EXECUTIVE SUMMARY

E-Infrastructures (i.e. the comprehensive ICT infrastructures that are needed to enable the complex, multi-disciplinary and globalised practice of modern science), with their resources and services are at the core of current changes towards Open Science and Open Innovation. They represent a global phenomenon in which Europe plays a key role through the funding of pillars such as the European Open Science Cloud (EOSC) and the European Data Infrastructure (EDI).

The “e-Infrastructures Project Area” collects the research contributes of scientists affiliated with Institutes of the CNR Department of Engineering, ICT and Technologies for Energy and Transportation (DIITET) in this context. These contributes spans across e-infrastructures in the area of Network, Computation and Data and along all the dimensions introduced by the recently published EC Staff Working Document on Implementation Roadmap for the European Open Science Cloud (i.e. Architecture, Data, Services, Access & Interfaces, Rules, and Governance).

The largest engagement is on topics related to the shaping and development of powerful and effective e-Infrastructures offering data management services. Part of the activities in this context concern the design of generic enabling facilities for data infrastructures and for the EOSC federation while others are related with domain specific solutions within Research Infrastructures facilitating scientists in operating in their respective fields. All these activities foster Findable, Accessible, Interoperable and Reusable (FAIR) data management.

The above research activities on data related infrastructures are nicely complemented by others on outstanding national project in network infrastructures and on distributed computing for Big Data.

As a whole, the institutes participating in this Project Area (ISTI, IIT, ICAR, IMATI, IREA) exhibit a good expertise that can be mobilized in supporting inter-department, national and European initiatives on e-Infrastructure, Research Infrastructures, Open Science and Open Innovation.

1. STATE OF THE ART OF THE RELEVANT SCIENTIFIC AREA

The dramatic increase in the amount and variety of data being produced and made available opens new scenarios to carry out research, develop and implement public policies, and expand the market with innovative products and solutions. In parallel, the emergence of the Open Science approach is changing the way in which scientific research is carried out, fostering a much more collaborative approach and new ways of diffusing knowledge.

E-Infrastructures\(^1\) are at the core of this change: they offer functionality that support communication and knowledge production workflows, from the basic data storage and curation, to data access, exchange and

analytics till data publication. They are set up to provide basic functionality that is common to more or less large, remotely disperse, community of practices. The functionality is offered “as-a-Service”, meaning that is made available by an online service operated by the e-Infrastructure provider. e-Infrastructures are completely transparent to their users. All the technical, organisational and operational tasks needed to deliver the service are hidden and up to the provider. These characteristics make e-Infrastructures enablers for addressing big scientific challenges, reducing costs, complexity and time required to develop and operate the necessary applications.

The role of e-Infrastructures as basic technological instruments able to boost innovation in the research, societal and economic contexts, was officially recognised by the Commission in its April 2016 Communication on the “European Cloud Initiative”\(^2\), as a part of the Digital Single Market Strategy. This Communication launched the **European Open Science Cloud (EOSC)** as a mean to enable data to be used throughout the value chain for scientific, societal and industrial purposes. A key concept of EOSC is to offer to every European researcher and citizen the possibility to access and reuse all publicly funded research data in Europe, across disciplines and borders. The same Communication laid the foundations for setting up a **European Data Infrastructure (EDI)**. EDI is envisaged to underpin EOSC by providing world-class HPC capability and high-speed connectivity as well as leading-edge services benefitting from them.

In the Commission’s vision EOSC is expected to be built as an open and evolving system-of-systems leveraging existing research supporting infrastructures and other resources, meaning with this generic network, compute and data infrastructures, domain specific research infrastructures (RIs)\(^3\) and thematic platforms, like for example, the Copernicus Data and Information Access Services (DIAS) offering services facilitating users in accessing Earth Observation data and information stemming from the Copernicus programme.

EOSC is also expected to (i) be able to support multidisciplinary and cross-disciplinary research, (ii) facilitate FAIR\(^4,5\) data management and (iii) enable cooperation and as early as possible sharing of research outcomes.

The realization of the EOSC vision is raising many new challenges in the e-Infrastructure related research sector. By exploiting the model proposed in the “EC Staff Working Document on the Implementation Roadmap for the European Open Science Cloud”\(^6\) endorsed by the EU Competitiveness Council of the EU\(^7\) these challenges can be organised according to six actions lines:

---

\(^2\) Communication: European Cloud Initiative - Building a competitive data and knowledge economy in Europe

\(^3\) RIs are domain-specific infrastructures called to provide their communities with facilities, resources and IT related services that are used to conduct top-level research in their respective fields.

\(^4\) **FAIR** is an acronym that stands for “Findable, Accessible, Interoperable and Reusable”.


\(^6\) Implementation Roadmap for the European Open Science Cloud

\(^7\) Draft Council conclusions on the European Open Science Cloud (EOSC)
[https://eoscpilot.eu/sites/default/files/council_conclusions_on_the_european_open_science_cloud_eosc.pdf](https://eoscpilot.eu/sites/default/files/council_conclusions_on_the_european_open_science_cloud_eosc.pdf)
1. **Architecture**: the EOSC architecture will be built as a federation of e-infrastructure, RIIs and other resources organised around a federating core and providing access to a wide range of publicly funded services supplied at national, regional and institutional levels, and to complementary commercial services. Implementing this vision requires, among the others, the identification of what are the optimal core services, how they can be offered at a sufficient level of quality, how interoperability at all levels can be achieved, and how to minimise the cost of using them by third-parties.

2. **Data**: EOSC will promote FAIRness oriented data management practices and tools. The EOSC settings calls for solutions dealing with the “big” dimension of data, in terms of volume and variety. Realising this vision requires the design and development of services and solutions (e.g. metadata, standards, mediators, preservation approaches) aiming at facilitating the findability, accessibility, interoperability, and reusability of data across the boundaries of the domain and communities they originate from.

3. **Services**: EOSC is expected to be a rich environment offering a wide range of services covering users’ needs. This requires to design and develop services and approaches providing researchers with innovative facilities improving the tasks they are called to perform in all the phases of a scientific life cycle: from the formulation of a research idea to the collection of data, their analytics and the FAIR “publishing” of the results.

4. **Access and interfaces**: EOSC will offer a broad range of mechanisms and interfaces for accessing its facilities including simple ways to deal with open data obligations and to access research data across different disciplines. This demand the design and development of a rich array of entry points for EOSC services access, e.g. an overall web-based portal, APIs to EOSC services, community tailored working environments.

5. **Rules**: Being a multi-stakeholder driven and evolving environment EOSC must be regulated by rules for participation for the different EOSC consumers and providers. These must take into account the diversity of actors in terms of role, location, maturity level, commitment level, and scientific discipline. This line includes models to represent these variety of rules, algorithms and solutions to automatically support the validation and monitoring of these rules.

6. **Governance**: EOSC needs a suitable governance framework aiming at ensuring EU leadership in data-driven science. This demands for mechanisms envisaging automatic representation, monitoring and reporting of KPIs, costs, activities, etc.

2. CONTRIBUTION TO THE RELEVANT SCIENTIFIC AREA

The contribution of DIITET Institutes to this PA is summarised below according to a matrix of competencies based on the six action lines driving the development of EOSC (cf. Sec. 1) and the three typologies of e-Infrastructures: **Network, Computing, and Data**.

<table>
<thead>
<tr>
<th></th>
<th>Network Infra</th>
<th>Computing Infra</th>
<th>Data Infra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture</td>
<td>IIT</td>
<td>ISTI, ICAR</td>
<td>ISTI, IREA, IIT, ICAR</td>
</tr>
<tr>
<td>Data</td>
<td>IIT</td>
<td>ISTI, IREA, ICAR</td>
<td>ISTI, IREA, IMATI, IIT,</td>
</tr>
</tbody>
</table>
IMATI is primarily working on Data Infrastructures. In particular:

- **Data**: the focus is on (i) definition and reuse of metadata and ontology models in the context of standardization efforts to ease data interoperability and information discovery; (ii) reuse and extension of Linked Data architectural pattern for mapping, connecting and indexing heterogeneous information from different sources; (iii) definition of methods addressing provenance provision for reproducibility and dependability.

- **Services**: the focus is on (i) 3D content-based multi-criteria search methods; (ii) data processing services; (iii) workflow composition and execution, to integrate heterogeneous software tools and to access and exploit a potentially large set of computational resources; (iv) methods to provide value-added information for applications (e.g. Medicine, Environment).

IMATI develops and maintains four e-infrastructures providing community tailored services: VVS supports research in visual data, i.e. on computational methods for the analysis, processing, visualisation, and reconstruction of 2D/3D digital representations. DRIHM and EXTras are based on the science gateway paradigm and rely on PRACE and EGI. GeCaRDC supports researchers in finding bibliographic resources.

ISTI is actively involved in the design and development of computing and data infrastructures:

- **Architecture**: It is leading the EOSCpilot project task called to design the architecture of EOSC. In doing that it is envisaging an architecture driven by few key principles, i.e. system-of-systems paradigm, as-a-Service provision, rules of participation regulating policies and procedures, FAIR principles. The proposed architecture model envisages a set of service typologies needed to comply with these key principles and to contribute to the implementation of the EOSC vision. ISTI is also leading the design and development of two pan-European infrastructures D4Science.org and OpenAIRE. D4Science.org is an infrastructure specifically designed to provide research communities with Virtual Research

---
8 Candela L et al. (2018) Initial EOSC Service Architecture. EOSCpilot Deliverable D5.1
https://eoscpilot.eu/content/d51-initial-eosc-service-architecture
Environments\textsuperscript{9}, i.e. web-based working environments customised to serve specific needs (services, data, capacity). OpenAIRE is contributing to the development of a global Scholarly Common supporting researchers, funders and other actors in research institutions to comply and monitor Open Access/Open Data/Open Science mandates in Europe.

- **Data and Services**: It is involved in the designing and development a rich array of services and approaches supporting all the phases of data management and implementing FAIRness and Open Science state of the art practices. Such services are contributing to two large software frameworks (i.e. D-NET and gCube) enacting, respectively, the implementation of D4Science.org and OpenAIRE. Services include (i) a data analytics platform\textsuperscript{10} specifically conceived to support researchers in sharing their analytics methods and tasks; (ii) a data integration and linking service\textsuperscript{11} supporting the development of a rich scholarly communication graph linking publications, datasets, people, organizations, projects, and funders aggregated from a variety of data sources; (iii) new generation repositories\textsuperscript{12} specifically conceived to support Open Science deposition.

- **Access and Interfaces**: It is involved in the definition and development of the concept of “Virtual Research Environment” as an approach to provide researchers with seamless and focused access to the facilities of their interest.

- **Rules and Governance**: It is involved in the EOSCpilot project activities called (i) to envisage the governance model of the forthcoming EOSC and (ii) to identify and develop a set of rules of participation for the various EOSC actors. In this role, it is conducting research on identifying what type of data models, algorithms, protocols and IT services are needed to automatically support the monitoring and assessment of these rules.

IREA mainly focused on tools and practices enabling FAIR data management and on open thematic services:

- **Data and Services**: IREA expertise is enhanced by its involvement in well-recognised ESFRI RIs, i.e. LifeWatch, EPOS and LTER, where all EOSC dimensions and recommendations are applied and tested. LifeWatch has been recognized as LifeWatch-ERIC (2017) and include Italy (Service Center of the e-infrastructure) as major pillar. In this context CNR has built the national Joint Research Unit LifeWatch-ITA. LTER is the global network of Long-Term Ecological Research sites. Its European branch, LTER Europe, is in the ESFRI roadmap to become an ERIC aimed at building its distributed and interoperable data infrastructure. EPOS is developing a pan-European research infrastructure in the field of Solid Earth Science. EPOS involves 24 countries and is included in the ESFRI roadmap to

---


become an ERIC, hosted in Italy, in 2018. At national level, the community is organized in a JRU, set up in 2017 and coordinated by INGV. The implementation of EPOS is co-financed by the EC (H2020 EPOS-IP). In EPOS-IP, IREA is coordinating the development of the **Thematic Core Service Satellite Data**

Within these RIs IREA research focus is on: infrastructures components; thesauri and ontologies; standard catalogue services; interoperable sensor services; data quality methods and services; modelling and GIS services; systematic and on-demand processing services for Earth Observation data.

**ICAR** performs activities mainly centered on the service action line applied to distributed computing for big data:

- **Services:** ICAR focusses on the development of a system based on a community of cores that is inspired by the collaborative model used in BOINC while implementing an ad hoc rewarding strategy similar to Bitcoin mining. As this solution does not require in principle any specific user skills (participants can join the network by simply providing their underused computational resources) the approach can be seen as a hybrid crowd as tasks can be solved by computer-based resources. ICAR researchers are also quite active in projects for IoT networks building and analysis. They have relevant skills on Data Mining, Machine Learning and Behavioural Analysis of huge amount of data coming from distributed environments.

**IIT** is primarily active in dealing with network infrastructures service development:

- **Architecture:** IIT is dealing with the design of the VoIP infrastructure of Tuscany Region, which allows Public Administrations to phone each other via Internet using the same numbers of PSTN. Moreover, IIT is cooperating with the DCSR5I-CNR\(^{13}\) to plan the evolution of the CNR network. Regarding network infrastructures design, evolution and monitoring, IIT is performing activities related with Rules and Governance domains by participation to: (i) IPv6 Italia, whose mission is to provide technical leadership and innovative thought for the successful diffusion of IPv6 in Italy, and (ii) GARR WGs, where new networking services and technologies are studied and experimented.

- **Data and Services:** regarding network infrastructure services, IIT is involved in (i) developing 6MoNPlus, a SW collecting router-advertisements and discovering IPv6 multicasts, ARP and DHCP broadcasts to neutralize rogue IPv6 routers and DHCP servers and (ii) Botnet discovering where the main goal is to determine how botnets spread out over the Internet and to collect, analyze and classify malware samples. Monitoring activities are based on honeypots.

IIT is also active in network and data infrastructures with studies and innovative technological solutions for the realization of complex Internet services: (i) the **.it Registry** is an international service with a very complex architecture. Great emphasis is given to the need to have a reliable, trustworthy, resilient and competitive system; (ii) a study on Internet diffusion is carried out, which uses the

---

\(^{13}\) “Direzione Centrale Supporto alla Rete Scientifica e alle Infrastrutture”, the executive office of the National Research Council of Italy responsible for the Institution infrastructure.
“domain name” endogenous metric, a unique methodology at a European level, based on the “Penetration Rate”. Diffusion is described at a national, macro-area, regional and provincial levels; (iii) a project is running to monitor the quality of .it DNS in order to create an integrated system able to point out abnormalities concerning the supply or use of the DNS service; (iv) an “in house” crawling system with a semantic engine has been developed to analyse, in real time, the Internet diffusion of the Agrifood, ICT and Tourism sectors in Italy.

The researchers involved in the describe PA also participate actively in several international working groups and initiatives aiming at identifying possible solutions, approaches and standards to be applied in the e-Infrastructure development, like W3C working groups (e.g. W3C-DWBP - Data on the Web Best Practice14, W3C-DXWG - Data Exchange Working Group15), RDA16 working groups, and INSPIRE17 working groups.

3. IMPACT

The CNR researchers involved in the e-Infrastructure PA are playing key roles in shaping the development of the forthcoming EOSC as a whole as well as of its constituents (e-Infrastructures and RIs). By leading the EOSC system architecture design in the EOSCpilot, ISTI is largely contributing to clarify and framing how it will be structured. On a more focused area IMATI, by participating in the EGI User Community Board, is contributing to delineate the EGI’s production infrastructure, to define usage policies and to prioritise requirements and issues. By improving and enriching the capabilities of e-Infrastructures and thematic Research Infrastructures the activities of the Project Area have also a large impact on different sectors of our society. The OpenAIRE infrastructure, for example, is now (June 2018) providing access to 24+ Million of publications, almost 600K datasets from 11,785 repositories and Open Access journal and it is supporting the European Commission and other funders in monitoring the Open Access mandate. The D4Science Infrastructure already showcases the potentiality of the forthcoming EOSC by supporting scientific and societal challenges in contexts like agro-climatic and economic modelling, food safety risk assessment, food security (AGINFRAplus), sustainable growth in the marine and maritime sectors, marine spatial planning, aquaculture economy (BlueBRIDGE), environmental data management (ENVRIplus, RI cluster), history, language studies, cultural heritage, archaeology, and related fields across the (digital) humanities (PARTHENOS, RI cluster), and smart cities, human mobility, poverty indicators, spatial analysis macroscopic human flows, etc. (SoBigData). In June D4Science operate 129 Virtual Research Environments


16 Research Data Alliance (RDA) www.rd-alliance.org

17 INSPIRE http://inspire.ec.europa.eu/
(VREs), serving 5000+ users from 44 countries with a 99.8 service availability. A number of agreements with international organisations, like FAO, are being signed to use what has been developed in the context of international fisheries and marine biodiversity monitoring activities.

Other societal sectors are addressed by IREA by contributing to development of LifeWatch and LTER RI through projects such as RITMARE (the Italian flagship Project in marine research), NextData (Italian Project of interest), eLTER (H2020 Long-Term Ecosystem Research in Europe), and ODIP (H2020 Ocean Data Interoperability Platform). Moreover, IREA’s activities in EPOS RI have important effects on the EOSC galaxy (namely, H2020 EOSC-hub and OpenAIRE-advance projects), and the ESA’s Geohazards Exploitation Platform project and DIAS initiative.

Some of the services and activities falling under the PA umbrella have also an important impact at National level. For instance, IREA plays the role of Center of Competence of the Italian Civil Protection Department and has established an agreement with MISE to monitor surface deformations affecting the Italian territory with particular emphasis on sites interested by activities to store or extract hydrocarbons. IIT is assisting the national level of Italian public administrations in managing and updating the networking infrastructure and the diffusion of the IPv6 protocol. Moreover, IIT is operating the .it Registry that (a) currently counts more than 3.1 million domain names and, in terms of number of domain names, it is the 6th Registry at European level; and (b) has more than 1,250 contracts with ISPs, Telco operators and public and private bodies. Finally, IIT has established a collaboration with the Italian Defense General Staff (Stato Maggiore delle Difesa).

The research activities described so far naturally map and interrelate with others related to the development of thematic Research Infrastructures performed in different CNR Departments, like the “Earth system science and environmental technologies”, the “Social sciences and humanities, cultural heritage (DTU) and the “Biomedical Science” ones.

Through the above described activities CNR can also actively contribute to the planned ICDI (Italian Computing and Data Infrastructure). This is a forum of major Italian Research Infrastructures and e-Infrastructures representatives, with the aim of promoting synergies at the national level and of harmonizing the Italian contribution to European and global challenges in various initiatives, including EOSC, EDI and HPC. In the long term, the vision aims of the forum is to create a national coordination body that could represent Italian infrastructures, being capable to interact with national and European initiatives and effective in facilitating the Italian participation to their design, governance and operation.

4. EMERGING RESEARCH CHALLENGES

The effective and massive use of e-Infrastructures and RIs across dsectors opens completely new application scenarios but also underpin an ample variety of research challenges. The Architecture of the overall landscape is still unclear. e-Infrastructures and RIs are evolving, new ones are emerging. How can all the provided services be used seamlessly, how they can be used across infrastructures boundaries in complex workflows and how these workflows can be executed by respecting their terms of use, trust, security and service time requirements are relevant open questions. Regarding data the challenges are even more complex. How infrastructures should ensure an effective and mostly automatic implementation of FAIR principles and, especially, how to fully support data products traceability and reproducibility are among the many open questions. The availability of data stored in repositories or
automatically collected by observing infrastructure users' behavior opens to an unlimited number of new services for knowledge extraction. Infrastructures are expected to offer transparently the capabilities for such, often complex extraction, based on a plethora of new paradigms and approaches.

5. CONCLUSIONS (2000/2200)

The CNR Institutes participating in the e-Infrastructures PA (i.e. ISTI, IIT, ICAR, IMATI, IREA) address complementary aspects of the Infrastructures research and development. The major involvement is on research related to data infrastructures, be them generic ones or embedded in thematic research infrastructures. Other activities pertain national projects in network infrastructures and on distributed computing for Big Data. The research groups involved witness the lively activity of the PA, with:

- the participation/coordination of twenty-five ongoing projects at international and national level;
- more than twenty hw/sw facilities, including clusters, servers, portals, platforms, FAIR enabling services, networking enabling architectures and applications, made available inside and outside CNR;
- an effort of eighty-three people;
- unique national facilities such as the .it Registry and the P-SBAS DInSAR Processing Chain (for Civil Protection).

These on-going research activities place DIITET in a relevant position for designing and building Networking, Computation and Data infrastructure services tailored for public and private bodies, as well as to serve in consultancy, evaluation and promotion of initiatives in the field. Through these activities the PA can also contribute to the Open Science challenge, to EOSC and to any national initiative in this area that may emerge in the future (e.g. ICDI). By increasing the quality, capacity and functionality of the infrastructures these activities also facilitate a more extensive and systematic exploitation of infrastructures in addressing a variety of societal challenges (e.g. civil protection, monitoring of environmental and economic indicators, food production assessment and monitoring) and in supporting Open Innovation (i.e. the early exploitation of research results in the Industrial sector).

The potential of the described activities is largely amplified by the multidisciplinary nature of CNR. There are many other activities in other PAs of the same and of other CNR Departments that can benefit and can complement what has been presented. A strict synergy and cross-fertilization among all these activities would indeed allow CNR to increase its key role in many National and International contexts.
**PROJECT AREA 4: E-INFRASTRUCTURE**

**Editorial team and Contact person (CP) Institute Email**

<table>
<thead>
<tr>
<th>Contact person (CP)</th>
<th>Institute</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>CARRARA PAOLA (CP)</td>
<td>IREA</td>
<td><a href="mailto:carrara.p@irea.cnr.it">carrara.p@irea.cnr.it</a></td>
</tr>
<tr>
<td>CASTELLI DONATELLA (CP)</td>
<td>ISTI</td>
<td><a href="mailto:donatella.castelli@isti.cnr.it">donatella.castelli@isti.cnr.it</a></td>
</tr>
<tr>
<td>D'AGOSTINO DANIELE</td>
<td>IMATI</td>
<td><a href="mailto:daniele.dagostino@ge.imati.cnr.it">daniele.dagostino@ge.imati.cnr.it</a></td>
</tr>
<tr>
<td>GEBREHIWOT ABRAHAM</td>
<td>IIT</td>
<td><a href="mailto:abraham.gebrehiwot@iit.cnr.it">abraham.gebrehiwot@iit.cnr.it</a></td>
</tr>
<tr>
<td>GIANNINI FRANCA</td>
<td>IMATI</td>
<td><a href="mailto:franca.giannini@ge.imati.cnr.it">franca.giannini@ge.imati.cnr.it</a></td>
</tr>
<tr>
<td>MANUNTA MICHELE</td>
<td>IREA</td>
<td><a href="mailto:manunta.m@irea.cnr.it">manunta.m@irea.cnr.it</a></td>
</tr>
<tr>
<td>MARTINELLI MAURIZIO</td>
<td>IIT</td>
<td><a href="mailto:maurizio.martinelli@iit.cnr.it">maurizio.martinelli@iit.cnr.it</a></td>
</tr>
<tr>
<td>MASCARI ELIO</td>
<td>ICAR</td>
<td><a href="mailto:elio.mascari@cnr.it">elio.mascari@cnr.it</a></td>
</tr>
</tbody>
</table>