutorials, access policies, or terms of use, etc. Therefore, it is crucial to establish an on-boarding team within the project with wide-ranging expertise and experience to cover all diverse aspects of the on-boarding process. Furthermore, this team has to produce a set of reusable guidelines, best practices, and other recommendations to reduce barriers the resource providers face in the EOSC integration. During the first six months of the project, we managed to assemble such a team, whose members are listed in Table 1. These colleagues provide direct support to the providers from their countries in the on-boarding process.

Table 1. NI4OS-Europe resource on-boarding team

Country code	Generic service support	Thematic service support	Repository support
GR	T. Zamani (GRNET)	T. Zamani (GRNET)	E. Papadopoulou (ATHENA)
CY	G. Tsouloupas (CYI)	C. Constantinou (CYI) A. Athenodorou (CYI)	S. Koukounidou (UCY)
BG	S. Yordanov (IICT)	S. Ivanovska (IICT)	P. Stanchev (IMI) G. Simeonov (IMI)
HR	K. Zailac (SRCE) E. Imamagić (SRCE) D. Davidović (RBI)	K. Posavec (SRCE) D. Davidović (RBI)	J. Stojanovski (RBI) K. Posavec (SRCE) D. Celjak (SRCE)
HU	T. Máray (KIFU)	K. Mohácsi (KIFU)	A. Szaldobagy (UD)
RO	A. Stanciu (ICI) A. Dinu (ICI)	A. Stanciu (ICI) A. Dinu (ICI)	A. Stanciu (ICI) A. Dinu (ICI)

SI	D. Harisch (ARNES)	D. Harisch (ARNES)	B. Klemenčič (UMUKM) D. Legat (UMUKM)
RS	P. Jovanović (IPB) Đ. Trajković (IPB)	P. Jovanović (IPB) Đ. Trajković (IPB)	M. Ševkušić (UOB) B. Kosanović (UOB)
AL	A. Malaj (RASH)	A. Malaj (RASH)	A. Malaj (RASH)
BA	M. Savić (UNI BL)	M. Savić (UNI BL)	M. Savić (UNI BL)
MK	B. Jakimovski (UKIM) V. Kjorveziroski (UKIM)	A. Mishev (UKIM) V. Kjorveziroski (UKIM)	B. Jakimovski (UKIM) A. Mishev (UKIM)
ME	Ž. Zečević (UOM)	Ž. Zečević (UOM)	Ž. Zečević (UOM)
MD	N. Iliuha (RENAM)	A. Golubev (RENAM) N. Cheradi (RENAM)	A. Golubev (RENAM) N. Cheradi (RENAM)
AM	W. Narsisian (IIAP) H. Astsatryan (IIAP)	W. Narsisian (IIAP) H. Astsatryan (IIAP)	W. Narsisian (IIAP) H. Astsatryan (IIAP)
GE	T. Maisuradze (GRENA) R. Kvatadze (GRENA)	R. Kvatadze (GRENA)	T. Gvenetadze (GRENA) T. Kvatadze (GRENA)

2.4 Overview of NI4OS-Europe landscape

In total, we have collected 97 resource descriptions, out of which 27 are generic services, 38 are thematic services, and 32 are repositories. These should be considered as candidates for on-boarding to the EOSC. From this pool of resources, we will select 20 generic services, 20 thematic services, and 15 repositories (KPIs defined in the DoA), which will be on-boarded into the EOSC within the framework of the NI4OS-Europe project. The selection will be based on their desired level of integration, technical capabilities to achieve such an integration, and relevance of a particular resource to pan-European scientific communities. Since the EOSC is still a dynamic environment, it was important for us to identify at least 50% more resource candidates than KPIs require. In this way, if we face any major technical obstacle that might be essential for the EOSC

integration, but due to practical reasons impossible or very hard to implement for a particular resource, we will be able to find an appropriate replacement.

The majority of **generic services** were identified during the project preparation. These kinds of resources within the NI4OS-Europe region have high visibility at the national level due to their generic capabilities to address technical needs common to various research areas. Also, many of them have already been utilized by the European research communities through various open calls - in many cases for production HPC use. On the other hand, the deployment of this kind of infrastructure requires high expertise in different technology fields, and its establishment requires significant effort and time. Therefore, we were able to identify almost all of them during the project preparation. Practically, only the ICIPRO Cloud infrastructure from Romania was identified through the WP2 landscape activity as a new generic service within the region. Its procurement was recently accomplished by the ICI (Bucharest), while its production deployment is expected to happen in the year 2021.

As far as **thematic services** are concerned, we tried to report as many as possible during the proposal preparation to create a pool of resources from which we will be able to choose the ones suitable for the EOSC integration during the project lifetime. Besides mature services, we also reported services that are in the pilot and testing phase, as well as services that were under development at the time. Mainly due to this fact, 29 out of 38 thematic services were identified at the proposal stage, and the rest of them were brought to our attention through the landscape activity. The situation with the **repositories** is a little bit different. Approximately half of them were identified during the project's proposal preparation (17 out of 32) and the other half during the first months of the project by the WP2 team. This is mainly related to technical aspects of the integration of repositories with the EOSC. At the moment, the usage of the OpeanAIRE platform within the EOSC working groups and task forces is recognized as a more appropriate way for measuring the repository performance. Practically, it means that all repositories already integrated within the OpenAIRE could be considered as on-boarded. Therefore, from our initial list of candidates for the on-boarding, we have excluded the OpenAIRE-integrated repositories. Fortunately, the WP2 landscape survey collected enough repository end-points, so it was easy for us to replace already on-boarded OpenAIRE repositories with the new ones, which will be incorporated within the OpenAIRE platform via the project as a part of the EOSC on-boarding activity.

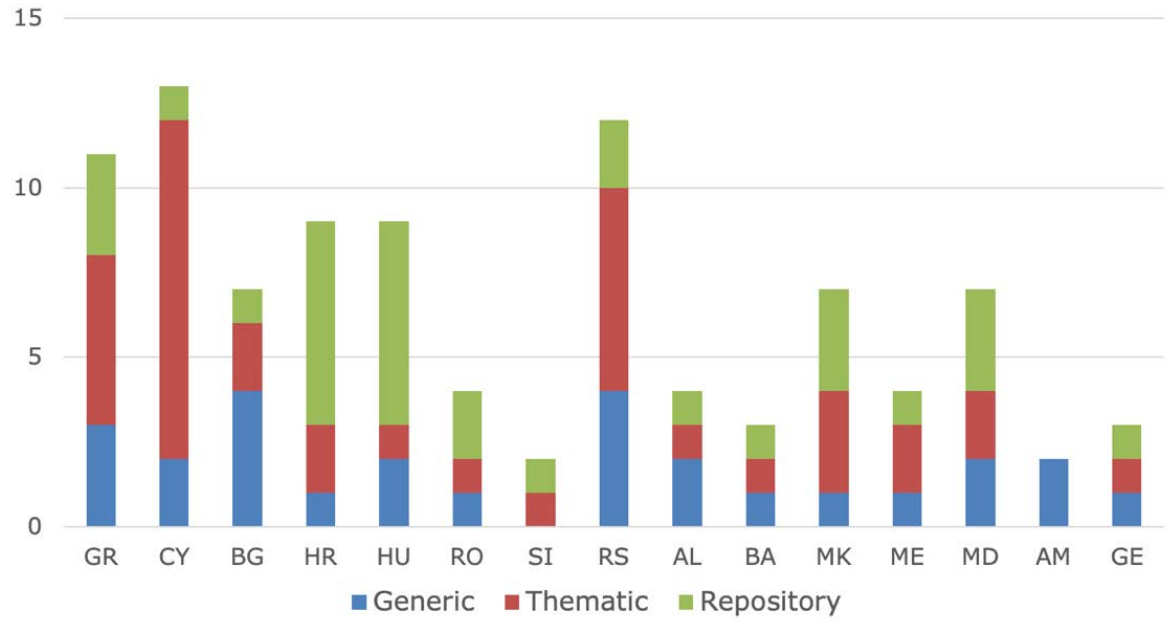


Figure 7. Per country distribution of the resource candidates to be on-boarded

Figure 7 illustrates the distribution of the resource candidates to be on-boarded per country. In this stacked bar chart, segments in blue, red, and green colours give a number of generic and thematic services and repositories per country, respectively. From each country, we have collected details about at least two resources, while in the bulk of cases, this number is much higher. The significant number of the reported repositories are from Croatia and Hungary (6 per country), the notable number of thematic services are from Cyprus (10), Serbia (6) and Greece (5) while Serbia (4), Bulgaria (4) and Greece (3) reported a majority of generic services.

3 Categorization of generic services

As described in the previous section, after the analysis of the service providers and services, we ended up with the list of resource-type specific metadata. We have organized these into dedicated templates for each resource type. In this section, we give a detailed description of metadata collected for the generic computing services.

Generic computing services are classified in 3 subcategories:

- HPC/HTC services
- Cloud services
- Storage services

Each subcategory has specific metadata collected, which describes specific service performance, capacity and usage capabilities. The data collected using these service templates will be useful for NI4OS thematic services, as well as NI4OS end users in order to adapt and appropriately choose specific resources for their use.

3.1 HPC/HTC services

High Performance Computing (HPC) plays an important role in the field of computational science, and are used for a wide range of computationally intensive tasks in various fields. These, so called supercomputers, require users to use their specific programming models, which enable scaling of computationally intensive applications. Therefor HPC services need to be described as specific as possible, in order end users to be able to identify specific programming and technology requirements.

High Throughput Computing (HTC), or also known as Grid computing, is the use of widely distributed computer resources to reach a common goal. A computing grid can be thought of as a distributed system with non-interactive workloads that involve many files. Usually HTC services represent smaller computing clusters, which are constructed with general-purpose computing resources, and standardized grid middleware software libraries. HTC services therefore are described using a subset of HPC metadata, and this is the reason why we have combined both HPC and HTC services in one subcategory.

Following are service specific metadata collected. The metadata is organized in six groups for more clarity.

- Basic information
    - Peak performance [TFlops] - Theoretical peak performance of the service in TFlops, including CPUs and Accelerators
    - Server specification – Vendor specific information about servers
    - Number of servers
- CPU details
    - CPU Specification – Vendor and model of CPU
    - CPUs per server
    - Cores per CPU
    - RAM per server [GB]

    - RAM per core [GB]
    - Total number of CPU-cores – Total number of CPU cores for the entire system
    - Max number of parallel processes – maximum number of parallel processes allowed for end users
    - CPU peak performance [Tflops] - CPU theoretical peak performance of the system
- Accelerator details
    - Accelerator specification – Vendor and model of the accelerator
    - Total number of accelerators
    - Accelerators per server
    - Maximal number of accelerators per server
    - Accelerators peak performance [Tflops] - Accelerators theoretical peak perfomance of the system
- Interconnection
    - Interconnect type – Interconnection technology between servers
    - Interconnect latency [µs]
    - Interconnect bandwidth [Gbps] - Interconnection bandwidth between nodes
- Filesystem details
    - Local filesystem type – Shared Filesystem used for interconnecting nodes
    - Total storage [TB]
- Software details
    - Operating system
    - Batch system/scheduler
    - Development tools
    - Libraries
    - Applications

3.2 Cloud services

Cloud computing services that are provided by resource providers can be classified as Infrastructure as a Service (IaaS). These services enable end-users to manage their virtual resources and dynamically deploy and scale their computational requirements. Cloud computing services are usually deployed on general purpose servers, which are interconnected to a dedicated storage facility, enabling fast and scalable storage management. Therefore, following are metadata collected for cloud services, divided in 5 groups.

- Basic information
    - Server specification - Vendor specific information about servers
    - Number of servers
- CPU details

    - CPU specification - Vendor and model of CPU
    - CPUs per server
    - Cores per CPU
    - RAM per server [GB]
    - RAM per core [GB]
    - Total number of CPU-cores - Total number of CPU cores for the entire system
- Filesystem details
    - Total storage [TB]
- Virtual machine specification
    - Minimum number of CPU cores per VM
    - Minimum amount of RAM per VM [GB]
    - Maximum number of CPU cores per VM
    - Maximum amount of RAM per VM [GB]
    - Maximum amount of storage per VM [GB]
- Software details
    - VM management type – cloud software stack deployed
    - Supported interfaces – Interfaces availabe for managing the cloud (Web, API)

3.3 Storage services

Storage services represent specific cloud services that enable users to safely and securely store their data in the cloud. These services can rely on different technology and software stack, and therefore support different access interfaces and different access models. Therefore, following are metadata collected for storage services, divided in 2 groups.

- Basic Information
    - Storage specification - software stack used for the storage service
    - Total storage [TB]
    - Storage technology – Technology used for storing the data (SSD, SAS, SATA)
    - Storage performance – IOPS of
- Software details
    - Supported interfaces
    - Supported storage types

3.4 Collected RDT for NI4OS-Europe generic services

In the table below, we list the basic information for each generic service as it's extracted from the resource description templates provided by the service providers that participate

in NI4OS. Specifically, we extracted the name of each service, the tagline, the description, country and service geographical availability as it's provided by the service provider.

Table 2. NI4OS-Europe generic services

Subcategory	Name of service	Tagline	Description	Country
Cloud	AVITOHOL-CLOUD	Retrieving data. Wait for a few seconds, then try cutting or copying again	The Avitohol Cloud service allows users to launch virtual machines on servers from the Avitohol supercomputer. It allows user groups to launch long running virtual machines with substantial flexibility. It is used by diverse research groups with needs for both advanced computing and data storage.	BG
Cloud	OpenStack	OpenStack	Provide computing resources running virtualized servers on the cloud. Helps users avoid downtimes, minimizes costs for maintaining hardware and infrastructure, allow users to focus on running their applications instead on running their infrastructure, allows easy expandability and scalability.	GR
Cloud	FINKI Cloud	FINKI Openstack cloud	Openstack cloud deployed at the Faculty of Computer Science and Engineering, UKIM.	MK
Cloud	Openstack	OpenStack	Will provide resources to research teams.	GE
Cloud	RSC (RENAM Scientific Cloud)	*OpenStack*	Provide computing resources running virtualized servers on the cloud for research, computational experiments, education, training, testing IT solutions and preparing services for further use in the activities of research and educational organizations.	MD
Cloud	ICIPRO	*Microsoft AVITOHOL Azure based IaaS at ICI Bucharest*	ICIPRO offers Infrastructure as a Service services (IaaS) to Public Sector beneficiaries that need flexibility, modularity, dynamics and access to state of art technologies. Tenants can self-provision scalable Windows Server and Linux virtual machines from a gallery of predefined images.	RO
Cloud	ETFBL-CC01	*A virtual home for your machines and services*	Virtual machine and service hosting. Support for classic long-term VM use, easy backup and live VM migration.	BA
Cloud	UoM Cloud	*UoM Cloud*	Allows users to run virtual machines and host services on servers from UoM Cloud. Support for long running VMs, high availability and live VM migration.	ME
HPC	AVITOHOL	*Avitohol supercomputer*	The supercomputer Avitohol was at 331st place in the TOP 500 list of supercomputers. It is built with HP Cluster Platform SL250S GEN8 (150 servers), Intel Xeon E5-2650 v2 8C 2.6GHz CPUs (300 CPUs), non-blocking InfiniBand FDR, 300 Intel Xeon Phi 7120P co-processors. It provides 412 TFlops of performance for diverse scientific and industrial applications. Users from science and industry with substantial computational needs use it to achieve their results faster and to solve bigger problems that are beyond the reach of ordinary clusters.	BG
HPC	ARIS	*ARIS*	GRNET (National Infrastructures for Research and Technology) provides high performance computing resources to the Greek and international scientific and research communities in order to conduct scientific research.	GR
HPC	Isabella	*Isabella*	Isabella provides Croatian researchers with access to high performance compute & storage resources.	HR
HPC	Data analysis service		The PARADOX Hadoop cluster consists of a single name node that runs the YARN resource manager, and three additional data nodes. The name node is hosted on a machine with 4-core Intel Xeon E3-1220v3 CPU running at 3.1 GHz, with 4 GB of RAM, and 500 GB of local hard disk storage. Each of the data nodes, which perform the computation and storage, are hosted on machines with 24-core Intel Xeon E5-2620 CPUs at 2.4 GHz, with 64 GB of RAM and 2 TB of storage. In total, the cluster provides access to 60 CPU cores, 180 GB of RAM and 5.3 TB of storage in HDFS. In the analysis of very large datasets, the movement of data can present a far more severe bottleneck than the actual computation. Therefore, the PARADOX Hadoop cluster is designed to overlap	RS

			computation and data storage operations, i.e., to enable performing of computation on the same machine(s) that store the corresponding data.	
HPC	PARADOX-IV cluster	*Serbian supercomputing cluster*	Fourth major upgrade of PARADOX installation (PARADOX-IV) became operational during September 2013. This upgrade consists of 106 working nodes and 3 service nodes. Working nodes (HP ProLiant SL250s Gen8, 2U height) are configured with two Intel Xeon E5-2670 8-core Sandy Bridge processors, at a frequency of 2.6 GHz and 32 GB of RAM (2 GB per CPU-core). The total number of new processor-cores in the cluster is 1696. Each working node contains an additional GP-GPU card (NVIDIA Tesla M2090) with 6 GB of RAM. With a total of 106 NVIDIA Tesla M2090 graphics cards, PARADOX is a premier computer resource in the wider region, which provides access to a large production GPU cluster and new technology. The peak computing power of PARADOX is 105 TFlops. Other technical information is provided in the table below.	RS
Storage	AVITOHOL-STORAGE	*Avitohol storage*	Provides shared storage for users of Avitohol supercomputer; reliable and high performance storage; available to all users of supercomputer Avitohol.	BG
Storage	Data discovery service	*Data discovery service*	The service provides for flexible searching for data discovery. This is a powerful dataset management system that provides publishing, sharing, searching and can use almost any type of data and metadata. It provides users with ability to publish metadata about their data and to find relevant datasets based on metadata information. The service is open and multidisciplinary.	BG
Storage	Archival service	*Archival service*	Data archiving is the practice of moving data that is no longer being used or are being used on a less frequent fashion into a separate storage device. It is a single set or a collection of historical records specifically selected for long term retention and future reference. The Archival service is based on the tape-based tertiary storage of the ARIS HPC system.	GR
Storage	RenamStor	*Servers for VMs backup and archive information*	Servers for VMs backup and archive information. Service based on Dell 740 servers with FreeNAS software	MD
Storage	PARADOX storage system		PARADOX provides a data storage system, which consists of two service nodes (HP DL380p Gen8) and 5 additional disk enclosures. One disk enclosure is configured with 12 SAS drives of 300 GB (3.6 TB in total), while the other four disk enclosures are configured each with 12 SATA drives of 2 TB (96 TB in total), so that the cluster provides around 100 TB of storage space. Storage space is distributed via a Lustre high performance parallel file system that uses Infiniband technology, and is available both on working and service nodes. Also, the storage can be used externally via gridFTP protocol.	RS
Storage	Simple storage service	*Data exchange service for the scientific communities*	The Simple Storage service allows research community members to keep and sync research data on various devices, as well as to share their data, thus making it a useful service in collaborative environment. The service allows versioning of all ingested files, which makes it useful for active data (data under development, frequently changed) management. Access is enabled via web browsers, desktop and mobile clients. The service is based on ownCloud platform.	RS

Based on this, in Figure 8 and Figure 9 we can see the distribution of Generic services by type and country.

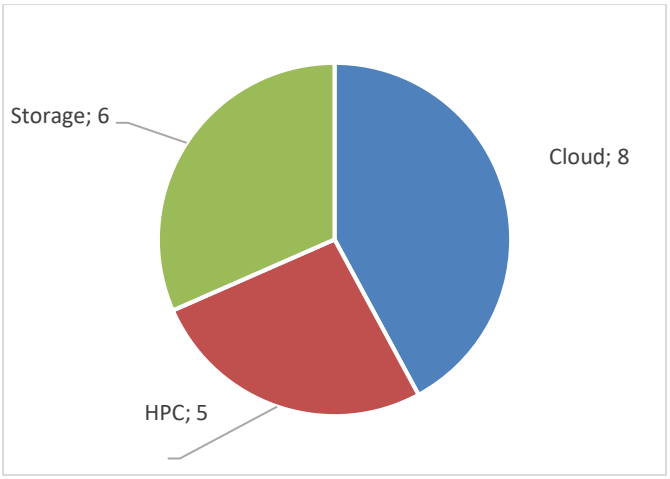


Figure 8. NI4OS-Europe generic services distribution by category

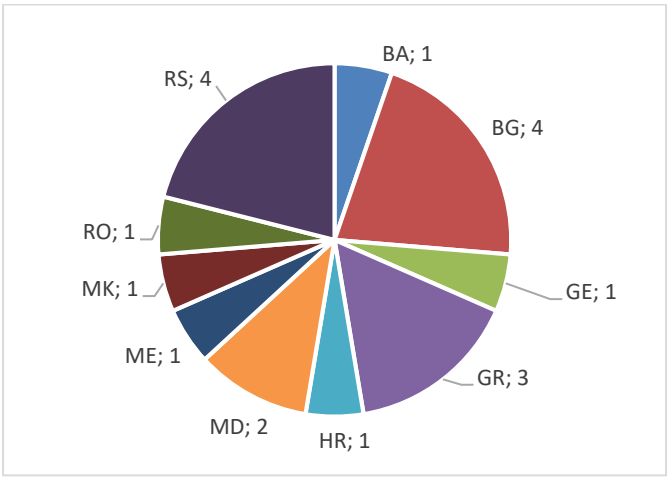


Figure 9. NI4OS-Europe generic services distribution by country

3.4.1 Quick-win generic services

The NI4OS-Europe project selected a subset of 6 resources, 2 generic, 2 thematic services and 2 repositories to be on-boarded first to EOSC. These so called quick-win services, chosen based on their maturity, will be used to validate the pre-production environment as well as the on-boarding procedures.

From the generic services, the quick-win services are the Data analysis service provided by the IPB and the FINKI cloud service provided by UKIM.

The Data analysis service is a Hadoop cluster-based service, hosted on the PARADOX cluster at the IPB. The cluster consists of a single name node that runs the YARN resource manager, and three additional data nodes. To avoid the possible bottleneck that can occur due to the movement of data, the cluster is designed to overlap computation and data storage operations, i.e., to enable performing of computation on the same machine(s) that store the corresponding data. The service is in TRL8 and can be used by the research communities immediately. In the later phases, it will also be integrated with additional

elements from the NI4OS-Europe pre-production environment, such as AAI, monitoring and accounting.

FINKI Cloud is an OpenStack cloud platform hosted by the FCSE, UKIM. It is a production service, with TRL9, already used by the research and education communities in the country. The service is already integrated with the NI4OS-Europe AAI and planned to integrate additional elements such as monitoring and accounting very soon. The 12 servers allocated for this purpose offer flexible VM management, their easy provisioning and deployment.

4 Categorization of thematic services

After performing a careful study and analysis of the thematic services proposed by a number of researchers coming from research institutions all across the region as well as taking into account knowledge based on services which have been embedded in previous e-infrastructure projects, we created a template with a list of metadata which could provide important information regarding the implementation of the services. These data can enable the categorization of the services according to their details.

Since the thematic services cover the long tail of science, it is expected to be inhomogeneous in the way they are implemented and applied. Thus, we need to be careful and include as much information as possible in order to have a clear understanding what the needs are by taking into account this diversity.

The thematic services, can be categorized in the way they are implemented, namely based on their computational needs. For instance some services could have the form of platforms providing tools running on virtual machines while some other services could include a set of software solutions which require the use of High Performance Computing to be executed.

The additional information collected is listed below.

4.1 Basic information

- Resource name – The service provider is requested to provide the name of the service as this is entered in the resource description.

4.2 Computational/storage requirements

- Computational requirements – This entry requires to provide the computational requirements of the Service.
- Requires VM(s) – Here the provider has to specify whether the service require VMs from generic service providers? The answer could be either Yes or No.
- Number of VM(s) – The reader should provide the number and type of Virtual Machines the service requires. The answer should specify the number of CPUs as well as their Physical Memory.
- Storage requirements [GB] – Unlike the physical memory which no matter how large it is, it cannot be a serious obstacle for the on-boarding of a service on hardware, what could consist a limitation is the storage memory. Thus, in this entry one should provide how much data storage the service requires. The answer should specify capacity and access protocol.
- Require HPC – The provider should specify whether the service requires the use of High Performance Computing.
- HPC resources – The provider should describe the amount of HPC resources and time schedule for execution.

- Require GPUs as part of HPC – In this entry the thematic service provider should explain whether the service requires the use of GPUs as accelerators as part of the High Performance Computing resources.
- Amount of GPUs – The Provider should insert the number of GPU cards necessary for the execution of the service.
- Require VMs with GPU acceleration – From our experience in previous e-Infrastructure projects, services running on VMs, required computational acceleration via GPU cards. Such applications were based on deep learning algorithms. Therefore, in this entry we would like to know if the service requires the use of VMs with GPU acceleration.
- Amount of VMs with GPUs – Here the provider should specify the amount of VMs with GPUs.

4.3 Further Information

- Deployed in an existing infrastructure – In this entry the service provider should explain if the service has been used in an existing Infrastructure. For instance some of the services which have been identified as services to be on-boarder in NI4OS-Europe in the proposal are already present in some other infrastructures such as VI-SEEM, ARIADNE PLUS, and others.
- Workflows available – Here the provider should explain if there are existing workflows available for this service.
- Pipeline available - Here the provider should explain if there are existing pipelines available for the service?
- Uses data stored in repository – In this entry the provider should explain if the service uses data stored in repository.
- Amount of produced data per year – Here the entry should be estimating the amount of data produced in a year.
- Metadata schema of the data – The thematic service provider should specify if the service is using a metadata schema and if yes, what this schema is.
- Ontologies of the data - The thematic service provider should specify if the service is using an Ontological solution and if yes, what this solution is.
- GDPR compliance – The provider should clarify whether the data follows the EU GDPR principles.

4.4 Software details

- License of the software – The service provider should identify what license would be used for legally binding guidelines for the use and distribution of software.
- Repository – The thematic service provider should provide the URL of the repository used for reading as well as for storing data.

4.5 Collected RDT for NI4OS-Europe thematic services

In the table below, we list the basic information for each thematic service as it's extracted from the resource description templates provided by the researchers who participate in NI4OS. Specifically, we extracted the name of each service, the TRL level, the tagline, the description, and the scientific domain as it's provided by the researchers/creators of the service.

Table 3. NI4OS-Europe thematic services

Name of service	TRL	Tagline	Description	Scientific Domain
Skeletal_Sex	1	Skeletal_Sex uses machine learning techniques to estimate the sex of unknown individuals using cranial and postcranial measurements.	This service estimates the sex (male, female) from skeletal remains. The user inputs standard cranial and postcranial measurements and Skeletal_Sex produces a sex estimation with an associated probability of an individual being male or female. The training sample is based on the worldwide William W. Howells Craniometric Data Set and the Goldman Osteometric Data Set, for cranial and postcranial measurements respectively. The aim of Skeletal_Sex is to offer an open access resource to forensic anthropologists and, secondarily, to bioarchaeologists.	Archaeological Science
Archaeological Simulation and Urban Modelling (ArcheoSUM)	2	Archaeological Simulation and Urban Modelling for the study of historical evolution of complex systems and social phenomena.	Data processing tools for computational modelling and simulation of historical evolution of complex systems and social phenomena. Supporting a Virtual Lab for testing multiple hypotheses regarding dynamic phenomena and conditions in time and space, such as urban modelling and the growth of settlements; historic city network interactions; and system dynamics modelling, with the use of agent-based modelling, parametric design, GIS, visualization and simulation of built environment-human interactions.	Urban Modeling and simulation, computational archaeology, Cultural Heritage and Built Environment
ClInHealth	2	ClInHealth produces comprehensive data about indexes and metrics that quantify the impact of atmosphere parameters and characteristics on the quality of life and health risks for the population.	The service will generate reliable, comprehensive and detailed studies of the impact of lower atmosphere parameters and characteristics on the quality of life and health risks for the population in our country. It uses a synergetic application of extensive computer simulations on the supercomputer Avitohol, combined with sophisticate analysis of the parameters and characteristics of near surface atmosphere. In this way the impact of the atmosphere on the human health and quality of life can be thoroughly investigated. The users are cross-disciplinary, from the domains of atmospheric physics and environmental science, allergology and epidemiology, as well as national and municipal policymakers.	Geophysics
Mapping Archaeological and Paleoclimate information (MAPi)	2	Mapping Archaeological and Paleoclimate information to facilitate interdisciplinary research between archaeology, geography and climate.	Mapping Archaeological and Paleoclimate information will enable researchers to combine, link and visualize datasets that are typically accessible only from within the confines of distinct disciplines, such as archaeology, geography and climate. It will provide an online resource that will facilitate interdisciplinary research, crossing the boundaries of these fields for the interpretation of past human-environment interactions by means of studying data of environmental factors together with cultural heritage digital assets, archaeological knowledge and remote sensing data, e.g., dynamic maps of geolocated archaeological finds, and their linked metadata (provenance), with climate visualizations.	Computational archaeology, Cultural Heritage and Built Environment

ClimCoSt	3	Providing comprehensive and reliable computer simulations of climate changes in regional/local scales and evaluation of their impacts on ecosystems and quality of life.	The service will produce reliable, comprehensive and detailed evaluations of possible regional/local climate changes and their consequences for different global change scenarios. Metrics and tools for evaluating some of the climate change impacts on environment and quality of will be available. Making use of vast computing resources, the service will enable scientist to perform in-depth assessment of the climate change impacts that cannot be achieved with the desired accuracy using local computing resources. Users are climate scientists and national and municipal policymakers.	Geophysics, climate
Effect-based toxicological assessment	3	How can I compare my data to others?	Service which computes Inhibition percentage that can be used to compare results among existing results in the database.	Environmental Toxicology
Toxicological evaluation	3	How toxic my sample is?	Service which computes effective concentrations or Toxic Units that can be used to compare results among existing results in the database.	Environmental Toxicology
EEGHUB.GE	4	EEG recordings collection.	The service contains time series of recorded brain electrical activity of healthy subject and subject with different disorder of Central Nervous system, with various gender and age group. Service has convenient search engine, which allows to find any recording which corresponds to specific requirements. The recordings are easily accessible and can be quickly downloaded for further exploration/exploitation. Possible customers are open-source groups of researchers/practitioners, lecturer/students, scientific institutions, Hospitals, Universities.	Neuroscience, psychophysiology, clinical neuroscience, medicine, psychology, neurophysiology, Cognitive and Social Science
Cyber Security Incidents Analytics	4	Forewarned is forearmed.	Cyber Security Incidents Analytics project, that in fact will be implemented as cyber security platform, should collect all security incidents data of a network, process it, enhance the data and make it available for analyze using one big data mechanism. Main Idea of this platform is to give access to the related incidents data analysis for large number of users based on their permissions, that means that both small ISP or organizations will have access to their incidents as well as governmental or international CERT's.	Security, Computer Science
IoT Cloud Platform	4	IoT platform for data collecting, processing and visualization.	IoT Cloud Platform will enable users to collect data from internet-connected hardware, and visualize them in near real-time. Collected data will be able to be analyzed online in Octave or R programming language (discovering relationships, patterns and trends in data), whereby users will be able to use a pre-written algorithm or to develop a new code directly in the web browser. IoT platform will support a variety of hardware devices, such as Arduino and Raspberry Pi platforms, mobile devices, PCs, etc. Beside web interface, API will be provided, so that data can be uploaded, stored, and accessed from the third party.	Interdisciplinary, Computer Science
AstroMatch	4		The AstroMatch service provides astrometric calibration of astronomical images taken with any telescope which contain celestial coordinates of a central (or any other) pixel and an approximate pixel scale. The service uses the Digitalized Sky Survey (http://archive.stsci.edu/cgi-bin/dss_form) to access archived astronomical images from the First and the Second Palomar Sky Survey and the Sloan Digital Sky Survey (https://www.sdss.org/dr14/imaging/imaging_access/) when possible. Astronomical images are matched to archived images of the same field of view through a set of translations and rotations until the objects overlap, which provides an initial astrometric solution. Next, all the objects from the input (uncalibrated) images are selected and searched for in many existing star catalogs to refine their celestial coordinates. Finally, the improved astrometric calibration is calculated and the WCS (World Coordinate System) solution is provided, and may optionally be written to the header of the images.	Astronomy

EML	4	Environmental machine learning.	The Explainable Machine Learning (EML) is a service for highly sophisticated machine learning (ML) and interpretation frameworks deployment aimed at delivering personalized and explainable predictive analytics solution to a research or commercial customer. Understanding and correctly interpreting models for predicting natural and social phenomena, parameterized with a large number of hyperparameters, such as random forests, deep neural networks, or an extreme gradient boosting, can be challenging. The EML is designed to implement the most accurate ML methods and to produce a posteriori explanation using consistent, locally accurate, individualized feature-attribution methods, thus shedding light on problems where human intuition and domain knowledge are often limited.	Environmental science
RS2C	4	Remote Sensing Scene Classification Service.	A web-based service for classification of remote sensing scenes into land use/land cover classes. The service is based on convolutional neural networks trained on publicly available datasets of high-resolution remote sensing images and implemented using TensorFlow, TensorFlow Serving, and Docker. The service enables faster analysis of large quantities of remote sensing images available from various sensors ranging from satellites to UAVs. These images are used for land cover and land use classification, monitoring urban growth, monitoring and forecasting climate changes, to name a few. Recently, using remote sensing for monitoring of ecosystems, insects and animals also gains in significance. The system in development aims to assist in making sense of the data acquired in these application domains. Main user communities envisaged to benefit from this service are in the areas of agriculture, food production, urban planning, and environment protection. However, we expect to identify new users' groups both during the project, as well as after it is finished.	Engineering
Open Mapping Application (OMApp)	5	Cloud application for automatic image mosaicking and georeferencing.	OMApp is cloud application for automatic image mosaicking and georeferencing. The application is designed to support several users, whereby every user is able to upload a set of captured images via a web interface, begin their processing and make an overview of already created maps. OMApp uses numerous open source image processing tools and libraries, where the most computationally demanding among them are able to perform multi-core parallel processing, which provides a better usage of the cloud resources.	Interdisciplinary, Computer Science
Analysis of Microscopy data for cells and embryos	5	MicroProc	A method that detects mitotic events as well as the orientation of the mitotic axis. Also follow the development of embryos.	Software for biological image analysis
Shrodinger	7	Service for advanced methods for solving of multidimensional stationary and time-dependent Schrödinger equation.	In many subdisciplines of computational molecular sciences, computational physics, chemistry, biology, materials science, exact treatment and analysis of a wide variety of phenomena has to rely on rigorous quantum description of the underlying processes. This, on the other hand, requires solving of either the stationary (time-independent) or the time-dependent Schrödinger equation. The effort required to complete such a task is heavily dependent on the dimensionality and complexity of the problem itself (e.g. the exact form of the Hamiltonian, number of the relevant degrees of freedom of the studied system etc.). Numerous methods have been proposed in the literature to achieve the mentioned aim. However, the available codes are most often user-hostile, the procedures for computation and generation of relevant data are non-standardized, and there is a clear lack of in-depth, thorough comparison of performances of various methods for solving the Schrödinger equation for various purposes. The proposed service will provide user-friendly (as much as possible) computational platforms for solution of stationary and time-dependent Schrödinger equation, implementing several algorithms. The main intention would be to rely on series of datasets generated by quantum mechanical computations, while the actually implemented algorithms will base on either sequential interpolation – variational calculation or on some variant of discrete variable representation technique (DVR), Fourier-grid Hamiltonian approach etc.	Technical and Natural Science

SciRoHub	7	Mirror of Copernicus Open Access Hub restricted to the Romanian territory.	The Service provides Copernicus's Sentinel1 and Sentinel2 products that cover the Romanian territory. Sentinel1 mission provides all-weather, day-and-night radar imaging for land and sea. Sentinel2 provides high resolution optical image data, including monitoring of vegetation, soil and water cover, as well as observation of inland waterways and coastal areas. It targets researchers from Romanian National Research Institutes and Universities that are interested in using satellite images for their research purposes. The Service is free to use for Romanian academic and research community. Registration is mandatory in order to access the data.	Earth Observation
CPMot	7		Density functional theory (DFT) is currently the most useful method for calculation of electronic structure of large molecules and materials. The basic quantity within DFT is electronic charge density. In typical DFT calculations one has to self-consistently solve the Kohn-Sham equations for single particle wave functions and the equation that expresses electronic charge density in terms of wave functions of occupied electronic states. Charge patching method (CPM) is the method based on DFT which avoids self-consistent calculations of electronic charge density. It is based on the idea that electronic charge density in the neighborhood of a certain atom depends mainly on its local environment and that it is the same in a large system that one wants to calculate and in some small system where that atom has the same local environment. These contributions are called charge density motifs. These motifs are essential input for electronic structure calculations using CPM. The CPMot service provides motifs for a variety of semiconducting and insulating materials and nanostructures.	Computational Physics
GQL	7		Guanine(G)-quadruplex is considered relevant for various cellular processes. Recent research indicates that it represents an active site in antitumor treatment, since the formation of a G-quadruplex-ligand complex is expected to prevent further cell division. GQL service identifies ligands that efficiently bind to G-quadruplex. The main advantage of GQL service is that it is based on analysis of recently (in 2017 and 2018) synthesized ligands (squaraine-based compounds, benzimidazole-carbazole molecules, and complexes with transition metals) which have not been thoroughly investigated. The service monitors interaction between a drug (ligand) and selected G-quadruplex structures from the PDB database using intensive molecular dynamics simulations. As a result, GQL produces datasets with the equilibrated molecular structures, which provide additional information to the available experimental NMR and X-ray diffraction structures, which are used as initial structures for further studies.	Computational chemistry
LMDB	7		LMDB is a distributed database of numerical results and codes pertaining to a large class of condensed matter theoretical models, namely the correlated lattice models. Correlated lattice models are used for the study of crystalline materials exhibiting high-temperature superconductivity, antiferromagnetic, Mott insulating states and various other emergent phenomena. These models abstract the details of the ionic lattice, but take account of the Coulomb interaction between the electrons. Therefore, they pose a quantum many-body problem which is intractable in general, but can be solved using a variety of numerical methods, to different levels of approximation. Typical many-body calculations take thousands of CPU hours, so sharing existing results is essential for the efficiency and making further progress in the field. LMDB provides a natural and high-level systematization of the models and methods which allows people to effortlessly share and view other people's results and reuse codes. Even more importantly, the service allows for automatized analyses of large datasets which is unprecedented in the field. The service consists of a central server which tracks numerical data on the data servers, hosted on HPC facilities where the data is produced and stored. The central server provides a web API for the communication with the database. Additionally, the central server provides a flexible web-based GUI for browsing and plotting results. A typical small-scale usage scenario involves no programming, yet allows the user to combine results from multiple remote sources into new analyses.	Computational physics

EpHEMERA	8	Endangered architectural and archaeological Heritage in the south Eastern MEediterRAnea area.	The World Heritage Convention, drawn by various international bodies in 1972, was designed to protect cultural or natural places of outstanding universal value so that future generations may be able to enjoy them. Responding to these principles as well as to the Charter on the Preservation of Digital heritage (Vancouver, 2003), this multidisciplinary project, which involves archaeologists, art historians, conservators and computer scientists, aims to create an open access, 3D interactive online geo-database of endangered architectural and archaeological heritage in the South Eastern Mediterranean basin; a region of tremendous cultural importance whose rich heritage is unfortunately threatened by both natural and human factors. A wide range of 3D modelling and topographic techniques have been applied to create accurate reconstructions of heritage sites, enriched by an extensive array of metadata. The Online 3D Database System for Endangered architectural and archaeological Heritage in the south Eastern MEediterRAnea area (EpHEMERA) is intended to serve as an infrastructure where it is possible to: Visualize online and through standard web browser 3D architectural and archaeological models classified according to a specific type of risk; Query the database system and retrieve metadata attached to each single virtual object; Extract geometric and morphological information.	Digital Cultural Heritage, Heritage, Environment, Energy, Super computing
AFMM	8	AFMM (Automated FrequencyMatchin gMethod) is a program package to aid in molecular mechanics force field parametrization for small organic molecules.	AFMM (Automated FrequencyMatchingMethod) is a program package for molecular mechanics force field parametrization. The method used fits the molecular mechanics potential function to both vibrational frequencies and eigenvector projections derived from quantum chemical calculations. The program optimizes an initial parameter set (either pre-existing or using chemically-reasonable estimation) by iteratively changing them until the optimal fit with the reference set is obtained. By implementing a Monte Carlo-like algorithm to vary the parameters, the tedious task of manual parametrization is replaced by an efficient automated procedure. The program is best suited for optimization of small rigid molecules in a well-defined energy minimum, for which the harmonic approximation to the energy surface is appropriate for describing the intra-molecular degrees of freedom. It can be used for small organic molecules.	Chemistry, Biology, Biochemistry, Biophysics, Medicine, Life Sciences
FEPrepare	8	FEP prepare is a webserver which automates the set-up procedure for performing NAMD/FEP simulations.	FEP prepare is a webserver, which automates the set-up procedure for performing NAMD/FEP simulations. Automating free energy perturbation calculations is a step forward to delivering high throughput calculations for accurate predictions of relative binding affinities before a compound is synthesized, and consequently save enormous time and cost.	Chemistry, Biology, Biochemistry, Biophysics, Medicine, Life Sciences
Subtract	8	Subtract accurately calculates the volume of protein binding sites, and works both for crystal structures downloaded from the Protein Data Bank and for protein structures arising from Molecular Dynamics simulations trajectories.	Subtract accurately calculates the volume of protein binding sites, and works both for crystal structures downloaded from the Protein Data Bank and for protein structures arising from Molecular Dynamics simulations trajectories. Subtract accepts an atom selection in the form of a PDB file and computes the three- dimensional convex hull of the atoms points with the help of SciPy library. The next step of the algorithm is to compute the volume of the convex hull and the volume of the atoms that are included in the solid based on their van der Waals radii. The subtraction of those two volumes yields the volume of the investigated cavity. The algorithm computes cavity volumes of trajectory frames in parallel for maximum efficiency and speed. It requires minimal usage of memory due to the fact that it follows a buffering strategy of reading file chunks and therefore there is no need to load the entire file into memory. There is a wide support of trajectory formats like Gromacs trajectory files and multi-model PDB files due to its dependency to the MDTraj library.	Chemistry, Biology, Biochemistry, Biophysics, Medicine, Life Sciences
BioConnect	9	Data discovery solution for omics data.	Bio-connect will offer a comprehensive data discovery solution intended to provide a general-purpose, web-based tool that can be used by any omics data owner. Bio-connect will allow omics data to be appropriately discoverable in a manner that data owners maintain control of the data, where they can set multiple levels of access. Furthermore, Bio-connect provides a set of tools used to assist with omics data validation, management, viewing and sharing of complex omics data.	Bioinformatics, Biology

Image Reconstruct	9	Image Reconstruct of medical images.	Image Reconstruct will offer a comprehensive solution intended to reconstruct medical images from Single-photon emission computed tomography (SPECT), Positron emission tomography (PET) or X-ray computed tomography (X-ray CT), using multiple methodologies, including but not limited to filtered back-projection (FBP), algebraic reconstruction techniques (ART) and maximum likelihood expectation maximization (MLEM). Furthermore, Bio-connect provides a set of tools used to assist with data validation, management, viewing and sharing of reconstructed data as images.	Medical Imaging, health
Cyprus Weather Data	9	Weather model forecast products and real-time observations from Cyprus.	Service which provides model forecast products and real-time observations from Cyprus. This service is of the interest of European researchers in climatology, interested in the Eastern Mediterranean climate.	Earth Sciences, physics
ChemBioServer	9	ChemBioServer is a publicly available web-application for effectively mining and filtering chemical compounds used in drug discovery.	ChemBioServer is a web-server for filtering, clustering and networking of chemical compound libraries facilitating both drug discovery and repurposing. It provides researchers the ability to (i) browse and visualize compounds along with their physicochemical and toxicity properties, (ii) perform property-based filtering of chemical compounds, (iii) explore compound libraries for lead optimization based on perfect match substructure search, (iv) re-rank virtual screening results to achieve selectivity for a protein of interest against different protein members of the same family, selecting only those compounds that score high for the protein of interest, (v) perform clustering among the compounds based on their physicochemical properties providing representative compounds for each cluster, (vi) construct and visualize a structural similarity network of compounds providing a set of network analysis metrics, (vii) combine a given set of compounds with a reference set of compounds into a single structural similarity network providing the opportunity to infer drug repurposing due to transitivity, (viii) remove compounds from a network based on their similarity with unwanted substances (e.g. failed drugs) and (ix) build custom compound mining pipelines.	Chemistry, Biology, Biochemistry, Biophysics, Medicine, Life Sciences
NanoCrystal	9	NanoCrystal is a novel web-based crystallographic tool that creates nanoparticle coordinates from any material crystal structure.	NanoCrystal is a novel web-based crystallographic tool that creates nanoparticle models from any crystal structure guided by their preferred equilibrium shape under standard conditions according to the Wulff morphology (crystal habit). Users can upload a cif file, define the Miller indices and their corresponding minimum surface energies according to the Wulff construction of a particular crystal, and specify the size of the nanocrystal. As a result, the nanoparticle is constructed and visualized, and the coordinates of the atoms are output to the user.	Chemistry, Biology, Biochemistry, Biophysics, Medicine, Life Sciences
ProTraits	9	Prokariotyc traits atlas.	ProTraits atlas is a resource containing ~545 000 novel phenotype inferences, spanning 424 traits assigned to 3046 bacterial and archaeal species. These annotations were assigned by a computational pipeline that associates microbes with phenotypes by text-mining the scientific literature and the broader World Wide Web, while also being able to define novel concepts from unstructured text.	Bioinformatics
REVIGO	9	Reduction and visualization of Gene Ontology.	REVIGO is a Web server that summarizes long, unintelligible lists of GO terms by finding a representative subset of the terms using a simple clustering algorithm that relies on semantic similarity measures. Furthermore, REVIGO visualizes this non-redundant GO term set in multiple ways to assist in interpretation: multidimensional scaling and graph-based visualizations accurately render the subdivisions and the semantic relationships in the data, while treemaps and tag clouds are also offered as alternative views.	Bioinformatics
OpenBioMaps	9	Open and free biological database service.	OpenBioMaps is a web-based, open-access database framework project which maintained by the OpenBioMaps Consortium and the databases involved at least partially open-access or contains free-content. The OpenBioMaps provides an open-access web application which designed to create and use open-content biological databases, specifically for scientists and conservationists, and its customizable toolset allows for the easy access and management of data.	biology

Dicom Network	9	Solution for Medical Images Archive, Access, Processing and Visualization.	"DICOM Network" service provide full set of functionalities for data collect, storage, distribution and exchange DICOM medical investigations. It's implements all the standard PACS interfaces as well as integrated security features. Various imagistic investigations like tomography, Roentgen, ultrasound, angiography, etc. Familiarization and working experience accumulation by medical specialists in using such systems offer obvious advantages in imagistic investigations and forming treatment decisions, allow supporting collaborative work and appealing for support from the best local and foreign specialists who have extensive experience in the field.	Medicine, Medical information system
Airpolution prediction	9	Predicting airpolution levels for the next 4 days.	Simulation system for generation of prediction of airpollution levels based on WRF-Chem software. The outputs are hourly levels of airpollution for PM10, PM2.5, NO and SO2	Natural Sciences
DREAM	9		The DREAM service simulates and predicts the atmospheric cycle of mineral dust aerosols. It is tuned for usage on high-performance computing infrastructures available today. A typical use-case is the production of a dataset with the aerosol optical thickness and surface dust concentration for a particular period and for a particular geographical region. The service supports different horizontal and vertical resolutions. The results produced by the DREAM service have been applied, using the human health impact function and calculated global fine particulate matter concentrations, for estimation of the premature mortality caused by the long-term exposure to airborne desert dust. The results have high sensitivity on the threshold concentration, which is a significant parameter of relevance to public health.	Climate Modelling
VideoLectures.NET	9	Exchange ideas & share knowledge.	VideoLectures.NET is an award-winning free and open access educational video lectures repository. The lectures are given by distinguished scholars and scientists at the most important and prominent events like conferences, summer schools, workshops and science promotional events from many fields of Science. The portal is aimed at promoting science, exchanging ideas and fostering knowledge sharing by providing high quality didactic contents not only to the scientific community but also to the general public. All lectures, accompanying documents, information and links are systematically selected and classified through the editorial process taking into account also users' comments.	multidisciplinary
Online Virtual Reality Environments Toolkit (OVRET)	3-4	Online Virtual Reality Environments Toolkit for the creation of immersive environments for virtual museums, interactive collections of cultural artefacts.	Users will be able to use the provided software tools, workflows and plugins (scripts of code) to create immersive environments for virtual museums, interactive collections of cultural artefacts, visual interfaces for geolocated interaction with DCH assets in physical space through mobile devices, and virtual visits of inaccessible, or demolished, heritage, monuments and historic sites.	Cultural Heritage
Clowder4DCH	5-6	A flexible and extensible online content management system for Digital Cultural Heritage.	Clowder4DCH, a highly extensible active curation-based research data management platform. It contains three major extension points: preprocessing, processing and previewing. When new data is added to the system, preprocessing is off-loaded to extraction services for extracting appropriate data and metadata. The extraction services attempt to extract information and run preprocessing steps based on the type of the data, e.g., to create previews. This raw metadata is presented to the user via a web interface. Users can upload, download, search, visualize research datasets and explore information linked to data. Users can link and organize datasets in online collections following the provided workflows for creating semantically structured data repositories specialized for digital cultural heritage. It enables users to form an online collaboration environment to support research communities and activities, and disseminate results.	Digital Humanities, cultural heritage and Built Environment

Gaussian	7	Service for integrating Gaussian process regression and other machine learning techniques with molecular dynamics simulations	Gaussian regression, along with other emerging machine learning techniques, has become more and more popular in computational chemistry, physics, biology and life sciences. In conjunction with the molecular dynamics simulations, these approaches have been shown to be rather useful for prediction of a wide variety of molecular and materials' properties and functionalities. However, due to the novelty of techniques, the procedures for their application as well as their validation are far from being standardized. Therefore, the proposed service would provide a user-friendly (as much as possible) environment for development and application of the Gaussian regression technique, along with certain other machine learning techniques, to predict physiological activity, various molecular and materials' properties, phase diagrams of complex materials, and also to compute their basic structural and spectroscopic properties, primarily on the basis of the results from molecular dynamics simulations. We strongly believe that such service could provide a useful platform that could be also used for standardizing the most contemporary machine learning techniques accounting for specificities in various areas.	Computational chemistry, computational physics, computational biology and life sciences, computational materials science.

The TRL 8-9 services fall into the broad categories of digital cultural heritage, life sciences, and climate science as shown in the Figure 10 below, while one service is multidisciplinary in nature.

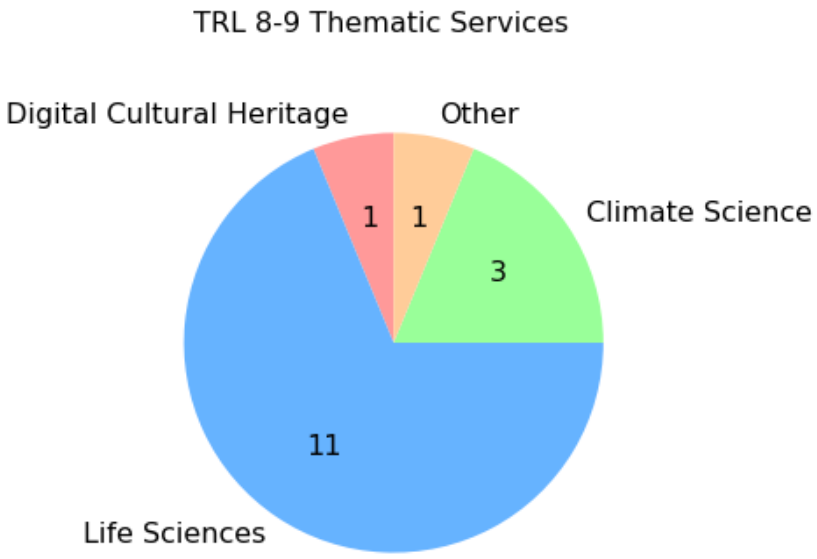


Figure 10. NI4OS-Europe thematic services categories distribution

5 Categorization of repositories

Repositories are online digital archives that collect, preserve and disseminate scientific intellectual output in the form of digital assets, such as publications, data and software. Repositories can be interoperable using protocols to which repositories should conform, such as OAI-PMH, which allows search engines and open access aggregators to index repository metadata and content and provide value-added services on top of this content. The interoperability feature and the open access to publications and data makes repositories one of the first and most mature characteristic of Open Science, driven by the Open Access movement.

Repositories that hold and preserve scientific literature and data will, unlike the generic and thematic services, be on-boarded using the OpenAIRE platform. The OpenAIRE Provider Dashboard acts as the gateway to the European Open Science Cloud for repository content providers. Therefore, we need to ensure the repositories achieve compliancy with the OpenAIRE Guidelines [5] which help repository managers expose the digital assets via the OAI-PMH protocol in order to integrate with OpenAIRE infrastructure, register the services and achieve validation using the OpenAIRE Validator Service.

Repositories are classified in 3 subcategories:

- Publication repositories
- Dataset repositories
- Software repositories

The same kind of metadata was collected for the description of both, publication and dataset repositories.

5.1 Publication repositories

Publication repositories enable the collection, preservation and dissemination of scientific literature. These repositories can rely on different software and may use different metadata schemas which may or may not be compatible with the Dublin Core or DataCite Metadata Schemas which OpenAIRE has adopted as the basis for harvesting and importing metadata about publications from repositories. Crucial information about the repositories are also OAI-PMH v2.0 implementation for harvesting dataset metadata and information about repositories registered in the OpenDOAR registry.

The data collected about the repositories focus on the FAIR principles (findability, accessibility, interoperability and reusability) compliance and the technical aspects that are crucial for OpenAIRE infrastructure compliance (underlying software, metadata schemas and the implementation of the OAI-PMH protocol). Following is the metadata collected for publication repositories:

- Basic Information
    - Type of content
    - Software and version
- FAIRness of the repository
    - Access method
    - Is the repository indexed in a directory/registry?

        - Is the repository registered or indexed in a searchable resource?
        - Are objects assigned globally unique and persistent identifiers?
        - What metadata standard does the repository use?
        - Are objects released with a clear and accessible data usage license?
- OpenAIRE compliancy
    - Does the repository support an OpenAIRE-compliant metadata schema? – Dublin Core
    - Does the repository use the OAI-PMH protocol for metadata exchange?
- Certification and policies
    - Does the repository hold a CoreTrustSeal or other repository trustworthiness certificates?
    - What policies does the repository have in place?

5.2 Datasets repositories

Dataset repositories enable the collection, preservation and dissemination of datasets. These repositories can rely on different software and may use different metadata schemas which may or may not be compatible with the Dublin Core or DataCite Metadata Schema which OpenAIRE has adopted as the basis for harvesting and importing metadata about datasets from data archives. Crucial information about the dataset repositories are also OAI-PMH v2.0 implementation for harvesting dataset metadata and information about research data repositories registered in the OpenDOAR re3data registry. The data collected about the repositories also focus on the FAIR principles (findability, accessibility, interoperability and reusability) compliance. Following is the metadata collected for data repositories:

- Basic Information
    - Type of content
    - Software and version
- FAIRness of the repository
    - Access method
    - Is the repository indexed in a directory/registry?
    - Is the repository registered or indexed in a searchable resource?
    - Are objects assigned globally unique and persistent identifiers?
    - What metadata standard does the repository use?
    - Are objects released with a clear and accessible data usage license?
- OpenAIRE compliancy
    - Does the repository support an OpenAIRE-compliant metadata schema? – Dublin Core/DataCite
    - Does the repository use the OAI-PMH protocol for metadata exchange?
- Certification and policies
    - Does the repository hold a CoreTrustSeal or other repository trustworthiness certificates?
    - What policies does the repository have in place?

5.3 Collected RDT for NI4OS-Europe repositories

Following, in the table below the basic information from the resource description templates about each of the repositories is gathered, as provided by the national NI4OS-Europe on-boarding support team representatives. Namely, we extracted the name and country of origin of each repository, the description, the scientific discipline, type of content, and whether it is listed in relevant repository registries.

Table 4. NI4OS-Europe repositories

Name of repository	Description	Discipline	Type of content	Registered in OpenDOAR / re3data
muzej.info, Bosnia-Herzegovina	A home for digital cultural heritage collections	Ethnology and Anthropology; Digital Cultural Heritage	Data; Digital Library	No
OpenScience-BG, Bulgaria	Indexing open access articles, reports, monitoring, usage statistics	Multidisciplinary	Publications	No
Institute of Public Finance Repository, Croatia	Repository provides access to publications and research data produced by the employees of the institute, papers published in scientific journals, conference proceedings, dissertations, books, manuals, guides and complete documentation related to the Institute's activities. Content of repository is open to all users for searching and downloading with clearly stated usage rights.	Social Sciences	Publications; Data	OpenDOAR
Meteorological and Hydrological Service of Croatia Repository, Croatia	Repository provides access to publications and research data produced by the employees of the institute. Content of repository is open to all users for searching and downloading with clearly stated usage rights.	Natural Sciences	Publications; Data	No
Repository of Faculty of Science, Croatia	Repository provides access to publications and research data produced by the employees and the students of the faculty. Content of repository is open to all users for searching and downloading with clearly stated usage rights.	Natural Sciences	Publications; Data	OpenDOAR
Repository of the Faculty of Economics in Osijek, Croatia	Repository provides access to publications and research data produced by the employees and the students of the faculty. Content of repository is open to all users for searching and downloading with clearly stated usage rights.	Social Sciences	Publications; Data	OpenDOAR
University of Zadar Institutional Repository of Evaluation Works, Croatia	Repository provides access to publications and research data produced by the employees and the students of the faculty. Content of repository is open to all users for searching and downloading with clearly stated usage rights.	Social Sciences	Publications	OpenDOAR

University of Zagreb Faculty of Forestry Repository, Croatia	Repository provides access to publications and research data produced by the employees and the students of the faculty. Content of repository is open to all users for searching and downloading with clearly stated usage rights.	Agricultural Sciences	Publications; Data	OpenDOAR
Opencience.ge, Georgia	Repository collects Georgia's research output (publications, theses, ect) and researchers' profiles. Additional functionaly like view/download statistics, bibliometric data for publications and researchers, links to SCOPUS, Pubmed, CORE and other databases. Our goeal is to collect all data related to research and academic publishing in Georgia. We plan to add Research Data, Patents, projects information in nearest future. Almost all data is in Open Access (we work on OA policy). Currently information is collected from a part of Georgain universities and research institutions. The repository is intended to be bi-lingual (Georgian, English). Service is in a development stage, meaning that we are studying user requrements, global practices and testing technologies at the same time.	Multidisciplinary	Publications	No
HELIX Data, Greece	HELIX Data is one of the services provided by the Hellenic Data Service. Hellenic Data Service, also known as HELIX, is a data catalogue and repository supporting knowledge management and scholarly communication in Greece. It is comprised of a diverse set of services, features and functionalities that facilitates responsible research conduct while easing compliance with Open and FAIR practices in the Greek research area. Moreover, HELIX succeeds in linking digital assets of publications, data and processes together thus contributing to information contextualisation and ultimately to building a European data economy of added value services and return on investments through improved research exploitation. HELIX content and tools are inclusive to all, from individual researchers and research communities to SMEs, data and citizen scientists and data enthusiasts. Their use is unlimited, as long as the licenses of the given/selected artifacts permit it, for research as well as educational or experimental purposes. HELIX Data was developed according to the principle of interdisciplinarity to serve the (complex) research needs of cross-disciplinary groups. For that, it supports a wide range of formats and standards and applies open and interoperable solutions for the sharing and preservation of its content in the long-term. Most importantly, HELIX encourages reuse and, since it is still a work in progress, at a later stage it provisions training activities around best practices on open and FAIR publishing and RDM for its users.	Multidisciplinary	Data; Publications	No
National Archive of PhD Theses, Greece	The National Archive of PhD Theses is a service that collects and provides access to the PhD theses from all Higher Education Institutions (HEIs) in Greece as well as PhD theses awarded to Greek scholars by foreign HEIs and certified by the Hellenic NARIC.	Multidisciplinary	Publications	No
Repository service, Greece	The repository service allows to the NI4OS users to deposit and share data via a user-friendly web interface. It can host publications and their associated data or software. It automatically generates a Persistent Identifier for each shared item. Access to shared items can be public or limited to selected repository users.	Generic	Publications; Data	No
KDK Repository, Hungary	The Research Documentation Centre of the Centre for Social Sciences at the Hungarian Academy of Sciences provides information on and access to research conducted at the Centre. The metadata and some of the documents of the Research Documentation Centre (RDC) are available to all visitors, but many are restricted to registered users. The interface can be set to many languages and contains RSS feeds to alert users of new content.	Social Sciences	Publications; Data	OpenDOAR; re3data

Pécsi Egyetemi Archívum (PEA), Hungary	Joint website of the repository of the University of Pécs. It contains legal documents digitized by the University Library and Knowledge Center of the University of Pécs, doctoral dissertations in the doctoral schools of the University of Pécs, full-text documents produced and preserved by the teachers and researchers of the University of Pécs.	Multidisciplinary	Publications	OpenDOAR
REAL Repository of the Academy's Library, Hungary	This site provides access to outputs of projects supported by the Hungarian Scientific Research Fund. Users may set up Atom and RSS feeds to be alerted to new content.	Multidisciplinary	Publications	OpenDOAR
REAL-EOD, Hungary	REAL-EOD contains the full text digitised books produced in the Library - including books digitized in the framework of the E-books On Demand project. This collection holds modern books published by the Library too. Books published by, or related to the Academy and its members belong to this collection as well. The collection holds scientific books, popular books, and digitised copies of items in the library collection - some of them not scientific, or even un-scientific. The repository is operated by the Library and Information Centre of the MTA.	Multidisciplinary	Publications	OpenDOAR
REAL-PhD, Hungary	PhD Thesis repository of the Library and Information Centre of the Hungarian Academy of Sciences. This archive is an orphan repository of hungarian PhD theses. The following Universities use REAL-PhD for archiving their theses: Andrássy Universität Budapest, Lutheran Theological University, Kaposvár University, Károli Gáspár University of the Reformed Church in Hungary, Franz Liszt Academy of Music, Jewish Theological Seminary, University of Pannonia, Pázmány Péter Catholic University, University of Physical Education.	Multidisciplinary	Publications	OpenDOAR
SZTE Repository of Publications, Hungary	The intention of the SZTE Repository of Publications is to make the full text of publications created as a result of scientific and artistic activities at the University available for the widest possible academic audience. Depositing works at the repository secures their long term archiving, and can also increase their viewability and number of citations. This latter is also due to the fact that uploaded documents are indexed by general search engines (e.g. Google, Google Scholar) and professional databases (e.g. BASE, MTA OAI).	Multidisciplinary	Publications	OpenDOAR
Institutional Repository in Medical Sciences – Nicolae Testemitanu SUMPh, Moldova	The Institutional Repository in Medical Sciences of the Nicolae Testemitanu State University of Medicine and Pharmacy of the Republic of Moldova - SUMPh provides access to the research output of the institution. IRMS – SUMPh collects, preserves and gives open access to conference papers, articles from scientific journals, reports, theses and dissertations, study aids, patents, student theses and other publications created by members of the Nicolae Testemitanu University of Medicine and Pharmacy.	Medical Sciences	Publications; Data	OpenDOAR
IREK – AESM: Institutional Repository of Economic Knowledge, Moldova	IREK – AESM: Institutional Repository of Economic Knowledge provides access to the research output of the the Academy of Economic Studies of Moldova.	Business and Economics	Publications; Data	OpenDOAR

Teze: Thesis repository, Moldova	Teze: Thesis repository collects, preserves, gives open access and promotes scientific results obtained in Republic of Moldova by publishing online all PhD theses in PDF format elaborated and defended in the Republic of Moldova. This repository provides access to the research output of PhD students.	Multidisciplinary	Publications	OpenDOAR
PHAIDRA UOM, Montenegro	The University of Montenegro "E-Theses Repository" is an online service developed to host the full-text of PhD and other research theses produced by postgraduate students of the University. They are systematically gathered at the whole of the university from the May of 2012. The material in the archive is available to be browsed, searched, read or printed by anyone interested in its content. All material in the University of Montenegro PHAIDRA Institutional Repository is copyright of the authors and it is authorized for use under Creative Commons (CC) licence regulations.	Multidisciplinary	Publications	No
UGD Academic Repository, North Macedonia	Institutional repository of The Goce Delčev University of Štip.	Multidisciplinary	Publications; Data	Yes
Digital Archive for Ethnological and Anthropological Resources, North Macedonia	Digital Archive of Ethnological and Anthropological Resources (DAEAR), available to all interested in the cultural heritage of Macedonia, are the central and only database of documents in the fields of ethnology and anthropology on Macedonia and wider. These archival databases include collections of ethnographic, folklore and other anthropological materials, special collections of objects, archival material, audio, photo and video materials currently being held at the Institute of Ethnology and Anthropology at the Faculty of Natural Sciences, "St. Cyril and Methodius" Universty in Skopje. All materials collected during research projects and fieldwork at the Institute are archived in these databases.	Ethnology and Anthropology	Data	No
UKIM Repository, North Macedonia	The repository of the University of Ss Cyril and Methodius in Skopje.	Multidisciplinary	Publications	Yes
ARTHRA, Romania	ARTHRA is the institutional repository of „Dunărea de Jos" University of Galați and makes the university scientific output freely available. The digital documents submitted in ARTHRA are also indexed in Google Scholar and they are automatically included in the researcher profile from GS. ARTHRA ensures preservation and management of documents and provides access to: articles of the academic community published in the annals of the university, doctoral theses and abstracts submitted in the university, conference papers/presentations, educational materials (tutorials, library guides), references, papers of library staff. The information retrieval in ARTHRA returns the results of the full content of all documents. For the time being, ARTHRA is included in Duraspace, OpenDOAR and Registry of Open Access Repositories (ROAR).	Multidisciplinary	Publications; Data; OER	Yes

ECOLIB, Romania	ECOLIB ensures preservation and management of documents and provides access to: articles of the ECOIND researchers published in national and international journals (full-paper/abstract), doctoral theses (abstract), conference and workshop papers, short description of the projects developed from 2006 till present, Proceedings of the International Symposium - Environment and Industry. The digital documents submitted in ECOLIB are indexed in Google Scholar.	Environmental Sciences	Publications	Yes
OPEN.ni.ac.rs, Serbia	Repository of publications of the all faculties in University of Nis.	Multidisciplinary	Publications	No
OPEN.uns.ac.rs, Serbia	Repository of publications of the all faculties in University of Novi Sad.	Multidisciplinary	Publications	No
University of Maribor Library Digital Repository, Slovenia	University of Maribor Library (UKM) keeps more than a million units of library materials and the library collection includes many olde and rare items. Many of them are now being digitized and are available mostly freely accessible in the UKM Repository that is accessible to public from 2016. The idea of an information system that would keep library materials in a digital and digitised form and would as such make them accessible to a wider public arose already in 2010. The repository is a result of the cooperation of the Informatics and Digital Library Department, the Acquisition and Technical Services Department and the Local History and Special Collections Department.	Multidisciplinary; Digital Cultural Heritage	Publications	No

As shown in the Figure 11, more than half of the repositories are publication repositories, while 38% supposedly contain publications as well as data or other types of digital assets, e.g. OER or digitized materials. The number of repositories that contain only or predominantly data is the lowest. According to the gathered information, none of the repositories contain software.

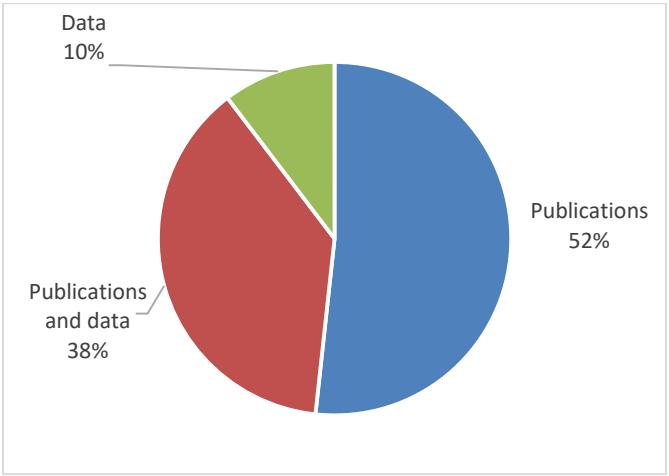


Figure 11. NI4OS-Europe repositories type of content

More than 60% of the repositories are multidisciplinary which is expected as many of these are institutional repositories managed by universities. The rest are domain-specific, 24% pertaining to social sciences (including business and economics and ethnology and

anthropology) and 14% to natural sciences (including medical and agricultural sciences). Figure 12 shows the information as provided in the collected RDTs.

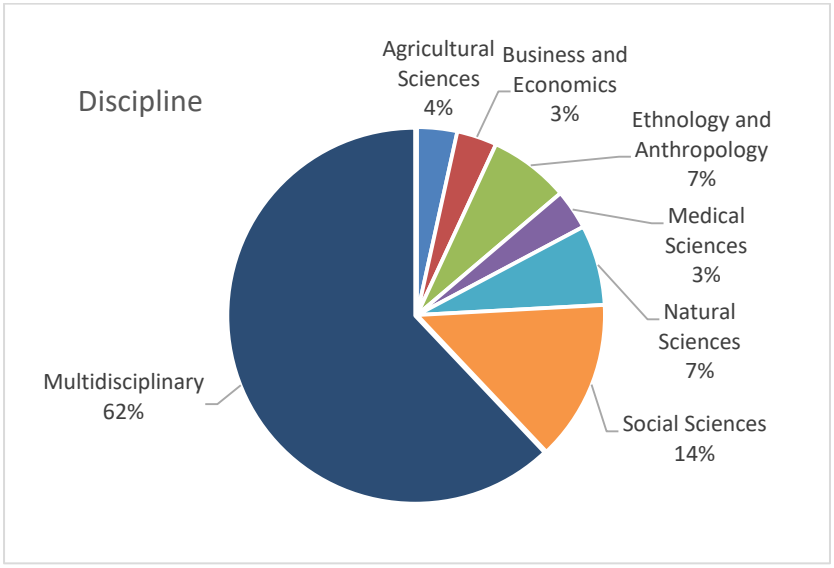


Figure 12. NI4OS-Europe repositories scientific fields

In FAIR terms, repositories included in registries make the repositories and their content findable by both humans and machines. As OpenAIRE allows for the integration of repositories listed in the OpenDOAR or re3data directories, we collected information on how many of the repositories are in fact already included in registries. Figure 13 shows that more than 60% of the repositories are already findable in the relevant registries, while the rest still need to meet this requirement to be able to become integrated in the OpenAIRE infrastructure.

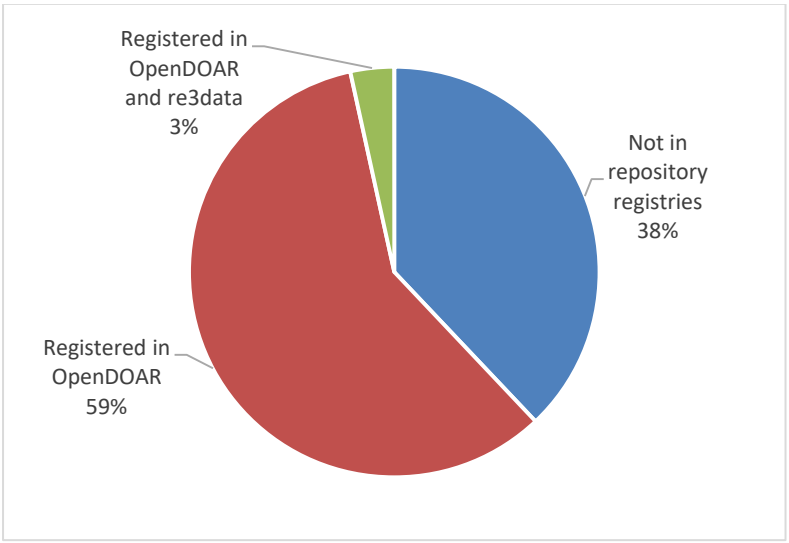


Figure 13. NI4OS-Europe repositories in registries

The features of the underlying software of the repositories may affect the level of interoperability of the repositories, which is an important aspect in achieving effective harvesting of the metadata and contents by major aggregators, such as OpenAIRE. As can be seen in Figure 14, the repository software is quite diverse, although the most represented are the widely used DSpace (various versions, including DSpace-CRIS), EPrints (again, various versions) and Islandora (used by Croatian repositories within the DABAR repository system). The rest use in-house solutions or other software, including ABEKT, ArchiveSpace, CKAN, NeoSite, Omeka-S and PHAIDRA.

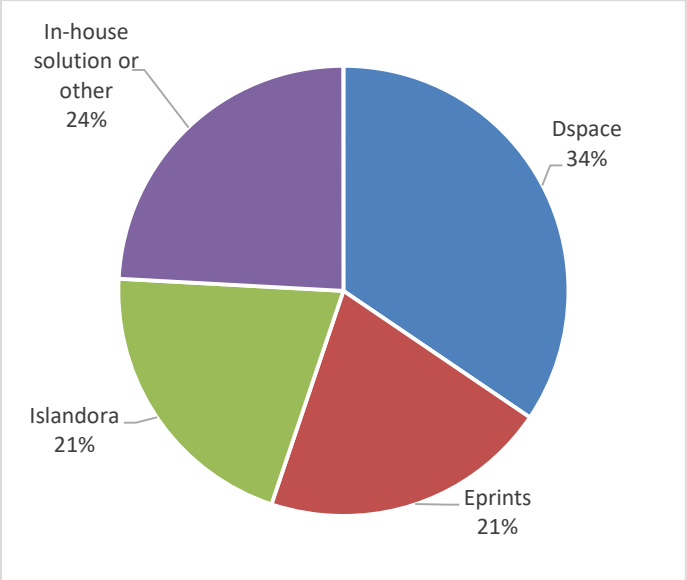


Figure 14. NI4OS-Europe software for repositories

6 Integration timeline

Based on the methodology and classification described in Section 2, here we list resource candidates to be on-boarded and outline an integration to the EOSC timeline of the NI4OS-Europe services and repositories. The development of such a timeline started by gathering the information using the resource description templates described in Section 2. Practically, the on-boarding team members (Table 1) collected this information in collaboration with resource providers, so the quality of the assembled data is expected to be of high degree.

Once the information about resources was gathered, the WP5 team performed an analysis of the collected responses. During the analysis, we considered the general resource description, technical specification of the resource, technology readiness, desired EOSC integration, and desired management integration. All services considered in this section have high-level TRLs (TRL > 7), and they are fully described in terms of the resource description template, whose details are given in previous sections. Since the resource management integration reflects the achieved level of integration with the EOSC, or, in other words, it imposes some concrete obligations on the resource providers based on measured performance, we considered the desired level of integration with the EOSC as a parameter of importance that mainly influences the resource integration timeline production. Therefore, we have sorted different types of resources per their desired level of integration. In addition to this, to reduce the complexity of the EILs, the results are illustrated using the cumulative level of integration, which is a more descriptive classification that recognizes low, medium, and high levels of integrations with the EOSC, as it was described in WP3.

Independently for each resource type, we have sorted resources based on the desired level of integration and within the RDT reported the expected integration timeline. Figure illustrate pools of generic and thematic services, and repositories organized by EILs. In each figure, the x-axis provides a timeline and y-axis gives accumulates the number of particular resources. With the red colour we have illustrated a low level of integration, with orange a medium level of integration, and with green a high level of desired integration with the NI4OS-Europe pre-production environment. Once a service moves from red to orange area, it is ready to be published within the project's catalogue, and we consider that it is on-boarded. It means that a resource satisfies minimal requirements, which, in particular, includes basic resource description, user manual, terms of use, privacy policy, and integration with the helpdesk. However, our aim is to establish a higher level of integration when possible, which includes additional unification with the PID, AAI, monitoring, and accounting systems. For this, additional work will be required, and the borders between orange and green areas provide estimates when such a level of integration is expected, per service. Once the service is fully integrated with the project's pre-production environment, which practically understands that the project is able to sufficiently monitor the performance of the service, it is ready to be published within the EOSC catalogue, and we consider that the service is on-boarded to the EOSC.

From these pools of resource candidates, we will choose KPI-defined number of resources (20 generic services, 20 thematic services, and 15 repositories) that will be on-boarded to the EOSC during the project's lifetime. These will be chosen by their importance to the scientific community, the capacity of the resource, and on the desired level of integration.

Therefore, integration with the EOSC will not be exclusively driven by EILs and MILs, but by the scientific impact of a particular resource and its capacity as well. Moreover, since the exact rules of on-boarding and participation change dynamically in the EOSC landscape, this NI4OS-Europe onboarding plan and levels that services reach during integration can change in the course of the project.

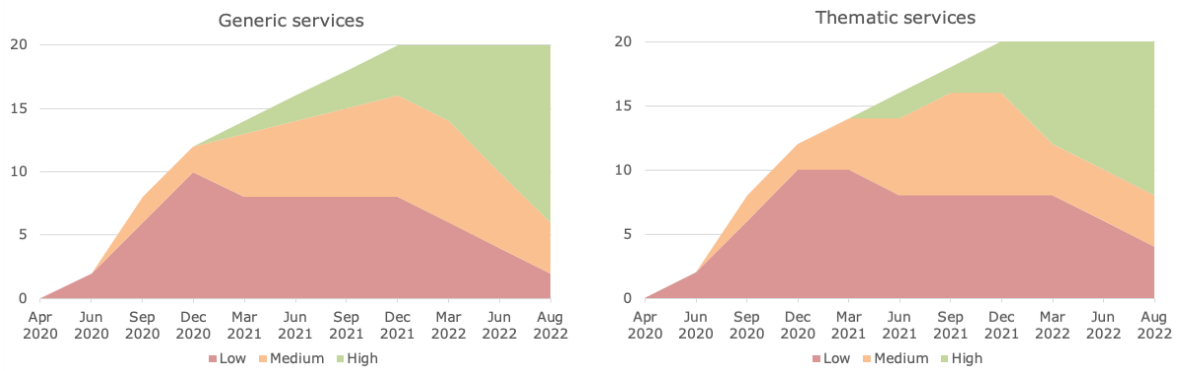
In all three figures (Figure 15), there are two resources whose on-boarding to the EOSC will start immediately, in April 2020. These we call quick-win resources, and from the operational perspective, the main purpose of their on-boarding is examination and verification of the current EOSC on-boarding procedures. Also, by their integration, the project will establish an initial connection with the EOSC on-boarding team, which we believe will give us a better understanding of the existing and close tracking of expansion and development of upcoming on-boarding-related procedures. The selected quick-win services were previously used in the production mode and actively monitored by the regional monitoring system, and their availability statistics were very high in the previous several years. The quick-win repositories are chosen primarily based on the underlying software, full-text search capability, and the number of items they contain.

Figure 15. Number of NI4OS-Europe resource candidates for on-boarding and their desired level of integration with the NI4OS-Europe pre-production environment

In parallel with the on-boarding of the quick-wins, the WP5 team will coordinate and provide the necessary support for three different types of resources. According to the integration timeline (Figure 16), in the period April - December 2020, we plan to fulfil all low-level integration requirements for approximately half of the resources from all three groups and to accomplish the middle-level of integration for quick-wins. Practically, we

expect that 12 generic services, 12 thematic services, and 8 repositories will be exposed through the project's resource portfolio/catalogue system, and equipped with a corresponding user manual, terms of use, privacy policy, as well as integrated with the project's helpdesk system. In addition to this, the quick-win resources will be fully incorporated into the project's pre-production environment, while the quick-win repositories integrated within the OpenAIRE platform. Practically, these six resources will be on-boarded to the EOSC by the end of this year. So, in the deliverable D5.2 (December 2020), in our first report on provider and repository integration, we plan to report on the integration with the project's pre-production environment of these 32 resources (12 generic services, 12 thematic services, and 8 repositories), out of which six will be already on-boarded into the EOSC catalogue. All 32 resources, integrated into our horizontal platform, will be offered for user community testing through the WP6 starting from January 2021.

Further on-boarding activities, in the period January - December 2021, will be mainly focused on the capacity building. In this period, we will aim to integrate with the project's pre-production environment the KPI-defined number of resources (20 generic services, 20 thematic services, and 15 repositories) with low EILs. That means that 55 resources will be registered within the project's catalogue, and we expect that 14 of them will achieve a high-level EIL, 20 of them a medium-level EIL, and 21 of them a low-level of integration with the pre-production environment. This is illustrated in Figure 16 for all three different types of resources under consideration. The figure accumulates the number of integrated resources during the project's lifetime. In these stack areas graphs, the expected number of resources with low-level, medium-level, and high-level of integration with the project's pre-production environment is presented with red, orange, and green areas, respectively. December 2021, when the KPI-defined number of resources should be exposed through the project's catalogue, is clearly visible. This is very important since it will allow us to offer all our resources (55 of them) through the WP6 open call for production use of services and repositories. Such a call, according to the WP6 plan, will be open in January 2022.



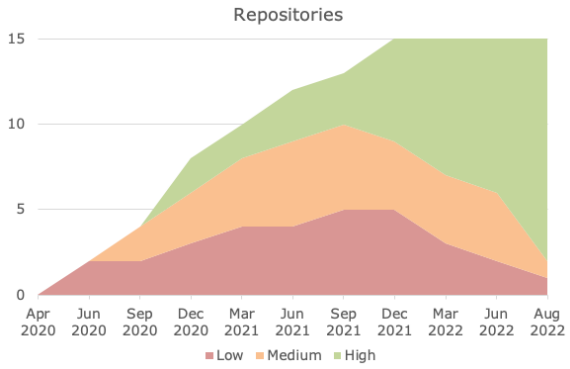


Figure 16. NI4OS-Europe integration timeline and the corresponding EIL for generic and thematic services, and repositories.

From December 2021 and until the end of the project, the on-boarding team will be focused on the capability of the integrated services and repositories, and their on-boarding to the EOSC catalogue. We intend to achieve a sufficient level of integration with the EOSC for the services and repositories registered within the catalogue. This will be performed not only by providing a direct support to the resource providers, but also by producing guidelines and know-how documents. In the period June - December 2021 the WP5 team will analyse feedback from the WP6 pre-defined user communities (in January 2021), which were provided with access to resources for testing purposes. After such an analysis, the concrete action list will be created, and operational aspects of the offered resources will be upgraded according to them.

From January 2022 until the end of the project, the WP5 team will try to improve and sharpen the services management procedures produced by the WP3 team, and in this way, to increase the integration level of all resources within the project's catalogue. In this period, achieving the higher level of integration, the resources will be gradually on-boarded to the EOSC and exposed through the EOSC catalogue. This will lead to a set of 55 services and repositories integrated into the EOSC, which was our aim from the beginning of the project.

7 Conclusions

The building of the EOSC catalogue of resources available to the European researchers at a single point of access strongly depends on the identifications of the national and regional resources and their integration into this common catalogue. All EOSC implementation projects, including NI4OS-Europe, among other goals, have the important task of recognizing the resources of the project partners, helping them to strengthen their capacities and providing the environment to on-board these resources to EOSC.

To better understand and evaluate the state and the maturity of these resources, this document provides several types of categorization. These categorizations include TRL, EOSC integration level and resource management levels. All three are combined to produce a matrix of targets that the resources are expected to achieve to be successfully on-boarded.

The three main categories of resources to be on-boarded to EOSC are generic services, thematic services and repositories. For each of these categories, a detailed categorization is proposed in this document. Based on that categorization, data for 90 resources are collected, using the resource description template, that will later be used to register these resources in the resource catalogue.

Based on all proposed metrics, as well as the collected data, a detailed integration timeline is given. The timeline defines the activities needed to be undertaken by a NI4OS-Europe on-boarding team containing representatives from each partner country for each of the types of resources.

This roadmap presents the best-case scenario and is subject to change depending on the evolution of on-boarding and participation rules and processes within EOSC – something which is not under NI4OS-Europe control. Similarly, the levels that services reach during the integration can also change in the course of the project depending on overall EOSC rules and developments.

Appendix I

Resource Description Template

Basic Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Name	*Single Line*	*X*	*X*		*Name of this specific service as assigned by the service provider. We recommend the name is fair/t short and descriptive.*	*In general take as-is because it is specific to the service and the service name selected by the Service Provider. Only ask for clarifications if the name does not correspond with the name of the service name provided on the Service URL or in case of obvious typo's.* *Examples:* *PARADOX-IV*	
Endpoint	*URL*	*X*	*X*		*Main URL to use the service (in the case of networked service)*	*Verify if the URL is a valid URL (use the URL within a browser), if the URL refers to the Web User Interface (WUI) of the service.*	
Identified	*Single selection list*		*X*		*When the service is identified: proposal stage, 1st year, 2nd year, 3rd year of the project.*		
Technology readiness level	*Single selection list*	*X*	*X*		*Technology readiness level* *https://wiki.ni4os.eu/index.php/Technology_Readiness_Level*		
Marketing Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Tagline	*Single Line*	*X*	*X*		*Short catch-phrase for marketing and advertising purposes (1 line). It will be usually displayed close the service name and should refer to the main value or purpose of the service.*	*In general take as-is because it is specific to the service and the service name selected by the Service Provider. Only ask for clarifications if the name does not correspond with the name of the service name provided on the Service URL or in case of obvious typo's.* *Examples:* *Serbian supercomputing cluster*	
Description	*Freetext*	*X*	*X*		*A high-level description in fairly non-technical terms of a) what the service/resource does, functionality it provides and resources it enables to access, b) the benefit to a user/customer delivered by a service; benefits are usually related to alleviating pains (e.g. eliminate undesired outcomes, obstacles or risks) or producing gains (e.g. increased performance, social gains, positive emotions or cost saving), c) list of customers, communities, users, etc. using the service.*	*In general take as-is because it is specific to the service provided by the Service Provider. Review the text, not to qualify the quality, but more on readability and for obvious typo's, if necessary, ask for clarifications. Also point out if it does not appear to address the potential customers and explain the main functionality of the service. If description is excessively long suggest being more concise.*	
Logo URL	*URL*	*X*	*X*		*Link to the logo/visual identity of the service.*	*Verify if the URL points to a logo picture or source including a logo that relate to the service in question, If the URL is not a valid URL or does not reference a valid logo, contact requester for clarification.* *Example:* *www.provider.eu/downloads/branding-pack.zip* *www.acmeIT.com/files/image.jpg*	
Website URL	*URL*		*X*		*Web page with information about the service. This will be hosted by the project (official web site and/or wiki).*	*Verify if the URL is a valid URL (use the URL within a browser), if the URL refers to the web page of the service, this is commonly the web page of the service in the Service Catalogue of the Service Provider. Verify if the name on the web page corresponds to the name provided as Service Name.* *Example: https://www.scl.rs/PARADOXClusterUserGuide*	
Target Users	*Multiselect list*	*X*	*X*	*Choose from the list below*	*Type of users/customers that commissions a service/resource provider to deliver a service.*	*Confirm they chose from list provided.*	
					Researchers		
				FALSE	Research groups		
				FALSE	Research communities		
				FALSE	Research projects		
				FALSE	Research managers		
				FALSE	Research organizations		
				FALSE	Students		
				FALSE	Innovators		
				FALSE	Businesses		
				FALSE	Resource providers		
				FALSE	Funders		
				FALSE	Policy Makers		
				FALSE	Research infrastructure managers		
				FALSE	Resource Provider Managers		
				FALSE	Resource Managers		
Classification Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Scientific domain	*Single selection list*	*X*	*X*		*The branch of science, scientific discipline that is related to the service/resource*		
Location Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Geographical availability	*Single selection list*	*X*	*X*		*Country that the service is available, if the access is regional use: Worlwide, Europe, Western Balkans*		
Language	*Single selection list*	*X*	*X*				
Service provider							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Service provider	*Single selection list*	*X*	*X*		*Acronym of the Service Provider*		
Service integration with NI4OS services							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Terms of use	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
User manual	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
Helpdesk	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
Privacy policy	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
AAI integration with NI4OS	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
PID for services	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
Monitoring	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
Accounting	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
Training resources	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
Code repository (for thematic services)	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
Generic service integration (for thematic services)	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
Datasets production (for thematic services)	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
API production (for thematic services)	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		
Workflow production (for thematic services)	*Single Line*	*X*	*X*		*If available, insert the URL, else indicate when the document/functionality will be produced and available online. For dates use MM.YYYY. If not planned at all use N/A*		

Resource Description Template – Resource Provider

Basic Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Acronym	*Single Line*	*X*	*X*				
Legal name	*Single Line*	*X*	*X*		*Full legal name of the organization providing/offering the service/resource and acting as main contact point. The Legal Name must correspond to the official legal name in the statute or the registration act/decree establishing the organization.*		
Legal status	*Single selection list*	*X*	*X*		*The legal status is usually noted in the registration act/statute of the organization.*		
Website URL	*URL*	*X*	*X*		*Webpage with information about the provider.*		
Marketing Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Description	*Freetext*	*X*	*X*		*The description of the provider.*		
Logo URL	*URL*	*X*	*X*		*Link to the logo/visual identity of the provider.*		
Classification Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Scientific domain	*Multiselect list*	*X*	*X*		*A named group of providers that offer access to the same type of resource or capabilities.*		
Type	*Single selection list*	*X*	*X*		*Defines if the provider is single-sited, distributed, mobile, virtual, etc.*		
Location Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Street name and number	*Single Line*	*X*	*X*				
Postal code	*Single Line*	*X*	*X*				
City	*Single Line*	*X*	*X*				
Country	*Single selection list*	*X*	*X*				
Latitude	*Single Line*	*X*	*X*				
Longitute	*Single Line*	*X*	*X*				
Contact Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
First name	*Single Line*	*X*	*X*		*First name of the provider's main contact person/provider manager.*		
Last name	*Single Line*	*X*	*X*		*Last name of the provider's main contact person/provider manager.*		
E-mail	*email*	*X*	*X*		*Email of the provider's main contact person/provider manager.*		
Public contact							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
E-mail	*email*	*X*	*X*		*Email of the provider's contact person to be displayed at the portal or general email to contact organisation.*		

Resource Description Template – Generic Service - Cloud

Basic Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Server specification	*Single Line*	*X*	*X*		*This might be*		
Number of servers	*Single Line*	*X*	*X*		*Number of*		
CPU details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
CPU specification	*Single Line*	*X*	*X*				
CPUs per server	*Single Line*	*X*	*X*				
Cores per CPU	*Single Line*	*X*	*X*				
RAM per server [GB]	*Single Line*	*X*	*X*				
RAM per core [GB]	*Single Line*	*X*	*X*				
Total number of CPU-cores	*Single Line*	*X*	*X*				
Filesystem details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Total storage [TB]	*Single Line*	*X*	*X*				
Virtual Machine specification							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Minimum number of CPU cores per VM	*Single Line*	*X*	*X*				
Minimum amount of RAM per VM [GB]	*Single Line*	*X*	*X*				
Maximum number of CPU cores per VM	*Single Line*	*X*	*X*				
Maximum amount of RAM per VM [GB]	*Single Line*	*X*	*X*				
Maximum amount of storage per VM [GB]	*Single Line*	*X*	*X*				
Software details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
VM management type	*Single Line*	*X*	*X*		*Cloud stack*		
Supported interfaces	*Single Line*	*X*	*X*		*Supported API interfaces. Web access is also an interface*		

Resource Description Template – Generic Service - HPC

Basic Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Peak performance [TFlops]	*Single Line*	*X*	*X*		*Theoretical*		
Server specification	*Single Line*	*X*	*X*		*This might be*		
Number of servers	*Single Line*	*X*	*X*		*Number of*		
CPU details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
CPU Specification	*Single Line*	*X*	*X*				
CPUs per server	*Single Line*	*X*	*X*				
Cores per CPU	*Single Line*	*X*	*X*				
RAM per server [GB]	*Single Line*		*X*				
RAM per core [GB]	*Single Line*	*X*	*X*				
Total number of CPU-cores	*Single Line*	*X*	*X*				
Max number of parallel processes	*Single Line*	*X*	*X*				
CPU peak performance [Tflops]	*Single Line*	*X*	*X*				
Accelerator details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Accelerator specification	*Single Line*		*X*				
Total number of accelerators	*Single Line*		*X*				
Accelerators per server	*Single Line*		*X*				
Maximal number of accelerators per server	*Single Line*		*X*				
Accelerators peak performance [Tflops]	*Single Line*		*X*				
Interconnection details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Interconnect type	*Single Line*	*X*	*X*				
Interconnect latency [µs]	*Single Line*	*X*	*X*				
Interconnect bandwidth [Gbps]	*Single Line*	*X*	*X*				
Filesystem details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Local filesystem type	*Single Line*	*X*	*X*				
Total storage [TB]	*Single Line*	*X*	*X*				
Software details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Operating system	*Single Line*	*X*	*X*				
Batch system/scheduler	*Single Line*	*X*	*X*				
Development tools	*Freetext*	*X*	*X*				
Libraries	*Freetext*	*X*	*X*				
Applications	*Freetext*	*X*	*X*				

Resource Description Template – Generic Service - Storage

Basic Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Storage specification	*Single Line*	*X*	*X*				
Total storage [TB]	*Single Line*	*X*	*X*				
Storage technology	*Single selection l*	*X*	*X*		*SSD, SAS,*		
Storage performance	*Single Line*		*X*		*IOPS*		
Software details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Supported interfaces	*Single Line*	*X*	*X*		*List of protocols*		
Supported storage types	*Multiselect list*		*X*		*Available*		

Resource Description Template – Thematic Service

Basic Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Service name	*Single selection l*	*X*	*X*		*Name of the service as entered in the service description*		
Computational/storage requirements							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Computational requirements	*Freetext*	*X*			*What are the computational requirements of the Service?*		
Requires VM(s)	*Yes/No*	*X*			*Does the service require VMs from generic service providers? Yes/No*		
Number of VM(s)	*Freetext*	*X*			*Number and type of Virtual Machines does the service require? Please specify vCPU, Memory*		
Storage requirements [GB]	*Single Line*	*X*			*How much storage does the service require? Specify capacity and access protocol. For local disks use local.*		
Require HPC	*Yes/No*	*X*			*Does the service require the use of High Performance Computing? Yes/No*		
HPC resources	*Single Line*				*Describe the ammount of HPC resources and time schedule*		
Require GPUs as part of HPC	*Yes/No*	*X*			*Does the service require the use of GPUs as part of HPC?*		
Amount of GPUs	*Single Line*						
Requre VMs with GPU acelleration	*Yes/No*	*X*			*Does the service requre the use of VMs with GPU acelleration?*		
Amount of VMs with GPUs	*Single Line*						
Further Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Deployed in an existing infrastructure	*Yes/No*	*X*	*X*		*Has the service been used in an existing Infrastructure?*		
Workflows available	*Yes/No*	*X*	*X*		*Are there workflows available for thes service?*		
Pipeline availables	*Yes/No*	*X*	*X*		*Are there pipeline availables for the service?*		
Uses data stored in repository	*Yes/No*	*X*	*X*		*Does the service uses data stored in repository?*		
Produce data	*Yes/No*	*X*			*Does the service produce data?*		
Amount of produced data per year	*Single Line*				*Estimated ammount of data produced*		
Metadata schema of the data	*Single Line*				*What metadata schema do the data support?*		
Ontologies of the data	*Single Line*				*What ontologies do the data support?*		
GDPR compliance	*Single Line*				*Does this data follow the EU GDPR principles?*		
Software details							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
License of the software	*Single Line*	*X*	*X*				
Repository	*URL*	*X*	*X*				
Form Status							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation cri**	**Notes**
Complete?	*Single selection l*	*X*	☐				

Resource Description Template – Repository

Basic Information							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Repository name	*Single selection list*	*X*	*X*		*Name of the service as entered in the service description*		
Type of content	*Single Line*	*X*	*X*		*objects: publications, data, e-publishing platform, digital library, OER, software ...*		
Repository software and version	*Single Line*	*X*	*X*		*eg. DSpace 6.3*		
FAIRness of the repository							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Access method	*Single Line*	*X*	*X*		*eg. open, restricted, authentication*		
Is the repository indexed in a directory/registry? If yes, please specify.	*Single Line*	*X*	*X*		**OpenDOAR** (https://v2.sherpa.ac.uk/opendoar/) **re3data** (https://www.re3data.org/)		
Is the repository registered or indexed in a searchable resource? If yes, please specify.	*Single Line*	*X*	*X*		*eg. OpenAIRE, BASE, CORE ...*		
Are objects assigned globally unique and persistent identifiers?	*Single Line*	*X*	*X*		*eg. DOI*		
What metadata standard does the repository use?	*Single Line*	*X*	*X*		*https://rdamsc.dcc.ac.uk/scheme-index*		
Are objects released with a clear and accessible data usage license?	*Single Line*	*X*	*X*		*e.g. CreativeCommons*		
OpenAIRE compliancy							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Does the repository support an OpenAIRE-compliant metadata schema? If yes, please specify.	*Single Line*	*X*	*X*		**Dublin Core** (https://www.dublincore.org/schemas/) **DataCite** (https://schema.datacite.org/)		
Does the repository use the OAI-PMH protocol for metadata exchange? If yes, please provide the OAI-PMH.	*Single Line*	*X*	*X*		*https://www.openarchives.org/pmh/*		
Certification and policies							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Does the repository hold a CoreTrustSeal or other repository trustworthiness certificates?	*Single Line*	*X*	*X*				
What policies does the repository have in place?	*Single Line*		*X*		*http://sherpa.ac.uk/policytool/*		
Form Status							
Entry name	**Entry type**	**Required**	**Public**	**Your answer**	**Guidance**	**Validation criteria**	**Notes**
Complete?	*Single selection list*	*X*					