EUROPEAN COMMISSION
Directorate-General for Research and Innovation
Research infrastructures

ANNEX 1 (part A)

Research and Innovation action

NUMBER — 857650 — EOSC-Pillar

Table of Contents

1.1. The project summary

Project Number [1]	857650	Project Acronym [2]	EOSC-Pillar

One form per project

General information	
Project title [3]	Coordination and Harmonisation of National Inititiatives, Infrastructures and Data services in Central and Western Europe
Starting date [4]	01/07/2019
Duration in months [5]	36
Call (part) identifier [6]	H2020-INFRAEOSC-2018-3
Topic	INFRAEOSC-05-2018-2019 Support to the EOSC Governance
Fixed EC Keywords	
Free keywords	National Initiatives, EOSC, Cloud, Research Infrastructures, e-Infrastructures,

Abstract [7]

EOSC-Pillar gathers representatives of the fast-growing national initiatives for coordinating data infrastructures and services in Italy, France, Germany, Austria and Belgium to establish an agile and efficient federation model for open science services covering the full spectrum of European research communities.
Our proposal aims to implement some of the main pieces of the EOSC jigsaw within a science-driven approach which is efficient, scalable and sustainable and that can be rolled out in other countries. National initiatives are the key of our strategy, for their capacity to attract and coordinate many elements of the complex EOSC ecosystem and for their sustainability, which will add resilience to the whole structure. We will combine these initiatives, who represent research communities in each country, with use cases of transnational networks working to implement FAIR data practices.
Through the coordination of national initiatives, EOSC-Pillar will be able to support the gradual alignment of policy and practice among countries and compliance to EOSC standards.
We are convinced that by federating national initiatives through common policies, FAIR services, shared standards, and technical choices, EOSC-Pillar will be a catalyst for science-driven transnational open data and open science services offered through the EOSC portal.
These initiatives will emanate the promotion of FAIR data practices and services across scientific communities, sharing best practice, and igniting opportunities for interdisciplinary approaches in the EOSC.
Above all, our vision is that national initiatives are key to involve user communities and research infrastructures both as test-beds for solutions, but also in their very design and sustainable evolution. For this reason, EOSC-Pillar's workplan is built around selected user-driven pilots from 7 scientific domains, that will show EOSC in action and provide valuable input to guide the roll out of services for other communities.

1.2. List of Beneficiaries

Project Number [1]	857650	Project Acronym [2]	EOSC-Pillar

List of Beneficiaries

No	Name	Short name	Country	Project entry month[8]	Project exit month
1	CONSORTIUM GARR	GARR	Italy	1	36
2	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR	Italy	1	36
3	CINECA CONSORZIO INTERUNIVERSITARIO	CINECA	Italy	1	36
4	ISTITUTO NAZIONALE DI FISICA NUCLEARE	INFN	Italy	1	36
5	FONDAZIONE CENTRO EURO-MEDITERRANEOSUI CAMBIAMENTI CLIMATICI	FONDAZIONE CMCC	Italy	1	36
6	UNIVERSITAT WIEN	UNIVIE	Austria	1	36
7	CINES	CINES	France	1	36
8	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS	CNRS	France	1	36
9	INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE	INRA	France	1	36
10	INSTITUT NATIONAL DE RECHERCHE ENINFORMATIQUE ET AUTOMATIQUE	INRIA	France	1	36
11	INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATION DE LA MER	IFREMER	France	1	36
12	INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE	INSERM	France	1	36
13	KARLSRUHER INSTITUT FUER TECHNOLOGIE	KIT	Germany	1	36
14	DEUTSCHES KLIMARECHENZENTRUM GMBH	DKRZ	Germany	1	36
15	FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.	Fraunhofer	Germany	1	36
16	HELMHOLTZ ZENTRUM POTSDAM DEUTSCHESGEOFORSCHUNGSZENTRUM GFZ	GFZ	Germany	1	36
17	UNIVERSITEIT GENT	UGent	Belgium	1	36
18	TRUST-IT SRL	TRUST-IT SRL	Italy	1	36

1.3. Workplan Tables - Detailed implementation

1.3.1. WT1 List of work packages

WP Number[9]	WP Title	Lead beneficiary[10]	Person-months[11]	Start month[12]	End month[13]
WP1	Management	1 - GARR	58.00	1	36
WP2	The human factor of the EOSC: Dissemination, Outreach and Community building	18 - TRUST-IT SRL	71.00	1	36
WP3	National Initiatives Survey	6 - UNIVIE	82.00	1	36
WP4	From National Initiatives to trans-national services	1 - GARR	92.00	1	36
WP5	The Data layer: establishing FAIR data services at the national and transnational level	15 - Fraunhofer	130.00	1	36
WP6	EOSC in action: Use cases and community-driven pilots	8 - CNRS	295.00	1	36
WP7	The Infrastructure layer: delivering horizontal data storage and computing services, from national to transnational	8 - CNRS	122.00	1	36
WP8	Ethics requirements	1 - GARR	N/A	1	36
		Total	850.00		

1.3.2. WT2 list of deliverables

Deliverable Number[14]	Deliverable Title	WP number[9]	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D1.1	Quality assurance, risk management and innovation management plan	WP1	18 - TRUST-IT SRL	Report	Public	2
D1.2	Data Management Plan and Policies	WP1	6 - UNIVIE	ORDP: Open Research Data Pilot	Public	5
D2.1	MoU signed with INFRAEOSC05 -a and -c projects	WP2	4 - INFN	Report	Public	3
D2.2	EOSC-Pillar Dissemination, Communication and Stakeholder Engagement Plan	WP2	18 - TRUST-IT SRL	Report	Public	4
D2.3	Initial report on the status of the relationship with the EOSC Governance and the EOSC related initiatives	WP2	4 - INFN	Report	Confidential, only for members of the consortium (including the Commission Services)	8
D2.4	Second report on the status of the relationship with the EOSC Governance and the EOSC related initiatives	WP2	4 - INFN	Report	Confidential, only for members of the consortium (including the Commission Services)	24
D2.5	EOSC-Pillar Dissemination, Communication and Stakeholder Engagement Report	WP2	18 - TRUST-IT SRL	Report	Public	17
D2.6	Final report on EOSC-Pillar dissemination, Communication and Stakeholder Engagement	WP2	18 - TRUST-IT SRL	Report	Public	33
D3.1	Summary Report on the National Initiatives survey	WP3	6 - UNIVIE	Report	Public	12
D3.2	Final Report on the communities ad usage of services	WP3	6 - UNIVIE	Report	Public	36

Deliverable Number[14]	Deliverable Title	WP number[9]	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D4.1	Legal and Policy framework and federation blueprint	WP4	13 - KIT	Report	Public	16
D4.2	Signed MoU with Regional initiatives	WP4	1 - GARR	Other	Public	6
D4.3	Roadmap for consolidating National Initiatives	WP4	4 - INFN	Report	Public	24
D4.4	National Service Registry Prototype	WP4	2 - CNR	Demonstrator	Public	24
D4.5	Business model and sustainability study	WP4	8 - CNRS	Report	Public	36
D5.1	FAIR Research Data Management Tool set - first release	WP5	7 - CINES	Other	Public	12
D5.2	FAIR research Data Management Workbench Operation Report	WP5	2 - CNR	ORDP: Open Research Data Pilot	Public	18
D5.3	Training Plan	WP5	17 - UGent	Report	Public	4
D5.4	FAIR-oriented Research Data Management: Support, Training and Assessment Activity Report	WP5	17 - UGent	Report	Public	18
D5.5	Collated set of ontologies as foundation for interoperability	WP5	15 - Fraunhofer	ORDP: Open Research Data Pilot	Public	36
D5.6	Fair Research Data Management Tool set update	WP5	7 - CINES	Other	Public	24
D5.7	FAIR Reseacrh Data Management Workbench Report update	WP5	2 - CNR	ORDP: Open Research Data Pilot	Public	36
D5.8	Training plan update	WP5	17 - UGent	Report	Public	20
D5.9	FAIR-oriented research data management: Support, Training and Assessment Activity Report update	WP5	17 - UGent	Report	Public	36
D6.1	State of the Art and Community Needs Report from Use Cases	WP6	8 - CNRS	Report	Public	8

Deliverable Number[14]	Deliverable Title	WP number[9]	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D6.2	Demonstrator and success stories from the use cases	WP6	8 - CNRS	Demonstrator	Public	24
D6.3	Final Report on Use Cases and community involvement	WP6	8 - CNRS	Report	Public	36
D7.1	Guidelines and Recommendations for the technical integration of resources and services into the EOSC	WP7	4 - INFN	Report	Public	24
D7.2	Report on service integration	WP7	13 - KIT	Report	Public	36
D7.3	Report on the validation statistics, operational infrastructure services and recommendations for future integration work	WP7	13 - KIT	Report	Public	30
D7.4	Report on transnational Service Usage	WP7	8 - CNRS	Report	Public	36
D8.1	POPD - Requirement No. 1	WP8	1 - GARR	Ethics	Confidential, only for members of the consortium (including the Commission Services)	4

1.3.3. WT3 Work package descriptions

Work package number [9]	WP1	Lead beneficiary [10]	1 - GARR
Work package title	Management		
Start month	1	**End month**	36

Objectives

- Efficiently manage and coordinate the project consortium and ensure timely execution;
- Ensure quality assurance, validation, risk management and innovative methods are adopted;
- Technically coordinate the project activities;
- Draft and implement a data management plan and handle all ethical and legal issues
- Monitor the results and implement corrective measures, if necessary
- This WP will liaise with the EC and the community.

Description of work and role of partners

WP1 - Management [Months: 1-36]
GARR, CINES, TRUST-IT SRL
This work package will be responsible for the coordination of project activities, ensuring the timely delivery of the programme and productive, fair, and effective collaboration between partners. The EOSC-Pillar project appears to be challenging from the management point of view due to the number of activities, partners and linked parties involved. Moreover, the project has to organize and develop several high level activities and monitor their impact. It is crucial to keep all partners on board by clear communication to partners All these aspects will be addressed, at the management level, by a correctly articulated structure which will avoid overloads and unnecessary bureaucracy. Moreover, the compact geographical coverage of the project poses several challenges in the interaction with other European regions, while the work-plan has activity lines, derived from the project's objectives, which can span all the regions involved in the EOSC. The managerial structure has thus been designed with a traditional technical management that develops horizontally across the countries and also vertically addresses the different regions taking into consideration their specificities. GARR will lead the project and all the activities which are strictly related to the overall financial and administrative management of the project, while the Technical Management will be performed by CINES. Managerial activities which are related to the other work-packages have been included in the respective work plans.

Task 1.1: Administrative and Financial (GARR) This task will cover the daily administration of the project, as well as the financial planning and control, including the reallocation of resources in case of need; it will be responsible for the delivery of periodical financial reports. T1.1 will take care of the administrative, financial and overall management of the consortium. This activity will be managed by GARR that will be responsible for the delivery of financial/administrative documents and periodical reports. It will include the daily management of the project, the communication amongst the consortium, the Project Office (PO) activities and the resolution of possible issues. Periodical Project Management Boards (PMB) and PO meetings will be held to ensure the above. All partners will be represented in the PMB, while the PO will be maintained by GARR.
Task 1.2: Technical Management (CINES) This task will oversee the project technical activity. It will be responsible of the daily planning and control of the activity, of the timely delivery of milestones and deliverables, of proposing changes and corrective actions in the workplan when needed. T1.2 will aim at the coordination of the overall technical activities, for the achievement of milestones and the delivery of the deliverables according to the foreseen work plan. It will mainly be performed by the Technical Manager (TM) who will monitor the results, check their quality and implement corrective measures, if needed. Periodical Technical Board (TB) meetings will be part of the coordination activity.
Task 1.3: Quality Assurance, Risk Management and Innovation management (Trust-IT) This task will be responsible of setting up quality management procedures and checking the quality of the project's outputs; will identify a risk mitigation and management strategy and evolve it if needed; will propose and implement an innovation management plan; will deliver the project's Data Management Plan.

Participation per Partner

Partner number and short name	WP1 effort
1 - GARR	30.00
7 - CINES	20.00
18 - TRUST-IT SRL	8.00
Total	58.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D1.1	Quality assurance, risk management and innovation management plan	18 - TRUST-IT SRL	Report	Public	2
D1.2	Data Management Plan and Policies	6 - UNIVIE	ORDP: Open Research Data Pilot	Public	5

Description of deliverables

D1.1: Quality assurance, risk management and innovation management plan [R, PU, TRUST-IT, M02]

D1.2: Data Management Plan and Policies [ORDP, PU, UNIVIE, M05]

D1.1 : Quality assurance, risk management and innovation management plan [2]

The document will outline the quality assurance, risk management and innovation plans for the project and set out the relevant procedures to be adopted by the consortium.

D1.2 : Data Management Plan and Policies [5]

This document will present the Data Management Plan (DMP) for the project, which will cover a: a) privacy/ethical processes of managing the data collected, processed or generated, and b) modalities for the publishing of documents. This will allow to properly share and publish all documents produced in the course of the activities, as this will affect future decision-making. It will define what methodology and standards will be adopted for handling any data; how this data will be shared and/or made open; and how the data will be curated and preserved during and after the project within the definition and scope of EOSC vision.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS1	Project KOM organised	1 - GARR	1	1st project meeting to officially present the project objectives and kick-off activities to the project participants, introduce the project management team, the activity leaders and project workers to each other.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS2	Project set up and running, Quality Plan and Risk Mitigation Plan Lauched	1 - GARR	2	This milestone marks the launch, through the appropriate project communication channels, of full-fledged project procedures, including QA, Risk and Innovation management.

Work package number [9]	WP2	Lead beneficiary [10]	18 - TRUST-IT SRL
Work package title	The human factor of the EOSC: Dissemination, Outreach and Community building		
Start month	1	End month	36

Objectives

• Build the liaisons of EOSC-Pillar with the different stakeholders in the EOSC landscape: develop a communication strategy, setup concertation channels with the EOSC Governance and with the other EOSC related initiatives, build up communities of end-users and implementers
• Deliver, develop the Communication and dissemination strategy and plan for EOSC-Pillar, that describe all activities performed amongst the different tasks to support the EOSC mission and corresponding stakeholders;
• Perform all regular communication, providing visibility for stakeholders, including social media presence
• Organise and develop an online consultation platform to support the surveys
• Ensure EOSC Pillar's presence at EU related EOSC events and the specific Working Group events organised as part of the EOSC Executive Board.
• EOSC Pillar will organise 3 co-located workshops around the project's results and one final event to showcase all results - this event may be co-located with the other funded projects within this call.
• Deliver a set of at least 10 focused webinars around the survey results and use cases.
• Design and production of all materials and other dissemination materials for the EOSC.

Description of work and role of partners

WP2 - The human factor of the EOSC: Dissemination, Outreach and Community building [Months: 1-36]
TRUST-IT SRL, GARR, CNR, CINECA, INFN, FONDAZIONE CMCC, UNIVIE, CINES, CNRS, INRA, INRIA, IFREMER, INSERM, KIT, DKRZ, Fraunhofer, GFZ, UGent
This work package will design and implement an overall communication and outreach plan for the project. Liaisons responsible for helping to achieve the project strategy will be identified. The main target audiences include: the national funding agencies; national and transnational user communities; the national (non commercial) providers of technologies (i.e. e-infrastructures), as well as EOSC-related and FAIR-related initiatives.
Task 2.1: Communication, Dissemination and Outreach Strategy (TRUST-IT) This task involves the definition and implementation of a sound communication strategy. This will include organisation and participation to events, web presence, social media communication, article writing (dissemination and scientific).This task will produce a detailed Communication, Marketing and Dissemination plan and strategy, according to which a coordinated, continuous set of actions will be carried out, to ensure coverage of all stakeholders and adequate reach and visibility across Europe. Regular social media activities will also be carried out. To facilitate Member States reading, some news pieces will be delivered in national languages. Production of communication materials, such as brochures, press releases, videos, webinars, and dissemination material that can also be used for distribution at selected events to be used at the EU and national level by the project as well as by the governance bodies.
Task 2.2: Concertation with the EOSC governance and related EOSC initiatives (INFN)
The EOSC landscape is expected to be rich of actors (EU projects, communities of end-users, national initiatives, etc.) each one of them with their views of the EOSC and its objectives and opportunities. The different projects will work at implementing EOSC trans-national services, in some case with similar scopes. Some concertation is required in order to create synergies, avoid duplication of effort and the multiplication of divergent paths.
This task will ensure the drafting and maintenance of the contractual collaboration agreements with the other projects of the INFRAEOSC05 -a and -c (regional initiatives in subtopic b are dealt with in task 4.2)
The focal point of the concertation effort will be the EOSC governance. This task will liaise with the EOSC governance in order to create a two-ways communication for the involved national initiatives and services/communities to adopt EOSC standards and achievements, while proactively providing feedback about their work of test and validation The task will also liaise with the existing EOSC-related initiatives in order to create synergies, build on their results and provide them feedback; in this way we shall be able of suggesting alternative solutions and standards whenever appropriate. This will include the interaction/collaboration with the working groups of the EOSC Executive Board, in particular the "Landscape Analysis" WG, and the participation in the EOSC Stakeholder Forum. Others initiatives may include the other H2020 INFRAEOSC projects and relevant eInfra and research infrastructure projects (e.g., eInfraCentral, EOSCpilot, EOSC-Hub, FREYA, OPENAire-Advance, RDA Europe) and international coordination fora a such as RDA, CoDate, WDS. Where relevant, synergies will be identified and exploited.

Task 2.3: Community building and stakeholder engagement (UNIVIE) This task will build a multi-disciplinary community. A first set of stakeholders also influenced by the OSPP recommendations could be considered: End-users : Researchers, large enterprises, SMEs, citizen scientists; Implementers: Research Infrastructures and e-infrastructures, universities, service providers, policy making organisations, research funding and national agencies, SDOs. The task will also synergise with ongoing efforts, such as the H2020 implementation projects and coordination actions and initiatives that currently offer services at different levels (generic vs. communityspecific) for different stakeholders involved in research data management as well as related policy work in Europe. The task will also build upon the outcomes and findings of WP3 (results of surveys).

Participation per Partner

Partner number and short name	WP2 effort
1 - GARR	4.00
2 - CNR	1.00
3 - CINECA	1.00
4 - INFN	6.00
5 - FONDAZIONE CMCC	1.00
6 - UNIVIE	21.00
7 - CINES	1.00
8 - CNRS	1.00
9 - INRA	1.00
10 - INRIA	1.00
11 - IFREMER	1.00
12 - INSERM	1.00
13 - KIT	1.00
14 - DKRZ	4.00
15 - Fraunhofer	3.00
16 - GFZ	1.00
17 - UGent	8.00
18 - TRUST-IT SRL	14.00
Total	71.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D2.1	MoU signed with INFRAEOSC05 -a and -c projects	4 - INFN	Report	Public	3
D2.2	EOSC-Pillar Dissemination, Communication and	18 - TRUST-IT SRL	Report	Public	4

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
	Stakeholder Engagement Plan				
D2.3	Initial report on the status of the relationship with the EOSC Governance and the EOSC related initiatives	4 - INFN	Report	Confidential, only for members of the consortium (including the Commission Services)	8
D2.4	Second report on the status of the relationship with the EOSC Governance and the EOSC related initiatives	4 - INFN	Report	Confidential, only for members of the consortium (including the Commission Services)	24
D2.5	EOSC-Pillar Dissemination, Communication and Stakeholder Engagement Report	18 - TRUST-IT SRL	Report	Public	17
D2.6	Final report on EOSC-Pillar dissemination, Communication and Stakeholder Engagement	18 - TRUST-IT SRL	Report	Public	33

Description of deliverables

D2.1: MoU signed with INFRAEOSC05 -a and -c projects [R, PU, INFN, 3]

D2.2 EOSC-Pillar Dissemination, Communication and Stakeholder Engagement plan [R, PU, M04, Trust-IT]

D2.3: EOSC-Pillar Dissemination, Communication and Stakeholder Engagement Report [R, PU, M17, Trust-IT]

D2.4: EOSC-Pillar Dissemination, Communication and Stakeholder Engagement Report [R, PU, M33, Trust-IT]

D2.5: Initial report on the status of the relationship with the EOSC Governance and the EOSC related initiatives [R, CO, M8, INFN]

D2.6: Second report on the status of the relationship with the EOSC Governance and the EOSC related initiatives [R, CO, M24, INFN]

D2.1 : MoU signed with INFRAEOSC05 -a and -c projects [3]

MoU with the projects from the subtopics -c and -a of the INFRAEOSC05 call, to regulate contractual collaborations between the parties.

D2.2 : EOSC-Pillar Dissemination, Communication and Stakeholder Engagement Plan [4]

This deliverable will describe the dissemination, communication and stakeholders engagement and strategy plan including input from all other tasks.

D2.3 : Initial report on the status of the relationship with the EOSC Governance and the EOSC related initiatives [8]

This deliverable will describe the status of the relationship with the EOSC Governance and the existing projects/ initiatives related to the EOSC and the foreseen synergies at the beginning of the project.

D2.4 : Second report on the status of the relationship with the EOSC Governance and the EOSC related initiatives [24]

This deliverable will describe the status of the relationship with the EOSC Governance and the existing projects/ initiatives related to the EOSC and the foreseen synergies at month 24; it will provide recommendations to improve the outcome during the last year of the project.

D2.5 : EOSC-Pillar Dissemination, Communication and Stakeholder Engagement Report [17]

This deliverable will describe the dissemination, communication and stakeholders engagement and strategy plan including input from all other tasks.

D2.6 : Final report on EOSC-Pillar dissemination, Communication and Stakeholder Engagement [33]

This deliverable will describe the dissemination, communication and stakeholders engagement and strategy for the final reporting period of the project.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS3	The EOSC-Pillar web presence online	18 - TRUST-IT SRL	1	The web platform of the EOSC Pillar internal web presence and landing page is available for first release.
MS5	Organisation and roll-out of the 1st workshop	18 - TRUST-IT SRL	14	Organization of EOSC-Pillar 1st workshop open to all stakeholders involved in research data management, e.g. scientific/ user communities, research institutions, research infrastructures and eInfras, as the stakeholders listed.
MS6	Organisation and roll-out of second and third workshop	18 - TRUST-IT SRL	36	Organisation of second and third project workshops, open to all stakeholders involved in research data management, e.g. scientific/ user communities, research institutions, research infrastructures and eInfras, as the stakeholders listed.
MS7	Organisation and roll-out of first webinars	18 - TRUST-IT SRL	6	This MS marks the delivery of first project webinars, addressing all stakeholders involved in research data management, e.g. scientific/ user communities, research institutions, research infrastructures and eInfras, as the stakeholders listed.
MS8	Organisation and roll out of 10 webinars	18 - TRUST-IT SRL	36	This MS marks the delivery of all 10 the foreseen EOSC-Pillar webinars, Open to all stakeholders involved in research data management, e.g. scientific/ user communities, research

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				institutions, research infrastructures and eInfras, as the stakeholders listed.
MS9	Final result-oriented event	18 - TRUST-IT SRL	36	This MS marks thge delivery of a final event aimed at showcasing the project results, co-located with the other projects funded under INFRA EOSC subtopic b) to maximise geographical coverage.

Work package number [9]	WP3	Lead beneficiary [10]	6 - UNIVIE
Work package title	National Initiatives Survey		
Start month	1	End month	36

Objectives

The overall objective of this WP is to design and conduct a set of surveys to assess the state of the art of the various aspects of national initiatives for open research data and services in each country, and then offer full report of the results.
Aims:
• Capture the state of the art in terms of national initiatives, research communities with strong open science, data and infrastructure requirements, non-commercial computing and data services.

• The surveys and their results will assess both, level of maturity, and compliance with the latest EOSC specifications, with the aim of identifying candidate services and tools that can be offered trans-nationally.

• The analysis will offer the state of the art of existing national initiatives and services for computing and data services in the countries involved, and support their consolidation.

• Creation of full reports, which will be offered in open access modality and in compliance with all FAIR principles.

Description of work and role of partners

WP3 - National Initiatives Survey [Months: 1-36]
UNIVIE, GARR, CNR, CINECA, FONDAZIONE CMCC, CINES, CNRS, INRA, INRIA, IFREMER, INSERM, KIT, DKRZ, Fraunhofer, GFZ, UGent, TRUST-IT SRL
A solid survey is the foundation of cooperative, open research and, thus, of its reproducibility and verifiability. It also delivers the basis for steering policies and decisions on the development of the results of this project and EOSC. This work package will provide a snapshot of the state of the art of national initiatives for open research data and services in each country. This will include: readiness level, business models, existing policies and technologies and their compliance with the EOSC, involved stakeholders, and existing infrastructures and services if available, access and usage, and/or thematic initiatives and communities.
Methodology - The survey will be based on institutional, discipline-specific surveys that have already been performed at universities and research institutions in other countries or contexts. Emphasis will be placed on the creation of specially developed questionnaires which will take into consideration the common basis created by the forthcoming EOSC as well as the differences between different “systems cultures” and languages. The questionnaires will be created in English by experts, who will be hired for this purpose: this will contribute to quality assurance. The questionnaires will be programmed by means of open source software. The implementation period is described in the Milestones of WP3. The completed questionnaires, will be statistically evaluated using an open source statistical software and an open source spreadsheet. The surveys will be designed in coordination with all WP3 tasks, taking into account the ethical and legal aspects, to reduce the number of surveys and improve the response from the targeted audience. The surveys will also be co-designed with WP4 and the research communities involved in WP6. The work will use the output and expertise from relevant initiatives such as RDA, CODATA, GoFAIR. The survey will collect and process personal data limited to contact details and roles of participants (no sensible information will be required). The handling and maintenance of the personal data will be subject to the GDPR. Participants in the survey will be notified and required to approve the project’s privacy policy. The minisation principle will be applied.
Ethical and legal aspects of accessibility and reuse - Consultancy for legal and ethical will help tuning question on these topics: • GDPR: special focus will be placed on questions of informed consent, access and subsequent use, security of processing as well as the legal parameters within which partners will operate for the handling of results. • Experts working on behalf of the project or being addressed by the survey activities will be guaranteed non-disclosure. • Prevention of liability in relationship to the use of "terms of use", "use agreements", "consent forms". • Legal Problems that may arise from survey activities.
Task 3.1: Business models, SLAs and user support (CNRS) A survey of business models, procurement policies, service level agreements and user support offered across countries. This task will design and set up surveys with focus on business models, procurement policies, service level agreements and user support from domain specific, institutional, and national data and compute infrastructures in different European countries. The survey in will also identify a number of community-driven European organisations and research infrastructures, as well as relevant international consortia or organisations.

Task 3.2: Access policies, authorization and authentication models, privacy and legal (KIT) Access policies are mainly regulated by two aspects: the legal aspect and the technical aspect. The interaction between these two elements define the possibilities and the granularity of access modality. This task is responsible for conducting a survey into access policies, authorization and authentication models, terms of use, re-use rules (licences), privacy and legal issues across countries and communities. Main aspects to be considered are also roles and responsibilities, validity of the policies, review, scope and coverage, objectives ("what and how"), retention and deletion issues concerning open data/ restricted access data/ non open data, access to storage, metadata curation, exceptions.
Task 3.3: Data management and data certification (UNIVIE) This task is responsible for conducting a survey into data management practices adopted or proposed across countries and/or transnational communities, and into data certification procedures and their diffusion. This will thus be addressed to representatives of as many disciplines and infrastructure holders as necessary (e.g. arts, humanities, and sciences) including also extramural research representatives. Also legal and ethical aspects will be addressed (e.g. user agreements, handling of sensitive data, accessibility, reusability, legal ambiguities in data use), as well as practices in data certification (or further related forms of certifications), general data handling, obstacles to sharing data, adoption of related services. The findings are expected to help also in reorientation on identifying resources in this strategic area, which corresponds to the expressed needs of people in the research process and in the common frame represented by EOSC.
Task 3.4: Communities, Services and Usage (UNIVIE) This task is responsible for conducting a survey into user communities, currently available services and their usage, as well as a gap analysis of necessary developments. The adoption of related services to the topics described in addressed topics in the survey described in Task 3.3, their maintenance, desired services, preferred data services, project-specific services, desired measures, obstacles to access services, training programmes, help desk services, machine actionable services, implementation of support services, desiderata that should become anchored as a service requirement. This task will remain open for the whole project duration, in order to deal with the evolution of the usage and communities throughout the project. This will help assessing the impact of the project on the communities and usage of services.

Participation per Partner

Partner number and short name	WP3 effort
1 - GARR	2.00
2 - CNR	4.00
3 - CINECA	4.00
5 - FONDAZIONE CMCC	1.00
6 - UNIVIE	34.00
7 - CINES	1.00
8 - CNRS	7.00
9 - INRA	1.00
10 - INRIA	1.00
11 - IFREMER	1.00
12 - INSERM	1.00
13 - KIT	10.00
14 - DKRZ	1.00
15 - Fraunhofer	3.00
16 - GFZ	1.00
17 - UGent	6.00
18 - TRUST-IT SRL	4.00
Total	82.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D3.1	Summary Report on the National Initiatives survey	6 - UNIVIE	Report	Public	12
D3.2	Final Report on the communities ad usage of services	6 - UNIVIE	Report	Public	36

Description of deliverables

D3.1: Summary report of the National Initiatives surveys [UNIVIE, KIT,CNRS; M12]

D3.2: Final report on the communities and usage of services [UNIVIE, KIT,CNRS; M36]

D3.1 : Summary Report on the National Initiatives survey [12]

Report on the analysis, discussion on the methods, tools and results, conclusions and recommendations.

D3.2 : Final Report on the communities ad usage of services [36]

Update report on the survey results at the end of the project, including the usage of services proposed by the project and related, with a discussion on the project impact on these findings. The document will analyse the changes registered throughout the project's lifetime, and include recommendations useful for the long-term sustainability of the proeject outcomes.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS10	Definition of survey targets	6 - UNIVIE	1	This deliverable marks the definition of survey targets, which will be co-designed with WP4 and the research communities involved in WP6, and provide input tho the rest of WP3 activity.
MS11	Legal consultancy in legal and ethical issues completed	6 - UNIVIE	3	This MS marks the completion of the consultancy for legal and ethical issues, including expert advice on GDPR, access and subsequent use, security of processing, non-disclosure, as well as prevention of liability in relationship with survey activities.
MS12	Determination and definition of distribution mechanism	13 - KIT	3	This MS marks the definition of the mechanisms and procedures to be used to distribute the survey to target respondents.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS13	Setup of the Technical infrastructure for the realisation of the surveys	6 - UNIVIE	3	This milestone marks the implementation of the technical infrastructure (i.e. forms, database etc) that will be used to deliver the survey.
MS14	Definition of questions, definition of questionnaires	6 - UNIVIE	4	This MS marks the definition of the questionnaires to be proposed to survey respondents.
MS15	Survey roll-out	6 - UNIVIE	6	This deadline marks the completion of the survey in the different countries.
MS16	Analysis and project internal presentation of results, findings, outcomes	6 - UNIVIE	7	This MS marks the completion of the survey data analysis and the preliminary presentation of the findings to the project participants.
MS17	Visualisation of results for external audiences	6 - UNIVIE	8	Creation and publication of graphics and other data visualisations to present data to external audiences.

Work package number [9]	WP4	Lead beneficiary [10]	1 - GARR
Work package title	From National Initiatives to trans-national services		
Start month	1	End month	36

Objectives

The specific objectives for this WP are:

• Build upon the findings of the set of studies performed by WP3 in order to highlight common policies and approaches adopted at the national level, that can be leveraged in coordinating the existing and forming national initiatives in the region;

• Highlight critical issues in the legal and policy framework of the countries that can hamper collaboration and interoperability, and suggest means to address these challenges;

• Explore cross-country federation opportunities and procedures, and define technical means in order to translate them into practice;

• Support the creation and strengthening of full-fledged national initiatives across the involved countries and improve their level of representativeness;

• Provide a lean policy framework for the longer-term coordination in the federation;

• Define common tools and frameworks to describe, enroll, and support national service/data providers;

• Pose the founding stone of a National Service Registry per country, to be operated by the National Initiatives as a single entry point for national and trans-national non-commercial service providers (i.e. typically services provided by the research communities themselves), and a bridge towards the EOSC;

• Coordinate with initiatives similar to EOSC-Pillar that address different European regions (e.g. Northern Europe, the Balkans), in order to seamlessly inter-federate and ensure the highest levels of interoperability for initiatives and transnational services;

• Study one or more possible business model(s) for the National Initiatives and their services, building on the results of the survey and in collaboration with the use cases (WP6) and pilots (WP7 – T7.4), where besides the technical aspects, cost studies will be performed, to obtain realistic information on implementation and operation costs of the services, as well as their scalability and sustainability in the longer term.

Description of work and role of partners

WP4 - From National Initiatives to trans-national services [Months: 1-36]
GARR, CNR, CINECA, INFN, FONDAZIONE CMCC, UNIVIE, CINES, CNRS, INRA, INRIA, IFREMER, INSERM, KIT, DKRZ, Fraunhofer, GFZ, UGent, TRUST-IT SRL
This work package will create a policy and legal framework for the coordination of National initiatives, that can then be extended to other countries. This work package will define a national service registry blueprint, to be implemented in each country and maintained by the national initiative. Registries will be federated with each other and fully interoperable with the EOSC. It is the ambition of this project that, in the longer term, the national registries will incorporate the large majority if not the totality of relevant scientific data services in each country. The starting point will be provided by the findings of the surveys carried out in WP3, which will be used to highlight common policies and approaches that can be leveraged to design a National Initiative model that can be implemented and adapted to the specific ecosystem of each country and its scientific communities. The National Data and Cloud initiatives in the involved countries are now moving their first steps and coordinating from these early stages, meaning that they are in a good position to incorporate useful changes that will help achieving interoperability by design.
The first step will be to coordinate these evolving National Initiatives. The reference model we are assuming for this coordination action is the federation, as the most suitable to maintain the maximum level of freedom for the national initiatives while allowing for interoperability and a lightweight model of shared management for the cross-national services. Federating the National Initiatives will involve not only the policy and legal framework, but also the technical aspects that will be addressed in detail in WP5 (data layer) and WP7 (infrastructure layer) and demonstrated by the use cases collected in WP6. The Federation is intended to become an interlocutor for the EOSC governance, in order not only to receive information on technical and policies evolutions in the EOSC, but also to convey the feedback of one of its key stakeholders: the national initiatives, and through them, the national research communities.

As EOSC-Pillar's long-term ambition is to unify under this federated model the National initiatives of the whole EU, the second step will be to establish and maintain high levels of concertation with similar regional initiatives, namely the other projects approved under this topic.
The third step will be to consolidate the national initiatives. This will include several actions to guide them towards:
• assuming more stable forms (e.g. from MoUs or SIGs to JRUs or even legal entities);
• improving/detailing internal policies and streamlining processes;
• improving their representativeness and sustainability perspectives through the involvement of new national stakeholders;
• identifying and adopting interoperable solutions and technologies;
• obtaining optimal levels of compliance with the EOSC standards.
The fourth step will be to define a common set of procedures and technical framework to enroll and describe services at the national level and facilitate the passage from local/national service dedicated to a specific community to trans-national service, extensible to new communities.
The last step will be to identify a viable business model to help ensuring the longer term sustainability of the National Initiatives and of their inter-federation. A special attention will be dedicated to the implications, in terms of business model, of trans-national services: while a national service could have a relatively simple business model, and well identified funding sources (for instance, a community can selffund the services they need, or a public funding agency can provide centralised funding to support services that are considered strategic for the research community), the case where a national initiative in one country supports a service that is used by other communities in other countries is far from being defined. Several possibilities should be studied here, including bi-lateral agreements between two or more national initiatives for the reciprocal exchange of services, outsourcing models for some services, de-centralised funding where communities pay for the resources they are consuming, etc. The WP includes 5 tasks.
Task 4.1: Policy and legal framework (KIT)
This task will set up a policy and legal framework by building upon the existing national policies and their similarities and propose a governance model. To this end, the following actions will be performed:
• ascertain requirements for open data, data protection and cross border data access;
• determine roles and stakeholders of infrastructures, services and security management, different funding models and funding agencies;
• establish boundaries for the legal interoperability of data;
• propose and setup an initial services management framework (with T4.4);
• indicate limitations and common practice of handling scientific data in national contexts
• participate in and shape the communication structure with the stakeholders (with T4.2 and WP2).
• propose a governance model that recognises all stakeholders and activities of the EOSC-Pillar (T4.2 and T4.3).
Task 4.2: Coordination with regional initiatives (GARR) This task will be responsible for establishing and maintaining the contractual collaboration agreements with the other projects of the INFRAEOSC05-b call and their proper execution. EOSC-Pillar will liaise with other regional initiatives in order to achieve high levels of coordination with them, avoid duplications and maximise synergy and syncronisation. The longer-term objective is to maximise the levels of interoperability between National Initiatives belonging to different regions and work towards a seamless Europe-wide interfederation.
Relevant actions will include:
• establishing and maintaining bilateral relations with other regional initiatives (to be regulated through MoUs or similar broad agreements)
• creating a coordination board that will meet periodically, to maintain multi-lateral coordination, exchange experiences, success stories and discuss critical issues and challenges; • collaborating with the relevant workpackages and tasks in the project to implement relevant coordination board resolutions;
• identifying common challenges to tackle and appointing cross-project task forces to address them, following their works and monitoring their results.
Task 4.3: Consolidation of national initiatives (INFN) This task will favor the constitution of new national initiatives and will provide support to the advancement, the enhancement and the readiness of the already started ones. The main actions will include:
• the involvement of a significant number of relevant stakeholders in each initiative;
• help in the definition of the proper policies and governance including their formalization, for example throughout MoU, taking into account the characteristics and the peculiarities of each nation;
• support to guarantee the proper representativeness to each initiative in the national and European panorama;
• facilitation to the access to funding opportunities at national and European level.
Task 4.4: National service registry (CNR) Building on the findings of the initial survey, this task will create and implement the model for a national Registry of services to be operated by the national initiatives. National registries will be interoperable with each other and with EOSC. As defined by ITIL (set of detailed practices for IT service

management - ITSM), the Service Registry contains information about service features, contact points, ordering, and request processes. It allows to easily discover and access the service card reporting for each service at least the service name, the service description, the minimum guaranteed availability, quality of service, the technological readiness level (TRL), the service owner, and the service terms of use. A service categorization will allow to group together similar services, thus promoting their discoverability and comparison. Service Level Agreements (SLAs) will help service providers to set expectations for their service requestors.
Task 4.5: Business model (CNRS) This task will build on the findings of the surveys (WP3 task 3.1) and of the science-driven use cases (WP6) and pilots (WP7 task 7.4). It will work in close coordination with the tasks 4.2 (regional initiatives) and 4.3 (national initiatives), together with WP2 for strategic linking with the EOSC governance. It will rely also on the results of the EOSC pilot project and integrate the experience of the EOSC-hub activities. This task will help in proposing different possible business models supporting transdisciplinary data, compute and open science services and infrastructures at the national and transnational levels across the EOSC federation, as well as potentially services from private providers when identified in the project. This task will also
• investigate/survey beacons of good practice from non European countries (e.g. USA, Japan, Australian) as well as from a number of science-driven international initiatives or organisations.
• iterate during the development of the possible business models within the EOSC Pillar project to get feedback and input
• organise one or two workshops gathering national stakeholders and relevant funding agencies, together with relevant international organisations and initiatives, including RDA and CODATA, to get their feedback and input during this process. This task will be carried out in close synergy with the user and science communities in WP6 and WP7 to insure these business models co-evolve with the research practices and the organisation of the communities at the national, European and international levels.

Participation per Partner

Partner number and short name	WP4 effort
1 - GARR	16.00
2 - CNR	9.00
3 - CINECA	1.00
4 - INFN	9.00
5 - FONDAZIONE CMCC	1.00
6 - UNIVIE	6.00
7 - CINES	6.00
8 - CNRS	9.00
9 - INRA	1.00
10 - INRIA	1.00
11 - IFREMER	6.00
12 - INSERM	1.00
13 - KIT	11.00
14 - DKRZ	1.00
15 - Fraunhofer	6.00
16 - GFZ	1.00
17 - UGent	1.00
18 - TRUST-IT SRL	6.00
Total	92.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D4.1	Legal and Policy framework and federation blueprint	13 - KIT	Report	Public	16
D4.2	Signed MoU with Regional initiatives	1 - GARR	Other	Public	6
D4.3	Roadmap for consolidating National Initiatives	4 - INFN	Report	Public	24
D4.4	National Service Registry Prototype	2 - CNR	Demonstrator	Public	24
D4.5	Business model and sustainability study	8 - CNRS	Report	Public	36

Description of deliverables

D4.1: Legal and Policy framework and federation blueprint [R, PU, M16, KIT]

D4.2: Template MoU with regional initiatives [O, PU, M03, GARR]

D4.3: Roadmap for consolidating National Initiatives [R, PU, M24, INFN]

D4.4: National Services Registry prototype [P, PU, M24, CNR]

D4.5: Business model and sustainability study [R, PU, 36, CNRS]

D4.1 : Legal and Policy framework and federation blueprint [16]

A study of the legal and policy state of the art in the involved countries, highlighting commonalities to be leveraged and challenges to be tackled, and proposing a blueprint for the federation's rules and procedures.

D4.2 : Signed MoU with Regional initiatives [6]

A multilateral MoU, or a set of bilateral ones to regulate the contractual collaboration with complementary grants 857652 EOSC-Nordic, 857647 EOSC-synergy, 857641 ExPaNDS, 857645 NI4OS-Europe. The memorandi of understanding will detail how synergies will be identified and be fully exploited and how work will be as much as possible complementary. The MoU will detail how the projects will continue to interact and exchange info throughout the duration of the projects. Collaboration will include aligning (and possibly co-organising) certain activities of dissemination (also towards the EOSC Executive Board, including its working groups) and outreach.

D4.3 : Roadmap for consolidating National Initiatives [24]

Guidelines and actions to be performed to consolidate National Initiatives.

D4.4 : National Service Registry Prototype [24]

A working prototype of a National Services Registry, including a complete set of access policies, definition of roles and profiles, registration rules, user support procedures etc.

D4.5 : Business model and sustainability study [36]

A study into possible business models for national initiatives and their national/transnational services, with an outlook on their longer-term sustainability.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS18	Requirements of open data, services and security management	13 - KIT	6	This MS marks the definition of requirements for open data, data protection and cross border data access and indicate limitations and common practice of handling scientific data in national contexts. This is a key step towards the policy and legal framework.
MS19	Setup of the regional coordination board	1 - GARR	6	This MS wikk mark the creation of the Regional Coordination Board and its first meeting.
MS20	Legal framework for trans-national data services	13 - KIT	14	The MS marks the definition of the legal framework to be used to implement and use trans-national open data services. It is part of D 4.1
MS21	Proposed service management structure for EOSC-Pillar	13 - KIT	14	The MS marks the presentation of a proposal for the management structure of the project services, that should be scalable enough to encompass other ones too. The proposal will be discussed by partners and the results will become part of D4.1
MS22	Business model round of feedback from National Initiatives	8 - CNRS	16	This MS marks the collection of input an feedback on business models from National Initiatves participants (through a workshop or a virtual meeting)
MS23	Business model workshop	8 - CNRS	24	A workshop to gather further feedback on the proposed business models, open to national stakeholders and relevant funding agencies, together with relevant international organisations and initiatives, including RDA and CODATA.

Work package number [9]	WP5	Lead beneficiary [10]	15 - Fraunhofer
Work package title	The Data layer: establishing FAIR data services at the national and transnational level		
Start month	1	End month	36

Objectives

The primary goal of this work package is to progressively reduce technical, societal and organisational barriers to ensure findability, accessibility, interoperability and re-use (FAIR) of research data considered in EOSC-Pillar.

Such a goal will be achieved by addressing the following objectives:
a) developing and operating an innovative set of services enacting the building and exploitation of a FAIR-friendly data space out of the content of various existing and future research data repositories,
b) setting up support and training activities facilitating the diffusion and adoption of mainstream standards and approaches for research data management,
c) boost, collate and agglomerate domain specific ontologies and related metadata as main basis for cross domain interoperability enabling efficient and active FAIR data exploitation.

Description of work and role of partners

WP5 - The Data layer: establishing FAIR data services at the national and transnational level [Months: 1-36]
Fraunhofer, CNR, INFN, FONDAZIONE CMCC, UNIVIE, CINES, CNRS, INRA, IFREMER, INSERM, KIT, DKRZ, UGent, TRUST-IT SRL
Building on activities performed by other past and ongoing projects and initiatives (e.g. INFRAEOSC -052018-2019 subtopic c, OpenAIRE-Advance, RDA, GO-FAIR, MarketPlace, FORCE, Oyster, APACHE, etc) this Work Package aims at creating the conditions for an effective exploitation and reuse of data across Member States.By facilitating and promoting the adoption of standards and harmonised solutions for FAIR data management and certification schemes for data repositories, this WP will make available a rich asset of data resources, with clear access and usage policies, to the researchers in the Member States represented in the project, and, more in general, to those accessing EOSC. It aims to develop tools to allow scientists to 1) find the information and data and 2) decipher the information within the data sets based on tools and standards developed by each domain. Such a capability requires using ontologies as common basis for metadata across domains, following the recommendations of the RDA as well as the current developments in various fields, e.g. materials science (see the establishment of a European Materials Ontology and Interest Group, e.g., at https://emmc.info/Intop2018). The aim is not to create a top-down common ontology, but to increase the collaboration of all ontologies created by relevant scientific fields and agglomerating them into one.
Task 5.1 Common tools and mechanisms for FAIR research data providers (CINES) By taking into account (a) the survey results emerging from Task 3.3, (b) the needs emerging from WP6 use cases, and (c) the initiatives and guidelines proposed by INFRAEOSC-05-2018-2019 subtopic c), this task will identify, develop and operate a number of services facilitating the creation of a federated FAIR research data space out of the contents of existing and future research data repositories and initiatives complying with the EOSC protocols and standards. Such services will be developed by building upon and complementing existing ones in EOSC-hub thanks to the projects resulting from EINFRA-12-2017 topic (e.g., EOSC-hub, OpenAIRE, FREYA) and other community based efforts. Envisaged services include (a) a “data repository compliance service” which provides the data provider (usually a data repository owner) with a report on the degree of accordance among the data and some defined policies and means to actual make the registry compliant with the EOSC; (b) a “data policy registry” which provides the data provider with machine actionable representations of data related policies (e.g. metadata schemas based on ontologies, controlled vocabularies, licences); (c) a “data enricher” which provides the data provider with additional information on the datasets he/she provided resulting from the overall data space (e.g. linking with other datasets, minting of DOIs). The developed services will be accessible via APIs in order to enable their exploitation for building more complex services. The description of these services will be included in the Service Registry. As a consequence of this action these services will be also accessible through the EOSC Portal catalogue.
Task 5.2 Common tools and mechanisms for FAIR research data consumers (CNR) This task is a mirror of Task 5.1 focusing instead of the end users, on the consumers of data. This targets access, and managing users own data sets, ownership, searching and re-use of existing data. By taking into account (a) the results emerging from Task 3.3, (b) the needs emerging from WP6 use cases, and (c) the initiatives and guidelines proposed by INFRAEOSC-05c, this task will identify, develop and operate a number of innovative services facilitating a FAIR-friendly use of the data space resulting from Task 5.1 activity. Such services will be developed by building upon and complementing the services currently

existing in EOSC-hub thanks to the projects resulting from EINFRA-12-2017 topic. Envisaged services include (a) a “community driven data catalogue” offering a unifying view of a community defined data space (resulting from a subset of the overall data space built by this proposal); (b) an “overall data catalogue” offering a unifying view of all the datasets federated by data providers partaking to this proposal. The developed services will be accessible also via APIs in order to enable their exploitation for building more complex services. The description of these services will be included in the Service Registry. As a consequence of this action these services will be also accessible through the EOSC Portal catalogue. Various technologies and related expertise from previous EU projects are exploited,including OpenAIRE , D4Science Infrastructure,EuDAT, EMMC-CSA, MarketPlace and FORCE.
Task 5.3 Helpdesk and documentation for promoting FAIR practices and support to FAIR-oriented data stewardship (UGent) This task will focus on research data management standards and on creating awareness in different data stewardship approaches in order to harmonize support offered to researchers. The first step will be to start developing a set of (inter)national guidelines and documentation on FAIR data stewardship addressed primarily to data steward teams and decision makers, and to support their dissemination. This task will collect different approaches on data stewardship, exchange practices and consult different stakeholders to create awareness on the available approaches and related issues. Secondly, documentation and online help will be offered to mainstream standards of research data management. While the call scope is not to create generic FAIR awareness campaign, as several funded projects and bottom-up initiatives cover this, we shall collaborate and coordinate and wherever possible liaise with existing FAIR-related initiatives (e.g. OpenAIRE, EuDAT, EOS-HUB, RDA, FAIRsFAIR, Go FAIR etc, and also community based actions as EMMC-CSA and MarketPlace in materials science community) and explore other collaboration opportunities to maximise the uptake of FAIR principles across the research communities (e.g. joint campaigns, co-organisation of events etc). The entire documentation will be uploaded to Zenodo and linked to related repositories (including on the FAIR UGent repository). Online targeted help will be offered This task interacts with Task 5.4 and will provide material for the training as well, while Task 5.4 will provide feedback on the needs for documentation.
Task 5.4 Training modules on FAIR-oriented research data management tools and solutions. (CNR) Training events (on site) and webinars and other e-learning means will be provided, in coordination with existing initiatives (OpenAIRE, RDA, GO-FAIR, MarketPlace and the projects resulting from INFRAEOSC-05-2018-2019 subtopic c), etc). These will be addressed to both data providers and consumers (a two-sided approach). The training will be performed on regional/ national levels and cover all needed components and services from the preparation of the data, metadata, ontologies, to the actual tools, legal frameworks and how to use them so that various end users can make their data FAIR. Results from other WP are needed, especially WP3, WP4, and other tasks in WP5. Training will be, when possible, performed in respective national languages to allow better understanding and foster participation. The use of specific tools (such as Mentimeter for active online workshop participation, and other available tools) will allow the training sessions to be highly interactive and to better involve the audience in the discussion. The training modules concentrate on the following themes (project results will be exposed to highlight the national/regional initiatives and situations): • Demonstrate added value for end users as a result of sharing, “how can I benefit from sharing?” • Best Practices and successful stories collected over the duration of the project as examples (e.g., from WP6) • Services and Tools • Data Stewardship (advice on, and practice of, responsible curation and handling of data, e.g., whether or not to archive raw data and when) • FAIR Research Data Management (RDM) • Policies (access rights, embargo period, mandating sharing in public funded projects, etc) • Legal aspects connected to IPR, GDPR (such as liability, data protection, privacy, etc.) • Ontologies for specific domains to boost interoperability between repositories. During the project any needs for additional aspects will be considered. Training activities will include general sessions embracing more than one theme and giving an overall flavour of all the topics, whereas specific sessions will be organized to deeply concentrate on a single argument or will be seized for a specific community or stakeholder group. The wide spectrum of training activities will be covered from short session (1-3 hours), short courses (one day) to long courses (2 days to one week). Short sessions will be provided particularly in webinar form whereas short and long courses will be carried out by physical meetings.
Task 5.5 Collating common metadata and ontologies: cross domain interoperability foundations. (Fraunhofer) The aim of this task is to collate all relevant ontologies available in all subdomains considered in this project into a common ontology that enables the development of common cross domain interoperability of all data repositories that will be available for other tasks to build upon. Tools to manage and handle the interoperability and integrate it as metadata in all services provided is developed. As a first step, the existing infrastructure of the EOSC hub is used to curate all ontologies. This includes contacting and interacting with all repository owners from all relevant domains, materials, bio, heritage conservation, geo, etc. In the second stage, a repository to collate all ontologies is created and published. The following communities are targeted initially: Materials and nano science, climatology, Health and geology.

Participation per Partner

Partner number and short name	WP5 effort
2 - CNR	21.00
4 - INFN	4.00
5 - FONDAZIONE CMCC	3.00
6 - UNIVIE	7.00
7 - CINES	14.00
8 - CNRS	12.00
9 - INRA	2.00
11 - IFREMER	2.00
12 - INSERM	4.00
13 - KIT	15.00
14 - DKRZ	2.00
15 - Fraunhofer	29.00
17 - UGent	13.00
18 - TRUST-IT SRL	2.00
Total	130.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D5.1	FAIR Research Data Management Tool set - first release	7 - CINES	Other	Public	12
D5.2	FAIR research Data Management Workbench Operation Report	2 - CNR	ORDP: Open Research Data Pilot	Public	18
D5.3	Training Plan	17 - UGent	Report	Public	4
D5.4	FAIR-oriented Research Data Management: Support, Training and Assessment Activity Report	17 - UGent	Report	Public	18
D5.5	Collated set of ontologies as foundation for interoperability	15 - Fraunhofer	ORDP: Open Research Data Pilot	Public	36
D5.6	Fair Research Data Management Tool set update	7 - CINES	Other	Public	24
D5.7	FAIR Reseacrh Data Management Workbench Report update	2 - CNR	ORDP: Open Research Data Pilot	Public	36

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D5.8	Training plan update	17 - UGent	Report	Public	20
D5.9	FAIR-oriented research data management: Support, Training and Assessment Activity Report update	17 - UGent	Report	Public	36

Description of deliverables

D5.1: FAIR Research Data Management Tool set [O, PU, M12, first release, CINES]

D5.2: FAIR Research Data Management Workbench Operation Report [ORDP, PU, M18, CNR]

D5.3: Training plan [R, PU, M4, first release, UGent]

D5.4: FAIR-oriented research data management: Support, Training and Assessment Activity Report [R, PU, M18, first release UGent, CNR]

D5.5: Collated set of ontologies as foundation for interoperability [ORDP, PU, M36; Fraunhofer]

D5.6: FAIR Research Data Management Tool set update [OTHER, PU, M24, second release, CINES]

D5.7: FAIR Research Data Management Workbench Operation Report update [ORDP, PU, M36, CNR]

D5.8: Training Plan update [R, PU, M20, second release, UGent]

D5.9: FAIR-oriented research data management: Support, Training and Assessment Activity Report update [R, PU, M36, second release, UGent, CNR]

D5.1 : FAIR Research Data Management Tool set - first release [12]

The bundle of service instance(s) resulting from T5.1 and T5.2 activities is released. This Toolset will be updated at least every 6 months after this initial release, and will have at least another major release at PM 24.

D5.2 : FAIR research Data Management Workbench Operation Report [18]

This deliverable documents the activities and results (e.g. indicators on integrated data providers, and datasets, indicators on datasets accesses) of the operation of the toolset service and EOSC-Pillar data space.

D5.3 : Training Plan [4]

This deliverable presents the overall project training plan, which will include 1. FAIR-related awareness; 2. training on project-related tools and procedures, addressing both project participants and the prospective user community; 3. use-cases related training addressing specific communities.

D5.4 : FAIR-oriented Research Data Management: Support, Training and Assessment Activity Report [18]

This deliverable reports on the activities and results of Tasks 5.3, 5.4

D5.5 : Collated set of ontologies as foundation for interoperability [36]

A repository collating the relevant ontologies available in all subdomains considered in this project into a common ontology that enables the development of common cross domain interoperability of all data repositories that will be available for other tasks to build upon.

D5.6 : Fair Research Data Management Tool set update [24]

Updated release of D15.

D5.7 : FAIR Reseacrh Data Management Workbench Report update [36]

Updated release of D16.

D5.8 : Training plan update [20]

Updated training plan (D17)

D5.9 : FAIR-oriented research data management: Support, Training and Assessment Activity Report update [36]
This deliverable reports on the activities and results of Tasks 5.3, 5.4 at the end of the project.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification

Work package number [9]	WP6	Lead beneficiary [10]	8 - CNRS
Work package title	EOSC in action: Use cases and community-driven pilots		
Start month	1	End month	36

Objectives

The need for "data FAIRisation" is important, diversified and specific for each community (nanotechnologies, environment -ocean, atmosphere, continental surfaces, health, humanities, biodiversity, solid earth). They are more or less structured and have developed tools and services adapted to their specific needs and constraints. The objective is to prepare scientific communities to be involved in EOSC.

Specific aims:
• analyse specificities of the scientific communities but also cross-cutting needs to avoid silos,

• develop cross-cutting approaches, platforms and demonstrators to better integrated different data from all components and domains and facilitate access, integrated and intelligent treatments and diffusion,

• identify from use-cases important tools and services, analyse their genericity and scalability,

• identify gaps and needs that can be fulfilled by the futureEOSC This WP is based on a bottom-up approach from use-cases covering several themes. Each use-case will analyse different tools and services for FAIRisation of data and services and type of governance according to the community needs. Pilot use-case analysis must take into account the entire chain of data use, from acquisition to scientific exploitation. Some use-cases deal (e.g. Earth System) with the topic of decompartmentalization of data independently of the scientific themes and organizations. In this context, it is important to consider a number of factors:

• the differences between data production from large centralized instruments (observation or experimental platform) or from decentralized networks (soil, sea, air, space),

• the different practices and organizations around data archiving and curation platforms (granularity, generic or community platforms, national or international, standardization of data formats, data/metadata exchange (including metadata) and services, practical certification),

• specific needs for quality control chains and data analysis (in particular dependence on virtualization The diversity of the data concerned : qualitative (e.g. humanities for example) and quantitative. It show a great diversity of structured or unstructured data, such as scalars, events, time series, images, space data, in-situ data geolocalized or not. These objects have very differentstructures and elementary volumes (e.g. events versus cube) that impact the services (referencing, versioning, moving data, ..).

Description of work and role of partners

WP6 - EOSC in action: Use cases and community-driven pilots [Months: 1-36]
CNRS, GARR, CNR, INFN, FONDAZIONE CMCC, UNIVIE, CINES, INRA, INRIA, IFREMER, INSERM, KIT, DKRZ, Fraunhofer, GFZ
This work package collects use cases based on the requirements of scientific communities. Pilots will run to validate the proposed solutions, which will be trans-national by design and general enough to be extended to other communities with minimal changes. The pilots will also include resources and costs study, in order to understand their feasibility and propose a viable business model for the resulting services. The selected use cases are connected to real scientific production and intend to improve the usage of data in a FAIR perspective.
Task 6.1: Use case 1. Defining Procedures/Service to enforce data provenance for thematic communities and beyond (CNR)
Provenance management is a key component in order to guarantee scientific data discovery, reproducibility and results interpretation. Provenance management should be able to define a set of metadata able to capture the derivation history of any stored data, including the original raw data sources, the intermediate data products and the steps that were applied to produce them (e.g., the analyses that were performed and the software version that was adopted). To enable a well-defined data provenance in the scientific experiments workflow, from data production to data usage, this task will:
Task6.1.1: elaborate cross-domain, FAIR-oriented procedures and recommendations to enforce data provenance. The set of procedures will be developed by taking into account needs coming from various communities such as

(i) nanoscience/soft matter domain which have started recognizing the need of such a definition of procedures, metadata schemas and services within some important European project, NFFAEUROPE and EUSMI where CNR-IOM is involved,
(ii) material science domain where the Fraunhofer team is already involved in most of existing actions within thematic NMBP actions (FORCE, EMMCCSA, MarketPlace, Oyster, to mention a few) and
(iii) climate change domain in the field of environmental science - where provenance management plays a key role both for numerical end-to-end simulations at the data center level as well as in the inner data analysis workflow (e.g. micro-provenance). Although this activity will start from the above specific communities, the output of this task will aim at addressing the user community at large.
Task6.1.2: implement the procedures and recommendations defined in Task6.1.1, by developing the proper adaptations/ extensions on top of existing scientific data services, made available by the participating partners, at national level. The implementation phase will cover several use cases, in particular and starting from the above communities. Expected outputs: the main output of this task will be twofold (i) the general guidelines and procedures regarding data provenance enforcement for scientific communities and (ii) some selected real pilot cases implementing them into the wider research data cycle. Recommendations and procedures in Task6.1.1 as well as design and implementation of open source software will contribute to the overall output of WP6.
Task 6.2: Use case 2. Agile FAIR data for environment and earth system communities (ocean, atmosphere, continental surfaces, solid earth, …) (IFREMER)
Earth Observation involves different domains: Continents, Ocean, Atmosphere, Solid Earth, Biosphere…. As a consequence, observations are conducted by a large amount of organisations including Spatial Agencies, Thematic Institutes, Research Institutes. Up to now, data management and valorisation are domain-dependant, relying on several European Research Infrastructures such as Actris and Iagos for the Atmosphere domain, SeaDataNet, EuroArgo, Jerico, for the Oceans, EPOS for the Solid Earth. In addition, data resulting from in-situ observations and remote sensing data (satellites) are often managed and preserved separately, involving difficulties for intercalibrations, intercomparisons and integrated uses. As an example, for the Oceans, two main data managements systems are now in place: SeaDataNet for insitu Observation Data and the DIAS (Data Information and Access Service) for Satellite data and Operational Oceanography (satellite data, model outputs…). However, these domains are interconnected (ocean/atmosphere, hydrosphere as a whole, ocean/solid earth) and these interfaces are of primary importance for several studies with large societal impacts: climate change, agriculture and food, human safety near the coast. Some initiatives have been launched tofacilitate comprehensive studies such as, e.g. the Community of Environmental Research Infrastructure (ENVRI). Taking in account these prior European initiatives, and relying on existing efforts at national level, like the new French Research Infrastructure IR-ST, this task will build a representative set of FAIR-oriented use cases based on Earth System communities engagement, using the following Agile-like process within a continuous dialog with involved parties (users and service providers):
1) stating an integrated view of the present situation in terms of FAIRization processes and dataflows,
2) recording specifications of users' needs, insisting on the needs about cross-domains harmonization and in-situ/spatial (e.g. EOSC/DIAS interfaces) integration,
3) providing an analysis of the possible improvements and impacts on existing infrastructures: e.g. harmonization of data catalogues and reference vocabularies, formats, access protocols, geographical metadata and data standards, use of cloud facilities for storage and processing, licences and citations (DOI),
4) a costs study (costs to deploy the process + operational costs)
5) the definition of the roles for the involved actors (i.e. labs, poles, research infrastructure, national data centers, etc.) and training. The process will be validated and extended to other entities/communities.
Task 6.3: Use case 3. Integration of data repositories into EOSC based on communities approaches (INRA)
Data repositories are a key component of the data governance as they are used by research organizations or scientific communities for data cataloging or depositing services with associated policies. Based on open source platforms such as dataverse they are easy to deploy and although they increase the accessibility and sharing of data, they also contribute, to some extent, to the fragmentation of the landscape. Thus, the integration of such repositories in a global interoperability framework is of critical importance in implementing the EOSC vision to build a federated data ecosystem. This was particularly highlighted by the e-ROSA H2020 project while mapping the agri-food data ecosystem to build a roadmap for an e-infrastructure for open science in agriculture and the MarketPlace project (https://themarketplace-project.eu) which targets federated concerted access to materials data repositories. Here special attention is needed for the specifications of deep metadata, i.e., that is in the data set it self (describing the content). Based on existing community data repositories, and specially agri-food, material sciences, heritage data, source code etc, this use case will explore the current limitations and provide concrete recommendations and tools to address interoperability of data repositories within the future EOSC infrastructure. The main goal is to connect and align dataverse and other repositories with EOSC specifications and to reduce fragmentation of the landscape (link with task 7.1 and 7.4)
• Develop a connector module between Dataverse and other repositories and the EOSC ETDR service

• Develop APIs between Dataverse and the EOSC/EUDAT B2SAFE service
• Perform the harvesting of dataverse by B2FIND.
• Integrate dataverse and D4Science VRE
• integrate data (dataverse) in Containers such as Docker
• Connect dataverse with (national) DMP services (task 5.2) This task will be based on existing dataverse implementations that will be tested and linked to scientific domains (Agri-food or Health sciences), but the developed tools and connectors will be "discipline or community agnostic" and will enrich the whole EOSC services ecosystem.
Task 6.4: Use case 4. Software source code preservation, reference and access (INRIA)
Leveraging the experience of Software Heritage, this task will design and pilot a solution for the preservation of massive collections of softwares source code (billions of files with links to publications) into EOSC eTDR service (European Trusted Digital Repository).
More specifically, this task will: • develop a deposit API for research software archival
• develop an access API for retrieval of archived software
• standardise a persistent identifier schema (PID) for referencing billions of archived software artifacts
• integrate the above APIs and Services with EOSC eTDR
• develop a pilot to fully host the software archive onto existing EOSC infrastructure

Task 6.5: Use case 5. FAIR principles in data life-cycles for Humanities (CNRS) The aim of this task is to build specific use cases based on SSH communities engagement. In order to do that, we will rely on consortia funded by Huma-Num and other partners from DARIAH (e.g. Italy and Germany). Another focus will be done on the link between data and publication. HAL, the french national open archive created by CCSD, provide a specific portal for SSH communities to deposit and deliver their publications in open access (https://halshs.archives-ouvertes.fr/). CCSD will work with Huma-Num to link publications on HAL to research data in data repositories, specially Nakala (https://www.nakala.fr/), the data repository from Huma-Num. This case will be a model for linking with other data repositories used in SSH communities. Finally the idea is to liaise with CO-OPERAS GO-FAIR Implementation Network to ensure that SSH needs are correctly taken in account in order to facilitate the integration of SSH in EOSC.
Task 6.6: Use case 6. Exploring reference data through existing computing services for the bioinformatics community (INSERM)
The aim of this use case will be to explore the possible interactions between already available Galaxy computing services and data repositories, in order to build an integrated and interoperable service for ELIXIR and the wider Life Science user community as a whole. The task will build on top of existing national services made available in France and Italy by participating partners, and which are described in the relevant sections of each partner description in section 4. It will aim at fulfilling the following objectives:
• Allow frictionless access to external data sources from different Galaxy deployments
• Facilitate the deployment of Galaxy instances close to the data
• Provide coherency between different existing Galaxy deployments
• Ensure health data security requirements are met throughout the process
The task will be split into the following subtasks:
T6.6.1 - Share and harmonize Galaxy deployment practices. The aim of this task is to develop and share experience on ongoing deployment solutions. Foreseen aspects include Ansible roles, Git repositories and Continuous Integration principles. The task will identify and develop best practices and improve them to cover the needs and the specificity of the different underlying infrastructures.
T6.6.2 - Facilitation of "Close to the data" computing. Connecting computing services to large datasets highlights the need to have strong connections between computing facilities and data repositories. This task will look at limitations of currently available services, and identify missing services for making the "computing close to the data" model feasible.
T6.6.3 - Data structure and formats. This task will look at existing data, their current formats and structures. It will aim at identifying, and possibly putting into practice, solutions allowing for a facilitated integration with computing services, especially with regards to storage strategies, exchange formats, and the use of metadata.
T6.6.4 - Health data security aspects. Legal and ethics requirements on the storage and handling of health data are becoming stricter (e.g. GDPR). In some cases, they can create obstacles to data sharing and interoperability. This subtask will look at how proposed solutions can be put in place while meeting those strong security requirements. Expected outputs: Although based on a community use case, solutions explored within this task will benefit the wider User community. Expressed requirements, proposed designs and envisaged solutions will contribute to the overall output of WP6, with a strong emphasis on data security aspect and correlations between data and computing. This in turn will feed the pilot deployment activity foreseen in WP7.
Task 6.7: Use case 7. Suitable data formats for seismological big data provisioning via web services (GFZ)
Within the seismological community, the last decades have been characterised by large amount of new additional permanent and temporary conventional seismic stations becoming available. It is already clear that there are some

game-changing technologies under development. The demand for new dense observations using new technologies is advancing fast. This includes Large N deployments consisting of huge numbers of easy-to-install geophones; cheap sensors (e.g. MEMs accelerometers) used by mobile devices as well as for infrastructure monitoring; and fibre-optic based technologies, that are each starting to show their great potential in providing quality data with a wide spectrum of applications ranging from Tsunami early warning to Infrastructure monitoring. Although data quality and resolution of the three groups mentioned above are different, for seismic monitoring and archival infrastructures, these groups have in common the potential to produce large volumes of data in a very short period of time due to both the extremely dense spatial and temporal resolutions.

• Large N deployments are becoming popular thanks to the availability of cost/power effective instruments. Typically of short duration, they may imply a very varying number of sensors (from a few tens to thousands), and are presently carried out with high frequency sensors.2028

• Cheap sensors are already used since about a decade in educational projects showing capabilities to provide satisfactory recordings for earthquakes. Recently, a number of crowdsourcing projects have been launched making available a growing number of additional sensors from private users to the seismological community. These crowd-sourced projects include tailor-made cheap sensors and digitizers for private consumers as well as usage of existing sensors in mobile devices. In both cases the easiness to access new data, sometimes in area where conventional seismic stations are not available, is making these data valuable for real-time applications in the context of rapid response and earthquake early warning.

• Fibre-optic can be used as seismic arrays measuring ground motion from earthquakes using the so called distributed acoustic sensing technique (DAS) with existing underground fibre-optic cables (normally used for telecommunications, such as internet, television, and telephone service) up to tens of kilometres long or deploying dedicated short fibre-optic cables. Alternative usage of fibreoptic communication cables can be also placed along their repeaters (typically 50-100 km intervals) conventional sensors and only use the fibre-optic for communication for example at the ocean floor using existing submarine cables. The former allows for example to image the internal structure of faults with high resolution as well as to infer creeping processes of faults at submicrometre step (Jousset, et al., 2018). The latter has the potential to provide an unparalleled global network of real-time data for ocean climate and sea level monitoring and disaster mitigation from earthquake and tsunami hazards. This considerable increase in the data generation is very challenging for all seismological data centres hosting data. Despite the FAIRness already achieved in this community large volumes involved when dealing with the above listed data-sets requires new development to properly and FAIRly serve these data. Moreover, for some data sets sensors are virtual demanding new tools able to convert on-the-fly big volumes of data from proprietary to standard formats with all the obvious challenges to remain FAIR. For the data centre side this could imply the usage of different formats to archive the data. An experiment with a different data format (based on HDF5) is being held at other institution (IRIS, Incorporated Research Institutions for Seismology, USA), but additionally to this a change in the way our web services process and provide such amount of data would be needed. Moreover, the change in the size and format of the data could imply a change in the approach to the replication by means of the B2SAFE service from GEOFON to KIT. For users requesting data related to an earthquake, new formats could be used (e.g. ASDF), which allow to merge/package a variety of sources and formats, allowing for instance to mixing earthquake parametric data, seismic waveforms, provenance information, etc. This task will focus on keeping FAIRness in this community through evaluating new data formats suitable for this use case, testing and implementing tailor made services where possible within the currently defined standards. Finally, service and data quality tests will be performed to assess their quality and usability.

Task 6.8: Use case 8. Virtual definition of big datasets at seismological data centres according to RDA recommendations (GFZ)

The management of digital objects remains an area of interest that crosses disciplines, institutions and infrastructures. In this context, the need for building aggregations or collections of such objects has become an essential element. Research data management practice requires not only to describe collections, but to make them actionable by automated processes to be able to cope with ever increasing amounts and volumes of data. GEOFON, the seismological data centre at GFZ, has archived seismic waveforms since 1993 and currently archives around 10.000 streams daily from seismic stations sending data in real-time from all around the world. The archive has today a size of 100+ TB and is composed of more than 35 million files.

The standard way in which seismological data centers provide data to users is based on specifications provided by the International Federation of Digital Seismograph Networks (FDSN). These specifications are followed by almost every seismological data centre in the world, making the interoperability between them ideal from the data consumer perspective. Data centres like GEOFON acquire time series data in real-time (but also offline experiments), which are some hours later archived and made available to users by means of the API defined in the FDSN-WS specification. This API lets users define the contents of the dataset and create them on-the-fly, but the specification does not contemplate the idea of pre-assembled datasets. Data requests could be classified in two big groups: the ones related to an earthquake and the ones related to an experiment. In general, most of the data requests are related to the time and location of

an earthquake. After any big earthquake thousands of data requests are received with a considerable overlap of data between them (similar short time windows and variable set of stations), but quite rarely exactly the same dataset. But there are also some users who request all data produced in an experiment, or all data recorded by a station. This results in a big amount of data requested (with long time windows and a fixed set of stations) to be later processed and not particularly related to any earthquake. GEOFON processes more than 8 million successful dynamically created requests/ year through its Dataselect-WS, which is part of the SeisComP3 software package, an in-house development running for more than 12 years at hundreds of data centres. It would be impractical, due to storage limitations, to replicate the requested datasets by keeping a copy of each of them. Therefore, there is no way for a user to reference the dataset for future use (publication, share with someone else). Today, the user can only share the request definition, but if data are modified the resulting dataset will be different from the original one. From the data center perspective, it is also difficult to offer big pre-assembled datasets to be downloaded, due to the resources needed for their storage. For instance, all data from a seismic network, which could have a size of many TBs, but it is being continuously updated. In this context, it is very appealing the idea of using a Data Collections System in order to define and save both types of data requests. In the case of the big pre-assembled datasets we can define collections containing only "pointers" (e.g. PIDs, URLs) to the files which are included. This would imply almost no extra storage, as only the pointers are stored. Therefore, the archive could be exposed through the definition of big datasets with a marginal increase in the space needed. In the case of the dynamic datasets the same approach could be followed. Namely, for each user request we could define a collection with the PIDs of the files which fall into the range of values defined by the user. If new data comes in the future it will not appear in the collection, because the content of the dataset will not be recreated based on the original query, but on the files which originally formed the dataset. In both cases, once a collection is created, it is possible for the user to reference it for later use (e.g. share it with others or use it as supplementary information for publications), which would imply a big step towards the reproducibility of results. This constellation of a data archive with continuous time series or not predefined datasets could also be found in other disciplines, not only seismology, and this generic approach could be immediately applied by anyone in other fields. Being the main advantage the possibility to define overlapping datasets, which in case of actual replication/instantiation would imply a considerable amount of storage compared to the size of the whole archive. GEOFON, in the context of the RDA Data Collections Working Group, has worked on an implementation of this specification since its early stages in order to manage the definition and storage of pre-assembled datasets, to register the user requests, and to offer the capability of downloading these datasets. This implementation has extended the Recommendations with methods to download members of the collection and also the collection as a whole and it has been used at GEOFON with more than 6000 collections and 1.5 million members for datasets pre-defined by the data centre operators. The main challenge is to link this system to the Dataselect-WS providing the customized user-defined datasets, in order to keep track of the user requests and to offer users the possibility to refer these in publications or to share them with others. Moreover, despite of the performant implementation, the design of the extensions is still oriented towards the seismological data, but it will be redesigned in a more generic way. The final aim would also be to make the Data Collections System generic enough and ready to be adopted by others within EOSC.

Task 6.9: Integrating heterogenous data on cultural heritage (INFN)

Heritage Sciences, i.e. the application of scientific experimental methods to the analysis of cultural heritage artefacts, produces a large quantity of numeric data that are only loosely related to the cultural object to which the analyses were applied. The lack of standard data models for the different technologies employed makes interoperability between datasets almost impossible. On the other hand, the same cultural objects and activities on them (studies, interventions, etc.) are documented in textual documents usually with very basic metadata. This situation requires the intervention of a human to link the documentation of scientific analyses to the documentation of the cultural object, e.g. chemical analyses and physical to a study by an art historian; this in the end prevents data re-use and data-driven research. The team in charge of the Use Case is setting up a very general model, based on an extension of the PARTHENOS Entity Model (PEM), compliant with CIDOC CRM, the standard ontology used for cultural heritage documentation. A text mining tool based on machine learning enriches the metadata of the texts which may be linked to the metadata accompanying the digital outcomes of the scientific analyses. At present, this is being experimented on a repository of Italian scientific data created by INFN, while the text mining tool is being applied to various kinds of documentation. The implementation domain currently in use is archaeology, where both datasets are available on-line. The goal of the present use case is to embed the text mining tool in a cloud environment, exploiting its multilingual potential to apply the same procedure to repositories of documents such as those existing in UK, the Netherlands, in Belgium, in France and more, in order to enrich their metadata. Datasets in such repositories are findable and accessible via the ARIADNE registry of archaeological datasets in Europe. By using the above-mentioned standard data model for scientific data, or possibly mapping to it different data schemas in use, it is anticipated that integration (and interoperability) is feasible not only state by state but also transnationally, availing of the already existing multilingual vocabularies to establish the correspondence of terms in different languages. Since the text mining tool is natively multilingual, the necessary activities include for it only a testing stage and exploiting cross-language machine learning. Instead, for scientific datasets the X3ML mapping tool will be used when necessary, i.e. when a different data schema is present. While the

mapping needs to be done (once for all) by experts, the conversion process may be done using the templates generated by X3ML. Note that with this approach cross-language and cross-national interoperability will be achieved, e.g. linking data produced in France with metadata in French to an artifact documented in Italy in Italian. Re-use of components created within other EU projects The text mining tool has been demonstrated in a cloud environment within EOSCpilot as a Science Demonstrator for heritage texts in Italian. There it is named TEXTCROWD. At present, the tool is being substantially re-engineered to transition from a rule-based algorithm (TEXTCROWD) to a machine learning-based one. The on-line archaeological registry has been created by ARIADNE and is currently operational, enabling access to 2.000.000 datasets. Data are usually open but may be subject to restrictions; those proposed for the pilot do not have this kind of limitations. The X3ML tool has been created within the PARTHENOS cluster project. Current status of the tools Text mining tool: operational, tested only on Italian documents. Not yet implemented in a cloud environment. Scientific repository: the Italian one is based on a PEM extension and allows the procedure described above. Not yet implemented in a cloud environment. Other implementations may require mapping via X3ML. Text documents: widely available via ARIADNE. Machine access via API to datasets currently tested only for some Italian documentation repositories.
Task 6.10: Open call for communities (KIT) This task will maximise the exploitation of the project results by enabling other thematic communities or research areas not already covered in the above list to bring their service into EOSC-Pillar. The initial set of scientific uses cases collected in will be expanded through an open call for participation, inviting communities and their developers to bring forward potential services that follow the goal of the project and enhance the portfolio of the project.
The task will:
• define the appropriate selection procedure and evaluation i.c. review process.
• Appoint an evaluation committee The number of selected proposals is not limited but depends on the viability of the proposal, the requested resources, the estimated duration of integration and the overall contribution to the project and to EOSC. The selected proposals will be invited to a workshop which will be organized by and attended by technical experts from WP6, WP5 and WP7. The workshop is intended as a full-immersion experience, held in the form of a hackathon, where WP5 and WP7 experts will offer offering the invited developers one to one support with the experts to integrate their specific service. In order to maximise the attendance to the events, within this task we will seek synergies with appealing events (e.g. GridKA school, ESOF2020, ESFRI community events, etc) and relevant project and initiatives, as well as other visibility opportunities (e.g. piggybacking specialisation/master schools etc).

Participation per Partner

Partner number and short name	WP6 effort
1 - GARR	2.00
2 - CNR	32.00
4 - INFN	36.00
5 - FONDAZIONE CMCC	12.00
6 - UNIVIE	1.00
7 - CINES	28.00
8 - CNRS	39.00
9 - INRA	15.00
10 - INRIA	20.00
11 - IFREMER	9.00
12 - INSERM	12.00
13 - KIT	19.00
14 - DKRZ	13.00
15 - Fraunhofer	24.00
16 - GFZ	33.00
Total	295.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D6.1	State of the Art and Community Needs Report from Use Cases	8 - CNRS	Report	Public	8
D6.2	Demonstrator and success stories from the use cases	8 - CNRS	Demonstrator	Public	24
D6.3	Final Report on Use Cases and community involvement	8 - CNRS	Report	Public	36

Description of deliverables

D6.1: State of the art and community needs report from the use-cases [R, PU, M08; CNRS]

D6.2: Demonstrator and success stories from the use cases [O, PU, M24; CNRS]

D6.3: Final report on use-cases and community involvement [R, PU, M36; CNRS]

D6.1 : State of the Art and Community Needs Report from Use Cases [8]

This deliverable reports on the findings of the Use Cases analysis, with a special focus in the state of the art, the community requirements and on the oppportunities to extend the use case.

D6.2 : Demonstrator and success stories from the use cases [24]

A working demonstrator encompassing the use cases included in this work-package, and success stories that demonstrate the added value of the EOSC-Pillar approach.

D6.3 : Final Report on Use Cases and community involvement [36]

This deliverable reports on the integration of use cases and the involvement of national and international thematic communities.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS24	Inventory of available use cases for the project in liaison with CO-OPERAS activity	8 - CNRS	8	The MS mark the identification of a list of use cases in the Social Sciences and Humanities that can be leveraged in this WP, in cooperatio with the CO-OPERAS OPERAS GO FAIR Implementation Network.
MS25	Existing tools available for FAIRization identified and classified	2 - CNR	10	Identification of existing tools that are used by the involved communities and evaluation of their state of the art (FAIR-ready, partly FAIR-ised, available for FAIRization) for the data provenance use cases.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS26	Communications PoC between NAKALA and HAL	8 - CNRS	18	Working Communication PoC between the Huma-Num to link publications on HAL to research data available in the Nakala data repository available.
MS27	Code repository for the Dataselect-WS	16 - GFZ	18	Repository with code for the Dataselect-WS extension to support the creation of a standard miniSEED data in runtime ready.
MS28	Code repository of improved Data Collection-WS ready	16 - GFZ	24	Repository with code of the improved Data Collection - WS ready to be used and adopted by other EOSC users
MS29	Selection process for the Open Call defined	13 - KIT	20	Selction criteria, process and commitee for the open call are ready and published
MS30	First hackathon organised	13 - KIT	24	First hackathon to involve new user communities takes place

Work package number [9]	WP7	Lead beneficiary [10]	8 - CNRS
Work package title	The Infrastructure layer: delivering horizontal data storage and computing services, from national to transnational		
Start month	1	End month	36

Objectives

• bring together user communities, use cases, e-infrastructure and service providers
• provide technical and strategic guidance for the integration and federation of services in the EOSC ecosystem define solutions for new services to access distributed resources
• mediate and coordinate between technical experts and communities in the project
• support the integration of federating solutions with services in the national frameworks
• set up and test AAI solutions and proof of interoperability
• perform the validation and prove readiness for deployment of federating services through the reallife experience of scientific communities
• steer federated services on a trans-national level, feedback to standardisation activities
• deploy service solutions as developed in WP6 in the federation framework or national infrastructure
• test specific services of European-wide interest for their ability to be integrated in the EOSC framework

Description of work and role of partners

WP7 - The Infrastructure layer: delivering horizontal data storage and computing services, from national to transnational [Months: 1-36]
CNRS, GARR, CNR, CINECA, INFN, CINES, INRA, INRIA, INSERM, KIT, Fraunhofer
Cloud computing and storage services, data handling and data analysis services as well as federating services are vital to an EOSC infrastructure. As identified in previous studies, e.g. the EOSCpilot and in the EOSC-hub, several interoperability challenges and non-technical barriers have to be overcome in order to provide seamless access to distributed cloud computing resources, data storage, repository solutions and to enable mobility of modeling and data analysis codes between infrastructures. Common policies have to be established, at the regional, national, and European level, and a lightweight management of the horizontal service layer has to be put into place. The services existing at national/regional level should be connected with the EOSC: this means, on one hand, to understand which EOSC services may be beneficial for the national communities and, on the other hand, to allow for scaling at European level by interconnecting the national/regional services with those in EOSC. The key activity consists therefore in supporting the integration followed by validation and finally providing instances of the federating/enabling services in a production ready environment. These services may be shared among different national infrastructures involved in this project. This activity is complementary to what is currently done within the EOSC-hub project. The produced solutions provide support for a wide range of communities and include services that will be part of the forthcoming EOSC infrastructure layer and will be proposed as candidate services for the inclusion into the EOSC-hub service catalogue.
Task 7.1: Guidance and procedures for integrating services (INFN) This task will work towards compatibility, interoperability and interfederation of existing horizontal services across national and thematic infrastructures.
The main activities will consist of:
• analysis from a technical point of view of the different solutions relevant at national level in the different countries participating to the project

• technical guidance on how to perform the integration/federation of services with the EOSC-hub services ensuring the use of standards and state of the art solutions

• definition of a clear path and of the associated processes for the inclusion of the relevant national services into the EOSC Catalogue and Marketplace and support for this

• coordination of the contribution to external standardization bodies and relevant initiatives according to the outcome of the project activities. Among the most important services for the connection of the national resources and services to the EOSC it is worth mentioning the AAI services which are the basic pillar for this; three different solutions are already available in the EOSC-hub service catalogue and an activity to make them interoperate is ongoing. The activity in this project will leverage the outcome of it.
Task 7.2: Support for the integration of national services (KIT) This task will support the integration and federation of services relevant to the national infrastructures as well as to the use cases addressed in WP6 according to the guidelines

and procedures defined in 7.1. As highlighted in task 7.1, the AAI services are paramount, so the integration of these services that exist at the national level with those in EOSC-hub will be one of the main work in this task. Additionally this task will support the integration of community services from T7.4 in the framework.

• support service owners and service providers to implement the integration/federation of their services; setup of a testbed infrastructure to deploy the integrated services as Proof of Concept, which will then be validated by task 7.3.

• analysis of the needed services for the use cases chosen as demonstrator: storage, data catalog, data repository, etc.
Task 7.3: Integrated services validation and operation of the federated services in a production environment (KIT) The workflow in WP7 foresees the testing and validation of generic services to be addressed in this task. For a large part the work will deal with validating the implemented AAI solutions. It will also include validating the exchange of data between data services and robustness of services tested. This task will perform the validation of the federated services in close cooperation with WP6 and other WP7 tasks using the real-life experience of scientific communities. The validation will be a key task in order not only to prove the viability of the proposed solutions, but also to provide feedback to the EOSC-hub and other relevant EOSC projects, since there will be services in this project that have not been integrated in EOSC elsewhere. The activity will also include validation of the operational readiness of the services needed to achieve the connection of the national resources/infrastructures to the EOSC in a real production environment.
Task 7.4: Services ready to use (CNRS) This task is aimed at deploying services already available for several scientific communities and support their integration in the EOSC framework. It will address their uptake from national to transnational level, and their enlargement to support new communities. The services of European-wide interest that will be tested are the following:

1. The use of acceleration resources (GPU, FPGA) in cloud computing to provide high-performance software-as-a-Service close to the end-user (e.g. tensorflow/pytorch).

2. The “Galaxy as a service” platform, dedicated to research groups needing powerful computational infrastructure to run complex analyses over large datasets within the familiar Galaxy environment, e.g. those using a modeling platforms running on Tier 2 or HTC infrastructure. This will notably cover WP6.6 use-case, for health related communities, as Galaxy-E initiative dedicated to biodiversity ones.

3. Virtual Laboratories/Virtual Research Environments supporting collaboration and co-operation among scientists working in institutions located in different Member States. The VRE will offer a transparent and high-level access to federated resources by guaranteeing that established policies are met

4. Virtual Research Environment-based service for research data publishing. This service provides a community-tailored, web-based working environment facilitating the publication of “research objects”. Such a working environment offers (i) a workspace VRE members can use to collaboratively organise the items to publish, (ii) a configurable catalogue VRE members can use to publish (and discover) items according to VRE-defined typologies of “research objects” and the metadata profiles accompanying each of such objects. The content of the catalogue resulting from each VRE is made available by protocols and facilitates its re-use (e.g. OAI-PMH, DCAT);

5. Converging to common deployment tools to facilitate code circulation and mobility between different infrastructures (HPC, HTC), like containers, deployment tools (NIX or GUIX, ...), notebooks, building a shared repository for containers image deployment 6. In the FORCE and Marketplace projects Fraunhofer co-develops a business decision support systems (BDSS) and a Materials Modelling Marketplace (MarketPlace and VIMMP) as a platform for collaboration and online materials modelling laboratories. The technology of BDSS is of interest for multiple communities, e.g., optimisation of health, climate, production and other resources. In particular the combination of Marketplaces and BDSS with EOSC-Pillar offers unprecedented opportunity for near full integration of the entire Materials community in Europe and allows significant advanced.. In this task we will integrate the FORCE/BDSS and Marketplace system in the EOSC framework in order to be run as an independent service for a wide range of communities and in particular support the VRE.

Participation per Partner

Partner number and short name	WP7 effort
1 - GARR	10.00
2 - CNR	9.00
3 - CINECA	10.00

Partner number and short name	WP7 effort
4 - INFN	23.00
7 - CINES	5.00
8 - CNRS	21.00
9 - INRA	6.00
10 - INRIA	7.00
12 - INSERM	4.00
13 - KIT	21.00
15 - Fraunhofer	6.00
Total	122.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D7.1	Guidelines and Recommendations for the technical integration of resources and services into the EOSC	4 - INFN	Report	Public	24
D7.2	Report on service integration	13 - KIT	Report	Public	36
D7.3	Report on the validation statistics, operational infrastructure services and recommendations for future integration work	13 - KIT	Report	Public	30
D7.4	Report on transnational Service Usage	8 - CNRS	Report	Public	36

Description of deliverables

D7.1: Guidelines and recommendations for the technical integration of resources and services in the EOSC [R, PU, INFN, M24]

D7.2: Report on service integration [R, PU, KIT, M36]

D7.3: Deliverable on the validation statistics, operational infrastructure services and recommendations for future integration work. [R, PU, KIT, M30]

D7.4: Report on transnational service usage [R, PU, CNRS; M36]

D7.1 : Guidelines and Recommendations for the technical integration of resources and services into the EOSC [24]

This report provides directions for the technical integration of resources and services to EOSC-Pillar project participants and related parties. Report about the guidelines and procedures released in MS7.1 and MS7.2. Provides also recommendations for future work.

D7.2 : Report on service integration [36]

This deliverable reports on the integration of the EOSC-pillar servcices into the wider EOSC. This will include details about the actions carried out in the integration activity, as well as technical information and documentation

that will be useful to maintain the services operational. Includes reports MS7.3 and MS7.4 their updates and supplies conclusions and recommendations.

D7.3 : Report on the validation statistics, operational infrastructure services and recommendations for future integration work [30]

The document will report of the validation criteria, processes and results (see MS 7.5, 7.6 and 7.7). This will include the validation of the operational readiness of the services needed to achieve the connection of the national resources/ infrastructures to the EOSC in a real production environment and will provide recommendations for future integration activities.

D7.4 : Report on transnational Service Usage [36]

This deliverable reports on the transnational usage of the newly-integrated services.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS31	Integration and Federation Guidelines	4 - INFN	9	This MS marks the release of guidelines for the technical integration/federation of resources and services with the EOSC.
MS32	Procedures to include national services in EOSC released	4 - INFN	12	The MS marks the release of procedures to include the national services into the EOSC-hub service catalogue and marketplace.
MS33	Interim integration report and gap analysis	13 - KIT	18	Issue of the report on the services integrated and, if applicable, a gap analysis on missing features which could improve the existing services. The MS is related to the final report on this activity (D7.2 that will be issued at the end of the project), of which it is a working version.
MS34	Report on the services integrated at month 34	13 - KIT	36	Issue of a report on the services integrated at month 34. It is related to D7.2, to be delivered at the end of the project.
MS35	Intermediate report on interoperability and validation	13 - KIT	8	Intermediate report on interoperability status and requirements, service validation criteria, and the tools and procedures selected for the service validation suite.
MS36	Validation suite ready and tested	13 - KIT	10	Release of the validation suite to test operational readiness of the services in a real production environment.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS37	Report on validation runs and first operational service readiness	13 - KIT	24	Issue of a report on the validation runs, including findings and recommendations, and first operational service ready.

Work package number [9]	WP8	Lead beneficiary [10]	1 - GARR
Work package title	Ethics requirements		
Start month	1	End month	36

Objectives

The objective is to ensure compliance with the 'ethics requirements' set out in this work package.

Description of work and role of partners

WP8 - Ethics requirements [Months: 1-36]
GARR
This work package sets out the 'ethics requirements' that the project must comply with.

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D8.1	POPD - Requirement No. 1	1 - GARR	Ethics	Confidential, only for members of the consortium (including the Commission Services)	4

Description of deliverables

The 'ethics requirements' that the project must comply with are included as deliverables in this work package.

D8.1 : POPD - Requirement No. 1 [4]

4.2 The host institution must confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research. For host institutions not required to appoint a DPO under the GDPR a detailed data protection policy for the project must be kept on file. 4.11 Detailed information on the informed consent procedures in regard to data processing must be kept on file. 4.12 Templates of the informed consent forms and information sheets (in language and terms intelligible to the participants) must be kept on file.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification

1.3.4. WT4 List of milestones

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
MS1	Project KOM organised	WP1	1 - GARR	1	1st project meeting to officially present the project objectives and kick-off activities to the project participants, introduce the project management team, the activity leaders and project workers to each other.
MS2	Project set up and running, Quality Plan and Risk Mitigation Plan Lauched	WP1	1 - GARR	2	This milestone marks the launch, through the appropriate project communication channels, of full-fledged project procedures, including QA, Risk and Innovation management.
MS3	The EOSC-Pillar web presence online	WP2	18 - TRUST-IT SRL	1	The web platform of the EOSC Pillar internal web presence and landing page is available for first release.
MS4	Second release of the EOSC-Pillar web presence online		18 - TRUST-IT SRL	4	Roll-out of all services on the web platform are available.
MS5	Organisation and roll-out of the 1st workshop	WP2	18 - TRUST-IT SRL	14	Organization of EOSC-Pillar 1st workshop open to all stakeholders involved in research data management, e.g. scientific/ user communities, research institutions, research infrastructures and eInfras, as the stakeholders listed.
MS6	Organisation and roll-out of second and third workshop	WP2	18 - TRUST-IT SRL	36	Organisation of second and third project workshops, open to all stakeholders involved in research data management, e.g. scientific/ user communities, research institutions, research infrastructures and eInfras, as the stakeholders listed.
MS7	Organisation and roll-out of first webinars	WP2	18 - TRUST-IT SRL	6	This MS marks the delivery of first project webinars, addressing all stakeholders involved in research data management, e.g. scientific/ user communities, research institutions, research

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
					infrastructures and eInfras, as the stakeholders listed.
MS8	Organisation and roll out of 10 webinars	WP2	18 - TRUST-IT SRL	36	This MS marks the delivery of all 10 the foreseen EOSC-Pillar webinars, Open to all stakeholders involved in research data management, e.g. scientific/ user communities, research institutions, research infrastructures and eInfras, as the stakeholders listed.
MS9	Final result-oriented event	WP2	18 - TRUST-IT SRL	36	This MS marks thge delivery of a final event aimed at showcasing the project results, co-located with the other projects funded under INFRA EOSC subtopic b) to maximise geographical coverage.
MS10	Definition of survey targets	WP3	6 - UNIVIE	1	This deliverable marks the definition of survey targets, which will be co-designed with WP4 and the research communities involved in WP6, and provide input tho the rest of WP3 activity.
MS11	Legal consultancy in legal and ethical issues completed	WP3	6 - UNIVIE	3	This MS marks the completion of the consultancy for legal and ethical issues, including expert advice on GDPR, access and subsequent use, security of processing, non-disclosure, as well as prevention of liability in relationship with survey activities.
MS12	Determination and definition of distribution mechanism	WP3	13 - KIT	3	This MS marks the definition of the mechanisms and procedures to be used to distribute the survey to target respondents.
MS13	Setup of the Technical infrastructure for the realisation of the surveys	WP3	6 - UNIVIE	3	This milestone marks the implementation of the technical infrastructure (i.e. forms, database etc) that will be used to deliver the survey.
MS14	Definition of questions, definition of questionnaires	WP3	6 - UNIVIE	4	This MS marks the definition of the questionnaires to

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
					be proposed to survey respondents.
MS15	Survey roll-out	WP3	6 - UNIVIE	6	This deadline marks the completion of the survey in the different countries.
MS16	Analysis and project internal presentation of results, findings, outcomes	WP3	6 - UNIVIE	7	This MS marks the completion of the survey data analysis and the preliminary presentation of the findings to the project participants.
MS17	Visualisation of results for external audiences	WP3	6 - UNIVIE	8	Creation and publication of graphics and other data visualisations to present data to external audiences.
MS18	Requirements of open data, services and security management	WP4	13 - KIT	6	This MS marks the definition of requirements for open data, data protection and cross border data access and indicate limitations and common practice of handling scientific data in national contexts. This is a key step towards the policy and legal framework.
MS19	Setup of the regional coordination board	WP4	1 - GARR	6	This MS wikk mark the creation of the Regional Coordination Board and its first meeting.
MS20	Legal framework for trans-national data services	WP4	13 - KIT	14	The MS marks the definition of the legal framework to be used to implement and use trans-national open data services. It is part of D 4.1
MS21	Proposed service management structure for EOSC-Pillar	WP4	13 - KIT	14	The MS marks the presentation of a proposal for the management structure of the project services, that should be scalable enough to encompass other ones too. The proposal will be discussed by partners and the results will become part of D4.1
MS22	Business model round of feedback from National Initiatives	WP4	8 - CNRS	16	This MS marks the collection of input an feedback on business models from National Initiatves participants (through a workshop or a virtual meeting)

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
MS23	Business model workshop	WP4	8 - CNRS	24	A workshop to gather further feedback on the proposed business models, open to national stakeholders and relevant funding agencies, together with relevant international organisations and initiatives, including RDA and CODATA.
MS24	Inventory of available use cases for the project in liaison with CO-OPERAS activity	WP6	8 - CNRS	8	The MS mark the identification of a list of use cases in the Social Sciences and Humanities that can be leveraged in this WP, in cooperatio with the CO-OPERAS OPERAS GO FAIR Implementation Network.
MS25	Existing tools available for FAIRization identified and classified	WP6	2 - CNR	10	Identification of existing tools that are used by the involved communities and evaluation of their state of the art (FAIR-ready, partly FAIR-ised, available for FAIRization) for the data provenance use cases.
MS26	Communications PoC between NAKALA and HAL	WP6	8 - CNRS	18	Working Communication PoC between the Huma-Num to link publications on HAL to research data available in the Nakala data repository available.
MS27	Code repository for the Dataselect-WS	WP6	16 - GFZ	18	Repository with code for the Dataselect-WS extension to support the creation of a standard miniSEED data in runtime ready.
MS28	Code repository of improved Data Collection-WS ready	WP6	16 - GFZ	24	Repository with code of the improved Data Collection - WS ready to be used and adopted by other EOSC users
MS29	Selection process for the Open Call defined	WP6	13 - KIT	20	Selction criteria, process and commitee for the open call are ready and published
MS30	First hackathon organised	WP6	13 - KIT	24	First hackathon to involve new user communities takes place
MS31	Integration and Federation Guidelines	WP7	4 - INFN	9	This MS marks the release of guidelines for the technical integration/federation of

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
					resources and services with the EOSC.
MS32	Procedures to include national services in EOSC released	WP7	4 - INFN	12	The MS marks the release of procedures to include the national services into the EOSC-hub service catalogue and marketplace.
MS33	Interim integration report and gap analysis	WP7	13 - KIT	18	Issue of the report on the services integrated and, if applicable, a gap analysis on missing features which could improve the existing services. The MS is related to the final report on this activity (D7.2 that will be issued at the end of the project), of which it is a working version.
MS34	Report on the services integrated at month 34	WP7	13 - KIT	36	Issue of a report on the services integrated at month 34. It is related to D7.2, to be delivered at the end of the project.
MS35	Intermediate report on interoperability and validation	WP7	13 - KIT	8	Intermediate report on interoperability status and requirements, service validation criteria, and the tools and procedures selected for the service validation suite.
MS36	Validation suite ready and tested	WP7	13 - KIT	10	Release of the validation suite to test operational readiness of the services in a real production environment.
MS37	Report on validation runs and first operational service readiness	WP7	13 - KIT	24	Issue of a report on the validation runs, including findings and recommendations, and first operational service ready.

1.3.5. WT5 Critical Implementation risks and mitigation actions

Risk number	Description of risk	WP Number	Proposed risk-mitigation measures
1	Disputes between partners (Likelihood: Low, Impact: Medium)	WP1	The Consortium Agreement and D1.1 will contain necessary conflict resolution procedures.
2	Failure to commit to the project workplan, resulting in execution delays (Likelihood: Low, Impact: Medium)	WP1, WP2, WP3, WP4, WP5, WP6, WP7	The WP Leaders and the Project Coordinator will impose specific corrective ations throughout the project lifecycle to provide the necessary flexibility ensured by a carefully designed workplan. Multiple consortium members are focussed on each given task such that a underperforming partner can be replaced by additional resources from another.
3	Failure of WP leaders to perform adequately or unavailability of the leader (Likelihood: Low, Impact: High)	WP1	Regular meetings will address this in good time. A deputy WP Leader has been appointed prior to the start of the project and will step-in in case of underperformance/failure of the WP coordinator, or of (temporary) unavailability.
4	Communication and outreach failures Likelihood: Low, Impact: Medium	WP2, WP3	WP2 leader has multiple years of experience in the field and a positive track-record in similar initiatives. As a contingency measure, the continuous communication, synchronisation, and engagement monitoring activities will ensure prompt corrections.
5	Lack of engagement in the consultation platform or surveys. (Likelihood: Medium, Impact: High)	WP2, WP3, WP4, WP5, WP7	The partners will leverage their vast network of national, European and trans-national contacts, utilised in several cutting-edge initiatives in research infrastructures, and will exploit their current advantageous position gained in the field thanks to their central role in critical initiatives, already identified and preliminary engaged.
6	Lack of participation to hackathons (Likelihood: low, Impact: medium)	WP6	The project will piggyback existing events that are appealing for the target community (GridKA, ESOF2020, ESFRI clusters' meetings). The format of the events will be designed to be an appealing training opportunity for young researchers that will be recruited thanks to the partner's liaisons with relevant master/ specialization courses. Travel coverage for a part of the participants is also considered.
7	Lack of interest from the research communites in contributing in, validating and using the developed solutions (Likelihood: low, Impact: high)	WP5, WP6, WP7	Communities are involved in the very design of this project. The use cases addressed belong to large cross-border thematic communities, and can build upon a strong user base and the appropriate channels to ensure the uptake of the new services are already in place. As a contingency plan, the national initiatives can tap into their participants in case there's a need to involve more users.
8	Lack of participation to workshops (Likelihood: Low, Impact: Medium)	WP2	Include also events with remote participation, e.g. webinars. Moreover, the workshops will be co-located with other EOSC related events.

Risk number	Description of risk	WP Number	Proposed risk-mitigation measures
9	Failure or major difficulties in deploying the services (Likelyhood: low, Impact: high)	WP5, WP7	The partners responsible for the services' integration and deployment are very experienced and have a positive track in delivering services and are involved in a number of related projects. Monitoring activities will ensure prompt corrections.
10	Failure to coordinate effectively with EOSC and EOSC-related initiatives (Likelihood: low, Impact: high)	WP2, WP5, WP7	Many of the consortium key partners have a direct involvement in the EOSC-hub. EOSCpilot, EOSCsecretariat projects as well as in thematic clusters. Through them, any lack of coordination can be promptly addressed.
11	Countries do not implement the proposed national service registry model (Likelihood: Low within the addressed countries, High for third countries; Impact: High)	WP4, WP5	The national service registry is considered one of the steps of the EOSC-Pillar strategy. Accordingly, partners (which are either the coordinator or one of the key stakeholders for the national initiatives in the respective countres) are committed to adopt it and make it as viable as possible. However, the project has no guarantees as to whether other countries will adopt the model, and the risk that it won't happen remains high. There are two strategies that can be put in place to mitigate this, and the project will work on both. 1) Make a good case for the adoption of the model, provide an early working prototype and propose it for the adoption, engage and offer training/support to initiatives interested in the adoption. 2) Co-ordinate with the other national initiatives (and with the INFRAEOSC-05-b projects collecting national initiatives) to understand if they propose a different model and ensure it is compatible, if not interoperable, with ours: in a federated environment like the one we are envisaging, there is room for different solutions as long as they can interact with each others. In this sense, there is no need that all the national initiatives adopt our model, as long as we are able to coordinate with them.
12	The national service registry is not sustainable/not of interest to the national initiatives and/or the communities [Likelyhood: Medium; Impact: High]	WP4, WP5	The national service registry answers a need which is strongly felt in the research community: that of uniforming the computing and data services available in a certain domain. While the effort of putting it together is considerable, its technical maintenance is not especially costly for the organisations involved. On the other hand, the risk is that the description of services are not kept up to date by the communities. To reduce the risk, the project will 1) make the maintenance of the relevant information as lightweight as possible and 2) try to encourage researchers/developers to a) enroll the services and b) keep the information on the services up to date by providing incentives (for instance by working with national governments to include the

Risk number	Description of risk	WP Number	Proposed risk-mitigation measures
			availability of FAIR data and related services in the evaluation of research).

1.3.6. WT6 Summary of project effort in person-months

	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	Total Person/Months per Participant
1 - GARR	30	4	2	16	0	2	10		64
2 - CNR	0	1	4	9	21	32	9		76
3 - CINECA	0	1	4	1	0	0	10		16
4 - INFN	0	6	0	9	4	36	23		78
5 - FONDAZIONE CMCC	0	1	1	1	3	12	0		18
6 - UNIVIE	0	21	34	6	7	1	0		69
7 - CINES	20	1	1	6	14	28	5		75
8 - CNRS	0	1	7	9	12	39	21		89
9 - INRA	0	1	1	1	2	15	6		26
10 - INRIA	0	1	1	1	0	20	7		30
11 - IFREMER	0	1	1	6	2	9	0		19
12 - INSERM	0	1	1	1	4	12	4		23
13 - KIT	0	1	10	11	15	19	21		77
14 - DKRZ	0	4	1	1	2	13	0		21
15 - Fraunhofer	0	3	3	6	29	24	6		71
16 - GFZ	0	1	1	1	0	33	0		36
17 - UGent	0	8	6	1	13	0	0		28
18 - TRUST-IT SRL	8	14	4	6	2	0	0		34
· COMMpla	0	0	0	0	0	0	0	0	0
Total Person/Months	58	71	82	92	130	295	122		850

1.3.7. WT7 Tentative schedule of project reviews

Review number [19]	**Tentative timing**	**Planned venue of review**	**Comments, if any**
RV1	20	On site or Brussels	
RV2	36	On site or Brussels	

1. Project number

The project number has been assigned by the Commission as the unique identifier for your project. It cannot be changed. The project number **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

2. Project acronym

Use the project acronym as given in the submitted proposal. It can generally not be changed. The same acronym **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

3. Project title

Use the title (preferably no longer than 200 characters) as indicated in the submitted proposal. Minor corrections are possible if agreed during the preparation of the grant agreement.

4. Starting date

Unless a specific (fixed) starting date is duly justified and agreed upon during the preparation of the Grant Agreement, the project will start on the first day of the month following the entry into force of the Grant Agreement (NB : entry into force = signature by the Commission). Please note that if a fixed starting date is used, you will be required to provide a written justification.

5. Duration

Insert the duration of the project in full months.

6. Call (part) identifier

The Call (part) identifier is the reference number given in the call or part of the call you were addressing, as indicated in the publication of the call in the Official Journal of the European Union. You have to use the identifier given by the Commission in the letter inviting to prepare the grant agreement.

7. Abstract

8. Project Entry Month

The month at which the participant joined the consortium, month 1 marking the start date of the project, and all other start dates being relative to this start date.

9. Work Package number

Work package number: WP1, WP2, WP3, ..., WPn

10. Lead beneficiary

This must be one of the beneficiaries in the grant (not a third party) - Number of the beneficiary leading the work in this work package

11. Person-months per work package

The total number of person-months allocated to each work package.

12. Start month

Relative start date for the work in the specific work packages, month 1 marking the start date of the project, and all other start dates being relative to this start date.

13. End month

Relative end date, month 1 marking the start date of the project, and all end dates being relative to this start date.

14. Deliverable number

Deliverable numbers: D1 - Dn

15. Type

Please indicate the type of the deliverable using one of the following codes:

R	Document, report
DEM	Demonstrator, pilot, prototype
DEC	Websites, patent fillings, videos, etc.
OTHER	
ETHICS	Ethics requirement
ORDP	Open Research Data Pilot
DATA	data sets, microdata, etc.

16. Dissemination level

Please indicate the dissemination level using one of the following codes:

PU	Public
CO	Confidential, only for members of the consortium (including the Commission Services)
EU-RES	Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
EU-CON	Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
EU-SEC	Classified Information: SECRET UE (Commission Decision 2005/444/EC)

17. Delivery date for Deliverable

Month in which the deliverables will be available, month 1 marking the start date of the project, and all delivery dates being relative to this start date.

18. Milestone number

Milestone number:MS1, MS2, ..., MSn

19. Review number

Review number: RV1, RV2, ..., RVn

20. Installation Number

Number progressively the installations of a same infrastructure. An installation is a part of an infrastructure that could be used independently from the rest.

21. Installation country

Code of the country where the installation is located or IO if the access provider (the beneficiary or linked third party) is an international organization, an ERIC or a similar legal entity.

22. Type of access

VA	if virtual access,
TA-uc	if trans-national access with access costs declared on the basis of unit cost,
TA-ac	if trans-national access with access costs declared as actual costs, and
TA-cb	if trans-national access with access costs declared as a combination of actual costs and costs on the basis of unit cost.

23. Access costs

Cost of the access provided under the project. For virtual access fill only the second column. For trans-national access fill one of the two columns or both according to the way access costs are declared. Trans-national access costs on the basis of unit cost will result from the unit cost by the quantity of access to be provided.