Title of the Project: Interoperable Data Lake (IDL)
Leading Partner: LEONARDO (Carolina Berucci, carolina.berucci@leonardo.com)
List of participants (ICSC public and private partners): • **INAF** • **INFN** • **Leonardo S.p.A.** • **Thales Alenia Space Italia S.p.A.**
Proposer Spoke: Spoke 3
Other Spokes involved: Spoke 2
Short Abstract: The Project aims at creating a Data Lake service, supporting a seamless access to space and ground-based observations and simulated data. The project addresses the design and commissioning of an interoperable, distributed data archive, relying on state-of-the-art open technologies, supporting both science and industry The service will specifically address the challenges related to the big data scenario, in terms of both data management, storage, access, identification and of access to computing resources necessary to process the data. The definition of a data model to describe and access the data will be designed according to FAIR and IVOA standards. New techniques of block-chain and web-based stacks like Object Storage, will be used, tuned and linked together to optimize big data storage and efficiency of data retrieval, exploiting state-of-the-art cloud-based technologies. A high-level functional architecture of the system (space + ground) will be deployed, ensuring the suitable level of secure access (SSA), focusing in particular on the processing of satellite data. Effective algorithms for data processing will be selected to create a mock-up simulator capable of generating a synthetic data set based on the identified use cases and operational scenarios, testing the algorithmic chain, and evaluating the computational load.
Starting TRL of the project: 4 **Target TRL of the project** (Technology Readiness Level): 5

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

Expected budget (total cost €): 775,576.19 €
% of total budget for activities in "Regioni del Mezzogiorno" (Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, Sardegna, Sicilia)**: 8%**

1. **Description of the project**:

In ten years, the space environment has radically changed, with an increase in investments, the emergence of new nations having or close to have space capabilities, the enlargement of multiple private operators, the growing use of new mode of business and operations, the introduction of new technologies, including dual capabilities, and finally the creation of space forces. The consequence of these developments is a significant increase in competition for access to space, the risk of collision and its catastrophic exponential effects, as well as hostile actions towards space and terrestrial infrastructures. The knowledge of space domain in terms of orbiting objects, especially those capable of offensive actions, and space debris population is mandatory.

The amount of data generated and processed by space observations is today huge and it will increase exponentially in the years to come, which raises challenges to provide data-intensive services in a timely and comprehensive manner. At the same time, the incoming generation of astrophysical observatories at various wavelengths is anticipated to have dramatically improved performance in terms of sensitivity and resolution, allowing the finding and characterization of novel classes of sources, so far barely detectable. Finally, ultimate HPC systems allows to run simulations of unprecedented accuracy and resolution, producing enormous data volumes to be stored, published, analyzed.

This will represent not only an invaluable opportunity for scientists, but also an outstanding technological challenge. The expected data volume will be hard to manage with traditional approaches.

This project addresses the challenges of big data data gathering, sharing, knowledge, safety, governance, Intellectual property, training and source of enhancement for current activities, research and services, to scale and to accelerate an efficient and sovereign data exchange using data federation standards. A federated data lake will embed services for the extraction of actionable insights from these large amounts of data and extend the use of space applications to new research lines.

For example, using the *GAIA-X industrial standard to implement a Space dataspace,* Leonardo aims at implementing a marketplace where the exchange of "Space" data represents a value for the different parties (data providers & communities) and where data and services can be made available, collated, and shared in an environment of trust. In Astronomy and Astrophysics equivalent standards are used to interoperate services and applications without concerns about different wavelengths and fields of application like simulations or observational data or time series or catalogs and are developed by International Virtual Observatory Alliance (IVOA).

Building on the top of those major standards, the IDL project will create a federated data infrastructure that follows the FAIR data principles. The resulting data lake service will

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

efficiently manage large data volumes and can interoperate with the different EOSC communities. In the perspective of growing archives, scalable resources need to be exploited and it is thus necessary to design robust, open-source solutions allowing data providers and data users not only to store but also process data of increasing size in a cloud-based environment.

The proposal addresses the usage of Cloud technologies to realize a digital platform to provide, as service, space factory infrastructure to users and communities both from the science and from industry. The access to computing capabilities, both in terms of HPC and virtualized environments (IaaS) can boost productivity and support complex computing intensive workflow. A cloud based secured data lake is critical to deploy data storage accessible from the factory, while being accessible through catalogue and archived.

The outcome of this project is to design and implement a prototype application performing the end-to-end processing chains to perform tests with a hybrid cloud/HPC deployment to exploit computing capacity to distributed data sources/services.

A key activity to obtain this objective is the definition and use of a metadata layer and standard, to be developed in close collaboration with data provider communities, such us, in our case, the Spoke3 scientists. A federated data lake will embed services for the extraction of actionable insights from these large amounts of data and extend the use of astrophysical applications to new research lines. The involved scientific and industrial communities will have access to a certified and federated data catalog, ready to be used by other data sectors.

We will implement a cloud-native distributed database with linear horizontal scalability. The purpose of this activity is to support the integration and query of data coming from different sources, complemented with simulated data, with a performance that enables novel application with real-time requirements in SSA use case, to achieve maximum effectiveness and efficiency in data provisioning and exploitation. The data will be processed in batches and then presented to the user as a set of predefined tiles and time-sliced images. This type of solution includes innovative analytics with advanced statistics and ML technique, data linkage and multi-sources data integration. A fully automated prototype application performing the end-to-end processing chains will be designed, implemented, and tested in the project. Tests will be performed on a cutting-edge infrastructure supporting horizontal scalability and hybrid cloud/HPC deployment.

Ownership and immutability of data are crucial for the project, this becoming clearer focusing on the creation of a marketplace of spatial data. A private, dedicated blockchain as a notary for tracking mutations of space data will be deployed. With the increasing volume and complexity of space-related information, ensuring the integrity, security, and accountability of data becomes critical. Our objective is to leverage blockchain technology to create an auditable trail of data actions in a secure and efficient manner, enhancing security, privacy, cost efficiency, governance and control, trust and accountability, collaboration, and consortium building within the domain of space data.

In the project the blockchain infrastructure will be integrated into the existing data pipeline through REST APIs. By leveraging the RESTful architecture of the software, the project will integrate the appropriate HTTP requests to enable seamless data transfer and synchronization between the blockchain and the space data systems. This integration will ensure that every data mutation is recorded securely and immutably on the blockchain, providing a transparent and auditable history of data actions.

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

To achieve enhanced security, the project will implement advanced cryptographic techniques within the blockchain infrastructure. Encryption algorithms will be employed to protect sensitive space data from unauthorized access and tampering. Access control mechanisms will be established to restrict data manipulation to authorized actors, ensuring the integrity and confidentiality of the information. The blockchain's decentralized nature will add an extra layer of security by eliminating the reliance on a central authority and reducing the risk of single points of failure.

Effective governance and control over data-related processes are critical. The blockchain infrastructure will establish predefined rules and protocols through smart contracts, enabling organizations to govern access, data sharing, and decision-making within the network. This ensures transparency, accountability, and responsible data management. Organizations participating in the network will have increased control over their data while adhering to industry standards and regulations.

To ensure the performance and functionality of the solution, a comprehensive verification phase will be conducted, involving the development of a software demonstrator or prototype. Given that the blockchain software will provide RESTful APIs, these APIs will be integrated into a more complex data pipeline.

During the verification phase, testing will be conducted to evaluate the integration of the blockchain infrastructure with the existing systems. This testing will involve simulating various data mutations to assess the system's responsiveness, scalability, and data integrity. By subjecting the solution to simulated scenarios, potential vulnerabilities and issues can be identified and addressed before deployment.

Moreover, the project will also study the Space sensors for SSA applications based on innovative technologies, new processing algorithms, extended use of on-board processing and AI/ml-empowered data fusion. These will provide a capability leap in terms of

- **24/7 operation and continuity:** space-based surveillance and tracking are not impacted by atmospheric and weather conditions and by day and night cycles
- **Accuracy**: space-based measurements are not affected by the impairments of the atmosphere
- **Global coverage**: space sensors are designed and deployed to complement and augment the coverage of ground-based assets, whose installation/operation is precluded in remote and oceanic areas
- **Responsiveness**: space sensors can improve the revisit of the space volume(s) requiring “constant” monitoring; further, space-based data relay (i.e. inter-satellite links) can provide real time tasking and fast telemetry transmission of/by space sensors.

The focus of the activity is on simulation of state of art algorithms for sensor data processing with the aim to evaluate computational load.

Enabling technologies:

The technologies that will drive the data-lake implementation of the mentioned use cases will be based on:

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

- Metadata management through suitable database (No-SQL + Relational);
- Data Management through the data lake (Third party copy services + Rucio);
- Data certification via Blockchains (REST API with an HTTPS);
- Cloud-enabled orchestration services.

Relationship between the proposed Project and the Spoke flagship objectives

The project uses technologies and solutions relevant to both involved Spokes and in particular:

- Spoke 2 WP5 (Data Management) and WP6 (Space Economy)
- Spoke 3 WP4 (Big Data Management, Storage and Archiving

1.1 Expected results

- Design and commissioning of an interoperable, distributed data archive in the perspective on the ICSC HPC, Cloud and Data infrastructure
- Availability (to both scientific and industrial users) of effective interoperable services for the storage and the processing of the data
- Experimentation and Demonstration of the use of blockchain sw stacks for the certification / tracking of valuable datasets
- Updating existing debris databases with new space –based observations
- Updating existing debris databases with observations from space
- Creation of a public data archive of state-of-the-art astrophysical simulations
- Experimentation of technological solutions capable to support science in large astronomy observatories like LOFAR and the SKA

2. Expected impact:

- Creating effective and long-term synergies among main industry and key research centers in order to give a precise direction towards deployment and operation of data lake solutions interoperable with main international data provider and EOSC services
- Strengthening and growing skills and competencies at the national level in the fields of big data management, storage, and access technologies, prompting the adoption of open software and standards following FAIR principles
- Consolidating national positioning in the European and International context both in industry and science, promoting the role of national leadership in the technological and research ecosystem.

3. Implementation

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

3.1. Work Packages structures and timeline of the project:

Define work packages and tasks, and involved partners

Specify list of deliverables. For each deliverable, clearly indicate delivery month and the milestone of Center in which it will be expected.

Gantt Chart (Not mandatory)

WP1 - Activity #1 (Data Lake)

WP title:	WP1: Data Lake		
Start date:	T0	**End date**:	T0+24M
Participants:	INFN (Spoke 2), Leonardo		
1. OBJECTIVES			
• Definition and PoC of the deployment model of the Data Management system. • Evaluation and benchmarking the Data Management system with domain specific data.			
2. INPUTS			
WP2, WP3, WP4, WP5			
3. ACTIVITY DESCRIPTION			
Activity description: INFN proposes to test solutions for managing data in a geographically distributed environment to effectively exploit the Datalake. The aim is to prepare an end-to-end prototype to demonstrate the Data Management (DM) capability. DM is the process through which analysis algorithms have access to the data they need to allow processing and analysis. In very general terms, in order to satisfy this requirement, there needs to be the ability to store and retrieve data. This data is assumed to be stored as files or data objects. This definition excludes data stored in specialist services, such as a database. Moreover, the prototype will show the feasibility to deploy Platform as a Service (PaaS) services for the actual processing of the data ingested into the Datalake. All this together will demonstrate the capability to analyze the astrophysical observations and simulations available data in a cloud environment. The proposed activity will be responsible to deployed a DataLake platform by means of a set of central services responsible for: - Third party copy - Orchestration of multiple and distributed storage endpoints - Management of the actual replicas of data - Basic metadata handling In addition, a number of endpoints for storing and archiving data will be provided in order to actually realize the Datalake infrastructure.			

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

During the first part of the activity the cloud-native deployment model for the overall environment will be prepared and tested. During the second phase of the activity, once the deployment is ready, the aforementioned system will be open to testing activities using domain specific data. From a technical perspective all the related services will be deployed in the form of containers managed via container orchestrators such as Kubernetes. The DataLake exploitation for the actual data processing will be enabled through PaaS level services implemented as container-based systems as well and served through a k8s system integrating the INFN-Cloud portfolio of services.

4. OUTPUTS				
Software			**Hardware**	
Definition		Del. date	**Definition**	Del. date
Documentation			**Others**	
DRL code	Definition	Del. date	**Definition**	Del. date
TN1	Report the deployment solution used to implement the Data Management and results of the functional tests	M12		
TN2	Report the deployment solution used to implement the Data Management and results of the functional tests	M24		

Hardware needs from ICSC

- A cluster of virtual machines for a total of 200 cores. These will host a k8s platform
- A total of 2 TB of disk space in form of persistent volumes to be attached to the above cluster of VMs
- 100 TB of distributed storages (at least 2 endpoints)

WP2 - Activity #2 (Data Models and metadata definition, data archiving)

WP title:	WP2: Data Models and metadata definition, data archiving and database		
Start date:	T0	**End date**:	T0+24M
Participants:	INAF, Leonardo		

1. OBJECTIVES

- Define a data model for organizing, find and access archived data
- Design and commissioning of a relational database for metadata management identification and provisioning of data

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

2. INPUTS
WP2, WP3, WP4, WP5
3. ACTIVITY DESCRIPTION
Study and implementation of a IVOA/FAIR compliant data model for the correct description and handling of the data sets. The data model will include all relevant information about data, software for data reduction, analysis, data policy, and all the information for data filtering for retrieval. A smart ingestion system to collect and organize data, tailored on data retrieval on specific use cases will be implemented in order to enhance the fast data discovery and the easiness of retrieval in the data lake environment. This WP will also take care of the FAIR and Open Science principles. - **Task 2.1: Datasets identification** Selection of data coming from radioastronomy, space debris observations, numerical simulations to implement the cases of use of the project and to guide the development of data models. - **Task 2.2: Data Models and metadata definition** Identification of a suitable data model and metadata definition about the provided data set of astronomical observations and the simulated/synthetic data. Implementation of an ingestion software to import the metadata descriptors into the database and publication of results in a VO compliant way. - **Task 2.3: Requirement analysis** Gather information necessary for the definition, implementation, testing and verification of the database technology to be integrated on top of the data-lake. Analysis and definition of the required data processing activities: preprocess, cleaning, normalize, rescale data, with a focus on data models and metadata definition specified in task 1. Analysis and specification of data query functionalities for database. - **Task 2.4: Database deployment in ICSC infrastructure** Implementation of the data ingestion service, to collect data efficiently from the instrument/sensor network. Integrate the components of the prototype system with their applications and data platform and deploy them on the Consortium infrastructure, by identifying the software plugins that are needed to make both DB and data processing pipelines design in WP1 broadly available. - **Task 2.5: Validation and testing** The goal is a performance test of the processing pipelines on a hybrid HPC/cloud infrastructure. Testing may highlight bottlenecks in the data processing pipelines which will be addressed by fine tuning the software produced in task 1.2. Verify model integration, real-time performance analysis, and optimization design.

4. OUTPUTS

Software: Database system for the management of metadata describing provided datasets		**Hardware**	
Definition	Del. date	**Definition**	Del. date
Requirement analysis Data provisioning	M12-MS9 M12 - MS9		

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

Database deployments Validation and testing		M18-MS10 M24-MS10		
Documentation			**Others**	
DRL code	Definition	Del. date	**Definition**	Del. date
TN1	Dataset identified and data model vs1	M12		
TN2	Technical report of the database	M18		
TN3	Final report on database and data model vs1.1			
Hardware needs from ICSC				
A cluster of at least 3 VMs dimensioned as follows: 50 GB of OS; 16GB RAM; 8 cores; 10TB disk space locally mounted.				

WP3 - Activity #3 (BlockChain system for the data lake)

WP title:	WP3: BlockChain system for the data lake		
Start date:	T0	**End date**:	T0+24M
Participants:	INFN (Spoke 2)		
1. OBJECTIVES			
• License purchasing of a commercial BlockChain • Deployment for the use of the project • Customisation definition tailored on projects requirements and configuration setup accordingly • Workflow setup for data objects using software REST APIs			
2. INPUTS			
WP1, WP2, WP4, WP5			
3. ACTIVITY DESCRIPTION			
The project involves the deployment of a blockchain network that utilizes a set of smart contracts specifically designed to execute fundamental CRUD (Create, Read, Update, Delete) operations, catering to the unique requirements of the project. This blockchain network will be accessible through a REST API with an HTTPS backend, facilitating seamless communication between the blockchain and any client. To ensure secure and authorized access to the network, an enrollment procedure will be established. This procedure will enable the creation of new users who will be granted permission to operate within the network. By implementing this enrollment process, the system can maintain strict control over user access and ensure that only authorized individuals can interact with the blockchain network.			

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

A web interface will be available for all the operations and to follow the objects mutation histories.				
4. OUTPUTS				
Software			**Hardware**	
Definition		Del. date	**Definition**	Del. date
Licensed software deployment and configuration. Client realization to perform all the operations described in the functional analysis		M9 M24		
Documentation			**Others**	
DRL code	Definition	Del. date	**Definition**	Del. date
TN1	Functional, Analysis, document	M6		
TN2	Workflow definitions for data object tracking, and implementation in the licensed software	M18		
Hardware needs from ICSC				
An odd number (> 1) of nodes (VM are ok) with root access and TCP/IP ports open, with public IP. No particular requirements for CPU. At least 16 GB of RAM each node. 1 TB storage each node.				

WP4 - Activity #4 (Blockchain deployment on the ICSC Datalake)

WP title:	WP4: Blockchain deployment on the ICSC Datalake		
Start date:	T0	**End date**:	T0+24M
Participants:	INFN (Spoke 2)		
1. OBJECTIVES			
• Definition and PoC of the deployment strategies on the Datalake; • Report on the BC technology tracking and comparison of solutions			
2. INPUTS			
WP1, WP2, WP3, WP5			
3. ACTIVITY DESCRIPTION			
INFN-CNAF has experience on the deployment of BlockChain sw stacks on Cloud environments [1]. In fact, the role of blockchain technologies on a cloud infrastructure can be multifold: they can be deployed as a managed infrastructure service, on which users can store immutable events, or they can be deployed on-demand.			

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

- Use the facilities, made available by the ICSC federated resources, to automatically deploy blockchain application stacks on different cloud providers federated with the Datalake; adapt the deployment strategy of the application stack to the ICSC.
- Perform a technology tracking study, which evaluates the readiness of the blockchain solutions available on the market, in the FOOS community and from the ICSC partners portfolios. The report can serve as a guideline for future blockchain use cases within and beyond the ICSC.

4. OUTPUTS

Software			**Hardware**	
Definition		Del. date	**Definition**	Del. date
Documentation			**Others**	
DRL code	Definition	Del. date	**Definition**	Del. date
TN1	Report the deployment solutions on the Datalake, with initial evaluation of results	M12		
TN2	Technology tracking report on the blockchain solutions maturity	M24		

Hardware needs from ICSC

- O (10) VMs on the Datalake, for R&D and PoC utilization
- O (20 VCPU) on the ICSC
- O (1-10) TB storage on the Datalake

WP5 - Activity #5 (Architecture and algorithms for data processing)

WP title:	WP5: Architecture and algorithms for data processing		
Start date:	T0	**End date**:	T0+24M
Participants:	Spoke 2-3, Thales Alenia Space Italia S.p.A.		

1. OBJECTIVES

Simulation of state of art algorithms for processing of space-based sensor data for SSA application evaluation of the computational load. The simulator will also allow to collect information to support evaluation against the optimal split between on board and on ground processing. Indeed, 1) on-ground processing can be implemented as a refinement of the on-board solution in positive detections or as a low-latency backup refinement 2) on-board detection may relax the communication bandwidth to ground. The results of the simulations will also provide a reference to evaluate by analyses the impacts of the processing in more complex scenarios with multiple sensors processing.

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

2. INPUTS
WP1, WP2
3. ACTIVITY DESCRIPTION
- Task 5.1: Sensors and data typologies identification for SSA applications Investigate existing and upcoming sensors typologies and technologies for space based SSA applications, identify measurable space objects characteristics they can detect and the data typologies they generate ***- Task 5.2: State of art algorithms for sensor data processing investigation*** Selection of a representatives subset of sensors and identification of the state of art algorithms for data processing ***- Task 5.3: Sensor data processing Simulator Design and Test*** Design and implementation of a mock-up (simulator) (on workstation) which is capable to generate a synthetic data set for the selected sensors, implement the algorithms of the processing chain and evaluate the computational load.
4. OUTPUTS

Software			**Hardware**	
Definition		Del. date	**Definition**	Del. date
Documentation			**Others**	
DRL code	Definition	Del. date	**Definition**	Del. date
TN1	sensors technologies and data typologies report	M8		
TN2.1	Algorithms and simulator design and verification report	M18		
TN2.2	Update - Algorithms and simulator design and verification report v.2	M24		

***3.2* Human resources** *(for each partner):*

- *FTE effort of junior (to be hired: MSc, PhD, Postdoc):*
- *FTE effort of senior (critical mass researchers):*

- INFN:
  - WP1 (on Spoke2):
    - 2 PM/y (staff)
    - 1 FTE x 2 years INFN to be hired
  - WP4 (on Spoke2):
    - 4 PM/y (staff) of INFN-CNAF staff
    - 2 FTE to be hired

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

- INAF:
  - WP3 (on Spoke2):
    - 4 PM/y (staff)
    - 0.7 FTE x 2 years to be hired
  - WP2 (on Spoke3)
    - 4 PM/y (staff)
    - 1 FTE x 2 years to be hired
- Leonardo:
  - WP1 (on Spoke2):
    - 0.5 FTE x 2 years
  - WP2 (on Spoke3):
    - 0.5 FTE x 2 years
- TASI:
  - WP1 (on Spoke2):
    - 0.4 FTE x 2 years
  - WP2 (on Spoke3):
    - 0.4 FTE x 2 years

3.3 Expected budget (total cost k€):

Use of the excel template for breakdown of the project's costs is mandatory.

Please, indicate in this section a short explanation of the budget (if not only personnel costs).

For WP3, on INAF-Spoke2, a budget of 100 k€ is dedicated to the purchase of a software license for two years. The software is a legacy blockchain solution for asset mutation tracking, with custodian vault for private keys and REST APIs to integrate in any software pipeline. This purchase will speed up operations, reduce the cost of personnel for development bringing in a production engineered software.

4. Other informations

- **HW/SW requirements**:*Computing site/kind/time, if any*
  - *See tables above*

- **Synergies with Other PNRR M4C2 projects** *(if applicable)*
- **Synergies with other national/international proje**cts (*(if applicable)*

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

- **If needed for a specific Spoke, add any other relevant information** *(es. Preferred development site, Cross-spoke relations:is the Ppproject cross-site? If yes explain how it unfolds)*

Annex:

I soggetti privati possono ricevere finanziamenti nel rispetto della normativa sugli aiuti di Stato

Ambiti finanziabili	Soggetti che svolgono attività non economica	Soggetti che svolgono prevalentemente attività economica			Soglie di notifica per soggetti che svolgono attività economica
		Grande impresa	Media impresa	Piccola impresa	
Ricerca fondamentale	100 %	100 %	100 %	100 %	40 *mln* € per impresa e progetto
Ricerca industriale *(Maggiorazione per collaborazione e/o diffusione)*	100 % (0 %)	50 % (15%)	60 % (15%)	70 % (10%)	20 *mln* € per impresa e progetto (15 *mln* € e per impresa e progetto)
Sviluppo sperimentale *(Maggiorazione per collaborazione e/o diffusione)*	100 % (0 %)	25 % (15%)	35 % (15%)	45 % (15%)	15 *mln* € e per impresa e progetto
Studi fattibilità	100 %	50 %	60 %	70 %	7,5 *mln* € e per studio
Investimenti in infrastrutture di ricerca	100 %	50 %	50 %	50 %	20 *mln* € per infrastruttura
Aiuti all'innovazione a favore di PMI (Maggiorazione per agevolazione relativa a costi consulenza per 200 *k* € in 3 anni per beneficiario)	0 % (0 %)	0 % (0 %)	50 % (50 %)	50 % (50 %)	5 *mln* € per impresa e progetto
Formazione (Formazione per disabili o svantaggiati)	100 % (0 %)	50 % (10 %)	60 % (10 %)	70 % (10 %)	2 *mln* € per progetto di formazione
Aiuti a Poli – Aiuti all'investimento	100 %	50 %	50 %	50 %	7,5 milioni di EUR per polo
Aiuti all'investimento – ex. Art. 107 paragrafo 3 lettera c Aiuti all'investimento – ex. Art. 107 paragrafo 3 lettera a	0 % 0 %	5 % 15 %	5 % 15 %	5 % 15 %	
Aiuti ai Poli – Aiuti al Funzionamento	100 %	50 %	50 %	50 %	

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf