



FROM

VOLUME

TO VALUE

SUMMARY

1. ■	INTRODUCTION TO THE NEW EDITION	5
1.1.	SCOPE, VISION, OBJECTIVE: THE ASSOCIAZIONE BIG DATA AS A SOURCE OF COMPETITIVE ADVANTAGE	7
1.2.	ACTIVITIES FROM 2015	9
2. ■	THE BIG DATA ASSOCIATION	11
2.1.	PROFILE OF THE ASSOCIATION	12
2.2.	THE ASSOCIATION IN FIGURES (2015-2019)	13
2.3.	HIGH PERFORMANCE COMPUTING FACILITIES IN EMILIA-ROMAGNA	15
2.4.	ASSOCIATION MEMBERS	16
3. ■	ENABLING INFRASTRUCTURE	17
3.1.	HPC, HTC & NETWORKING	18
3.2.	OTHER RELEVANT FACILITIES	21
3.3.	FUTURE TRENDS AND DEVELOPMENT	23
4. ■	SCIENTIFIC AND APPLICATION DOMAINS	25
4.1.	DIGITAL ENABLING TECHNOLOGIES	27
4.2.	HEALTH AND LIFE SCIENCES	29
4.3.	AGRI-FOOD, BIOBASED INDUSTRY AND BLUE GROWTH	31
4.4.	SMART CITIES AND SECURITY	33
4.5.	MATERIALS SCIENCE	35
4.6.	INDUSTRY 4.0	37
4.7.	EARTH SCIENCE AND CLIMATE CHANGE	39
4.8.	GLOBAL SOCIAL SCIENCES AND ECONOMICS	41
4.9.	HUMANITIES, CULTURAL HERITAGE	43
4.10.	PHYSICS, ASTROPHYSICS AND SPACE SCIENCE	45
5. ■	EUROPEAN PROJECTS LIBRARY	47
6. ■	CONTACTS	53

Palma Costi

Regional Minister for Economy and development, energy and green economy, post-earthquake reconstruction

Patrizio Bianchi

Regional Minister for Coordination of European policies for growth, education, vocational training, university, research and employment

With great pleasure we welcome this edition of From Volume to Value.

In 2016 we set up the research group on big data and supercomputing, and we started by checking the infrastructures and competences on the territory of the Emilia-Romagna region.

Emilia-Romagna has always been a crossroads. Here almost a thousand years ago, after dark centuries, the first "students" eager for new knowledge found themselves in Bologna, founding that institution which still today remains the crossroads of many university experiences.

Here, over the years, the supercomputing infrastructures of the national scientific system have been located and from here, we started to set up an association named Associazione Big Data that represents a place of convergence of skills and knowledge relating to big data, their applications, their impact on the daily lives of all citizens.

We are honored to have contributed to its birth: now the association is open to the whole country, aware of having to contribute to the full integration of our scientific system in the great European and world community of knowledge.



1

INTRODUCTION TO THE NEW EDITION

The management of large amounts of data via high computing capacity in terms of performance and memory available, provides remarkable opportunities in various research areas and applicative domains such as, among others, financial analysis, environmental monitoring and geophysical simulations, cultural heritage management, precision farming, multimedia, etc.

Many world-class institutions in supercomputing and big data are located in Italy and in the Emilia Romagna region. They are able to store, manage and analyze large amounts of data. Therefore, the impact of a joint exploitation of the expertise, know-how and facilities of the national and regional public and private actors is very relevant.

With the aim to exploit such potential, the "Associazione Big Data" was recently launched. Based in Emilia Romagna, but mobilizing the best world class actors of the whole country, it aims at:

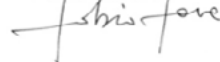
- Sharing and jointly exploiting existing results, knowledge, capacities, and research and innovation initiatives and frameworks;
- Fostering cooperation between national and regional public and private actors to also maximize

the leverage effects of research investments already used at regional, national and EU levels;

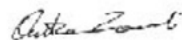
- Promoting joint actions including the coordination, planning and programming of relevant research and innovation activities in the big data domain;
- Supporting researcher careers, training and mobility, and developing skills in relevant sectors to ensure the necessary highly qualified workforce needed to underpin an effective and sustainable exploitation of big data analytics in research and industry.

The document collects figures, expertise, technologies and facilities available in the frame of the association in each of the knowledge and innovation domains of major relevance of the Associazione Big Data.

The president
Prof. Dr. Fabio Fava



The vice president
Prof. Antonio Zoccoli



1.1. SCOPE, VISION, OBJECTIVE: THE ASSOCIAZIONE BIG DATA AS A SOURCE OF COMPETITIVE ADVANTAGE

High performance computing, big data analysis, deep and machine learning algorithms, high bandwidth networks are technological pillars of digital society. Thanks to them, a wide range of future scientific, technological and societal challenges can be met to pursue growth in many sectors of the economy and society. The opportunities are many and yet not all predictable. These challenges will require the development of new skills and of new technologies.

Large investments in ICT infrastructures and technologies, such as high performance computing, cloud computing, artificial intelligence and big data, as well as in research and innovation, inclusiveness and skills are the priorities for Europe in order for it to benefit from the digital economy potential.

In Italy, research bodies such as INFN, CNR, ENEA, GARR and CINECA have already implemented big data via high performance computing and e-infrastructures to support major research and academic communities. Most of these Italian high performance and high throughput computing resources are concentrated in the North of Italy, especially in Emilia-Romagna, where they are well integrated into the local knowledge and business systems, consisting of universities and research centers, large-cap and mid-cap companies, as well as international players that are new entries in the local scenario. The data centre of the

European Centre for Medium-Range Weather Forecasts (ECMWF), which is going to be set up in Bologna, will bring other prominent computing and analytics capabilities, with the potential of further exploiting available expertise, infrastructures and facilities, by generating new local economic and societal opportunities.

Since 2015, stakeholders involved in supercomputing and big data production and management have engaged in mapping and analyzing their potential impact on scientific, social and business domains: thus, the Associazione Big Data was set up to interconnect and jointly exploit the knowledge, capacities and research and innovation potentials of this community to leverage the effects of actions and investments made so far and to maximize their impacts, locally but also at national, EU and international levels.

The "Associazione Big Data" mission is to facilitate:

- the sharing and exploiting of existing results, knowledge, capacities, research and innovation initiatives and frameworks;
- cooperation between public and private entities, maximizing the leverage effects of R&I public and private investments;
- joint actions including coordination, planning, facilitating and promoting relevant R&I actions, policies and international cooperation activities;



- researcher careers, training and mobility and, in general, the development of skills in the big data and AI domains.

More precisely, the Association shall operate in the following economic and societal sectors and disciplinary contexts:

- the economic system, with special attention to mid-cap companies, in order to facilitate its access to big data processing and related analytic tools, high bandwidth networking and high performance and high throughput computing in order to increase competitiveness and stimulate job creation and economic growth.
- the public administrations, in order to support awareness and implementation of well-developed data and computing based infrastructures to provide better services to citizen in the education, health, environment, mobility and security, in order to foster a more widespread, effective and citizen-friendly e-government system at national and local level.
- the knowledge community, in order to enhance its ability to analyze large amount of data from experiments, observations, monitoring and sensing systems and to process more detailed and complete data series to assess or gain insight of the investigated phenomena.

1.2. ACTIVITIES FROM 2015

The Big Data Community started its operations informally in 2015 working on a common framework including experiences, skills, facilities and opportunities under the auspices of Emilia-Romagna Region, and in the context of the regional policy for research and innovation infrastructures.

Since that time, the community enlarged the field of action from regional to national and international dimension, and a number of significant achievements have been reached by the members of the Big Data Community.

On September 2016 an all-Italian partnership between ENEA and CINECA was selected by EUROfusion, the European Consortium for the development of fusion energy, to deliver high-performance computing and data storage to support European research on fusion for two years. The 29 member countries of the Consortium are using a partition of MARCONI, the Italian supercomputer hosted by the CINECA. This machine has replaced the Japan-based supercomputer HELIOS, set up in the International Fusion Energy Centre in Rokkasho.

In December 2016, the regional government of the Emilia-Romagna Region submitted the Italian proposal for the relocation of the ECMWF Data Centre in Bologna. It was then announced as the winner in June 2017. The relocation site is located in the area of the Tecnopolo di Bologna Big Data hub owned by the Regione Emilia-Romagna. The Big

Data Community greatly supported the preparation of the proposal, sharing data and information that improved the attractiveness of the proposal as a whole. In May 2018 the Italian government announced the selected proposals to create the new Italian Competence Centre foreseen in the Industry 4.0 national plan. Among them was the proposal BI-REX: Big Data Innovation & Research EXcellence, with a cofinancing of 9.2 M€ and a large research partnership mostly gathered from the Community.

All the actors in the big data field that launched the "Big Data Community" formalised their partnership in an association called "Associazione Big Data". The charter of the association was signed by all the participants in June, 2018. The Association aims at promoting a community of research centers and infrastructures of excellence in the field of supercomputing (High Performance Computing) and the treatment of Big Data. Moreover, it provides the possibility of acting at the international level as a single subject.

Furthermore, in November 2018, some of the members participated in the H2020 call "HPC and Big Data Enabled Large-Scale Test-beds and Applications" with the project IO-Twins - Distributed Digital Twins for industrial SMEs: a big-data platform that has been successfully funded with a budget of 17M euros. The IO-Twins project will deliver large-scale industrial test-beds, leveraging and combining data related to the manufacturing and facility management optimisation domains,

coming from diverse sources, such as data APIs, historical data, embedded sensors, and Open Data sources.

In January 2019 the president of the association signed a memorandum of understanding with the Slovenian HPC Sling Association. The aim of the MoU is to collaborate through a number of possible activities, such as the sharing of access to digital infrastructure and establishment of partnerships for developing digital infrastructure; joint research and innovation projects and actions; research exchange programs; support for research groups with common activities or projects; enhancement of the capacity to attract funds and resources from third parties for the respective members at the national and/or international level; training, dissemination and outreach, joint summer schools, and any other activities deemed to be of mutual interest and agreed upon. It is hoped that additional international agreements will be signed in the next few months.

In 2019 the SUPER project will start, where SUPER stands for "Supercomputing Unified Platform". It is funded by the Emilia-Romagna ERDF's funds and will integrate CINECA and INFN platforms for first, followed by Tier1-CRESCO6-ENEA ENEA and the CMCC Supercomputing Center, with cloud access delivering and enabling value in multiple application domains. Once the infrastructure is available, a number of test beds will be developed in a subsequent "phase II" of the project.

The background of the top half of the page is a solid orange color with a faint, light-colored pattern of circuit board traces and circular nodes, resembling a microchip or data network.

2

THE BIG DATA ASSOCIATION

2.1. PROFILE OF THE ASSOCIATION

The Big Data Association operates in a wide ecosystem of initiatives carried out at the regional, national and European level, regarding Big Data and the relevant enabling technologies. At the regional level, the Association has an important role in the harmonisation of initiatives led by the stakeholders, e.g. the setup of the Tecnopolo di Bologna Big Data hub or the participation in BI-REX competence centre.

At the national level, all the members are involved in projects and initiatives related to Big Data, Artificial Intelligence and digital innovation, such as the Human Technopole Foundation, leading partner of the Human Technopole Project, aimed at developing personalised medicine and nutrition to tackle cancer and neurodegenerative diseases by means of genomics and big data analysis. Moreover, several members have provided support to industry in the framework of the national plan Industria 4.0 for innovation in Italian industry. Some member representatives are also personally involved in the national policy-making process, taking part in two High Level Expert Groups on Artificial Intelligence set up by the Ministry of Economic Development and by the Ministry of Education, University and Research, aiming at defining a national strategy for the applications of AI and at producing guidelines to inject AI into the educational process.

The Association's members are also involved in European and international

initiatives and projects as described below. It is worth mentioning that CINECA and INFN have a key role in the Italian participation in the European High-Performance Computing Joint Undertaking (EuroHPC), that will pool European resources to develop top-of-the-range exascale supercomputers for processing big data, based on competitive European technology. Another significant contribution is made to the European Open Science Cloud (EOSC), with participation in both hub and pilot projects. EOSC is the cloud infrastructure for research data in Europe, responsible for infrastructures and policy-making. All the initiatives and projects above are part of an important commitment by Europe to the fields of big data and artificial intelligence, as stated in the EU Declaration on Cooperation on Artificial Intelligence, signed in April 2018, aimed at coordinating national policies and at establishing cooperation between the adhering countries on the exploitation of AI skills, features and potential applications.

2.2. THE ASSOCIATION IN FIGURES (2015-2019)



Supercomputing resources
for public research
in Italy managed by
association members

>90%



H2020, FP7
or CEF projects¹

140



Total Cost of
the projects

> 2,1 BLN€



Funding received
from EC

160 MLN€

¹ Project funded by H2020, FP7 or CEF programs that were on 2015 or started after and involves at least one member of the association.

GARR-X

RESEARCH NETWORK

- Up to 200 gbps
- VPN for WLCG high energy physics community
- VPN for PRACE HPC Infrastructure
- VPN for Human Brain project FENIX Federating HPC Infrastructure

LEPIDA

THE EMILIA-ROMAGNA REGIONAL NETWORK OF THE PUBLIC ADMINISTRATIONS

- Up to 100 gbps capacity
- More than 140,000 km optical fibers and approx. 3,000 access nodes
- 43 Pop nodes + 4 integrated regional data centers.
- Active in 10 International Public Peering Exchange Points and in 4 Private Peering Facilities.
- Current coverage of fast internet availability in Emilia-Romagna: 76% of households
- Provision of dark fiber special link between CINECA and INFN-CNAF at 2 terabits /s.

2.3. HIGH PERFORMANCE COMPUTING FACILITIES IN EMILIA-ROMAGNA

CINECA

Tier0: MARCONI system. 3500 low latency server nodes + 3500 scale out server nodes; ~10 Petabyte of memory RAM; ~20 Petaflops peak performance

Tier0 in PRACE Partnership for Advanced Computing in Europe Eurofusion HPC service facility

Tier1: GALILEO system. 1000 low latency server nodes; ~2 petaflops peak performance

Cloud service: MEUCCI system. 250 low latency server nodes

Openstack virtualization middleware; containers, urgent computing

AI & ML: DAVIDE system. 45 hybrid nodes x 2 Power8 + 4Tesla P100; ~1 petaflops peak performance

STORAGE REPOSITORY: PICO system. 50 server nodes for visualization and data management services; multi tier storage repository;

~20 petabytes of net disk space;
~20 petabytes of tape library

A high-throughput computing (HTC) facility which hosts the WLCG Tier1 at the **CNAF-INFN** in Bologna, with the following capabilities:

CPU: ~30000 computing cores;

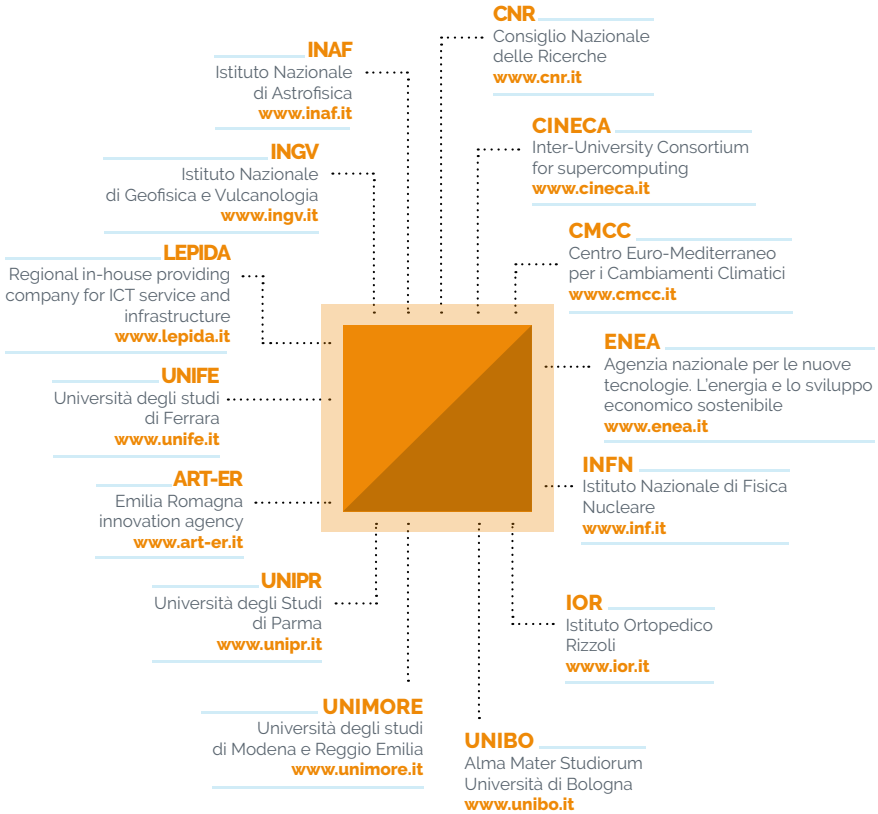
STORAGE: ~40 PB of online disk space;

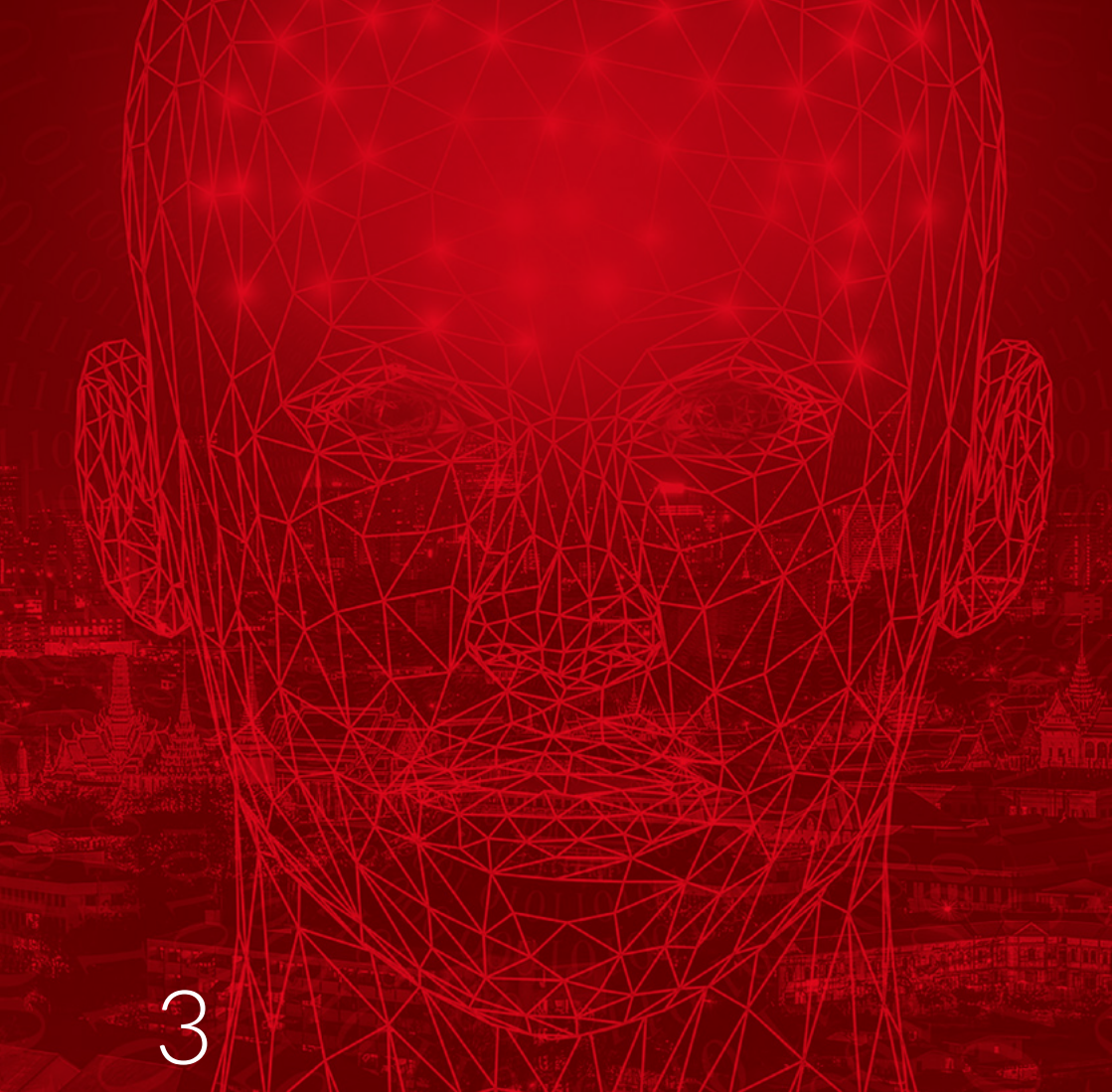
LIBRARY: ~90 PB of nearline tape space.

GARR-X/LEPIDA

Fast and effective nationwide and international network connection, mainly provided by GARR and Lepida.

2.4. ASSOCIATION MEMBERS





3

ENABLING INFRASTRUCTURE

3.1. HPC, HTC & NETWORKING

CINECA is the national supercomputing facility. Thanks to the long-term vision of the Ministry of Education, University and Research, which has persistently supported the progressive development of the CINECA Interuniversity Consortium, over time CINECA introduced in Italy the most innovative world class HPC system architectures. The first vector supercomputer was introduced in Italy in the 1980s, the first parallel HPC architectures in the 1990s, the first cluster at the beginning of the 2000s, the large massively parallel architectures, the first petascale system in Italy, ten years ago; and lastly, the modular cluster, the current multipetascale HPC system actually in production. At the time of their installation, most of the CINECA supercomputer systems were ranked in the top part, from the 7th to the 12th position, of the TOP500, enabling and supporting national and European excellence in science and technology innovation. The high end tier 0 systems have been continuously integrated in a complex environment that complements the tier 0 supercomputers with a high quality of service tier 1 system, a large scale data repository, and, from time to time, innovative prototypes to keep pace with the cycles of innovation and to develop and test new architectures with the objectives of increasing and improving the effectiveness and the efficiency of the next cycle highest performing computing system.

The current computing architecture facility hosted by CINECA integrates a

tier 0 high end computing system with a peak performance on the order of 20 petaflops, ranked number 12 in TOP500 at the time of its installation, a tier 1 system for quality of service, an HPC cloud system, and a prototype production system for artificial intelligence and machine learning applications. In all, the current computing architecture integrates more than 9000 server nodes. A large-scale data repository, with a full capacity in excess of 50 petabytes, completes and integrates the computing architecture as well.

The CINECA HPC facility enables a wide range of scientific research through open access granted by independent international peer reviewed processes by mean of the PRACE association at the European level, and the ISCRA at the national level.

CINECA provides services to the Worldwide Eurofusion community, being the contractor of the European tender for that specific service that will run till the end of 2023, and also provides operational computing service for weather forecasting for National Civic Protection under the supervision of Emilia Romagna ARPA-SMR. CINECA is part of the European digital infrastructure for many ESFRI RI facilities and initiatives, among others, EPOS: European Plate Observing System, led by INGV, ELIXIR: Infrastructure for Life Science, and the HPB European Human Brain Flagship Project. Also, with reference to those RI initiatives (but not only those), CINECA is a core partner of the European

Open Science Cloud HUB infrastructure and the European Data Infrastructure. At the national level, many joint development partnership agreements are in force with INFN, ENEA, INAF, SISSA, ICTP and many R&D collaboration actions and agreements are in force with qualified national research institutes, universities, and public administrations. CINECA received recognition as a 'Golden Digital Hub for Innovation' from the Big Data Value Association, is a Competence Centre for supporting innovation towards industries, and has led many proofs of concepts in collaboration with industries and private organisations. CINECA entered formal partnerships for added value services and R&D activities with ENI and manages and operates the ENI corporate supercomputing facility, one of the world's largest industrial supercomputing infrastructures.

The HTC facility is hosted at CNAF Bologna, which is one of the INFN National Centres defined in the INFN charter. CNAF has been charged with the primary task of setting up and running the so-called Tier-1 data centre for the Large Hadron Collider (LHC) experiments at CERN in Geneva. Nowadays it hosts computing not only for LHC but also for many experiments, ranging from high-energy physics to astroparticles, dark matter searches in underground laboratories, etc. CNAF also participated as a primary contributor in the development of grid middleware and in the operation of the Italian grid infrastructure. This facility operates within

the framework of a national INFN HTC infrastructure consisting of the CNAF Tier 1 and 10 smaller facilities, called Tier 2 centres, placed all over Italy.

The CNAF data centre operates about 1,000 computing servers providing ~30,000 computing cores, allowing the concurrent run of an equivalent number of general-purpose processing tasks. All the computing resources are centrally managed by a single batch system (LSF by IBM) and dynamically allocated through a fair-share mechanism, which allows full exploitation of the available CPU-production with an efficiency of about 95%. Part of the CNAF computing power is currently hosted at CINECA, and connected to the main site via dedicated high-speed network links traversing the city of Bologna, allowing the remote computing nodes to access the mass storage located at CNAF with maximum throughput and minimal latency, as if they were local.

CNAF operates a very large storage infrastructure based on industry standards, for connections (about 100 disk servers and disk enclosures are interconnected through a dedicated Storage Area Network) and for data access (data are hosted on parallel file systems, GPFS by IBM, typically one per major experiment). This solution allows the implementation of a completely redundant data-access system. In total, CNAF hosts about 40 PB of online disk space, with a total I/O bandwidth of about 1.5 Tb/s and 90 PB of nearline tape

space arranged in a robotic library read out by 20 tape enterprise-level drives. The 10 additional Tier 2 centres distributed across Italy have a similar aggregate of resources, both in terms of computing and disk storage, but not tape. They are connected to the CNAF Tier1 through dedicated high-speed links, and the resources can work in a coordinated way in the global INFN computing ecosystem.

The user community of the CNAF Tier1 facility is primarily composed by research groups from INFN and most of the Italian universities (including UNIBO, UNIFE, UNIPR), working on nuclear, subnuclear, astroparticle and theoretical physics. In the coming years, the HTC computing resources of INFN CNAF will be increased at least by an order of magnitude in order to meet the computing and storage requirements of forthcoming experiments and upgrades.

In addition, LEPIDA, within the framework of the Digital Agenda of the Emilia-Romagna, has been entrusted with the design, implementation and provision of four data centres, geographically distributed and natively connected by the LEPIDA network, for the use of public administrations and healthcare systems. These data centres offer advanced computing services, storage, data protection, business continuity with native disaster recovery functions and a specific focus on energy management.

3.2. OTHER RELEVANT FACILITIES

As previously mentioned, the concentration in Bologna of national infrastructures for HPC and big data constitutes the interchange and reference hub for the national and European public and private research community. The national panorama also presents a vitality of tier 1-class infrastructures and highly complex cloud computing services.

As shown schematically below, with positioning data, which, due to its continuous updating process could suffer from some inaccuracy, but which will represent the national panorama, the distributed presence of digital infrastructures is certainly valuable. Attention is drawn to the distributed infrastructure of the INFN sites, to the infrastructure implemented through the action of ReCaS (calculation network for SuperB and other applications), and to the distributed system for cloud computing services implemented by GARR. In addition to the distributed infrastructures, there are tier1 systems at the peta scale class at the Portici computer centre of ENEA, the CRESCO systems, at the University of Salento computer centre, which houses the CMCC supercomputing system, and the computing centre of SISSA / ICTP, which hosts the SISSA computing system. By mean of "SUPER – Supercomputing Unified Platform Emilia-Romagna" the HPC and Big Data national ecosystem will be federated to enable good practices of data access and processing based on FAIR (Findable, Accessible, Interoperable,

Reusable) principles and guidelines and to move towards the implementation of the digital single marketplace at the European level provided by the EOSC – European Open Science Cloud initiative.

National HPC and Big Data infrastructure

Institution	Location	Site Name	# of cores	# of server nodes	Storage (PB)	Tape (PB)
INFN	Bari	NFN-BARI T2	174	12	1,51	
	Bari	RECAS	4.000	250	3,55	
	Bari	GARR	1152	58	3,00	
	Bologna	CNAF-T1	30.000	1875		
	Bologna	CNAF-LHCB-T2	384	12	40,00	90,00
	Catania	INFN-CATANIA	390	20	1,20	
	Catania	RECAS	2.562	128	0,32	
	Cosenza	RECAS	3.500	175	0,90	
	Frascati	INFN-FRASCATI	150	8	,13	
	Milano	INFN-MILANO-ATLAS	180	12	1,31	
	Napoli	INFN-NAPOLI-ATLAS	334	21	1,82	
	Napoli	RECAS	4.956	310	4,96	
	Padova	LNL	452	28	1,76	
	Pisa	INFN-PISA	1.250	78	2,49	
	Roma	INFN-ROMA1	160	10	1,34	
	Roma	INFN-ROMA1-CMS	192	6		
Torino	INFN-TORINO	78	6	0,06		
Cineca	Bologna	MARCONI	406.000	7000		
	Bologna	GALILEO	72.576	1512	20,00	20,00
	Bologna	PICO	2.400	50		
CMCC	Lecce	CMCC	13.824	288	4,00	
CNR	Roma	Data Center	1.034	64	1,20	
ENEA	Casaccia	ENEA-Casaccia	192	12	0,02	
	Frascati	ENEA-FRASCATI	480	24	0,08	
	Portici	CRESCO4	4.864	04		
	Portici	CRESCO5	512	32	4,00	5,00
GARR	Portici	CRESCO6	20.832	434		
	Catania	GARR	1152	58	3,00	
	Cosenza	RECAS	384	32	1,00	
	Napoli	GARR	384	24	1,00	
INAF	Palermo	GARR	1152	72	3,00	
	Trieste	OATS	256	16	1,00	
SISSA	Trieste	SISSA	4.000	250	1,50	

3.3. FUTURE TRENDS AND DEVELOPMENT

The digital infrastructures, high performance computing architecture, hyper scale cluster architecture and high bandwidth network infrastructure will have a tremendous and drastically innovation in the next few years.

The global race towards exascale computing is imparting a strong acceleration to the deployment of supercomputing systems with a performance in excess of many hundreds of petaflops and the target for the availability of the first exascale computing system, a supercomputer with a performance of at least one thousands petaflops, is more and more close. The USA, China, and Japan are investing many billions of dollars for supporting R&D with the target of having an exascale system in production by 2022/23.

The European Union and 28 member states created a joint undertaking with the purpose of mobilising a budget of billions of Euros for R&D for the development of European HPC technologies and for achieving towards-pre-exascale and exascale systems to open production in Europe in the same time frame. The Italian federal government, the local ecosystem of universities and research institutes, CINECA and other agencies, and regional public administrations are collaborating and working for hosting in Italy one of those supercomputers, on the basis of the excellent track records of CINECA and of the vital persistent development in Italy of the methods of computational

science.

From this point of view, the development roadmap of CINECA foresees, as a baseline target, systems with dozens of petaflops in a short time and hundred-petaflop systems in the medium term. The medium term target will be definitively increased in case of success in acting as an EuroHPC Hosting Entity.

The needs and the objectives of capping the consumption of electrical energy of supercomputer systems has led to the development of microprocessor technologies characterised by an extremely narrow performance profile, such as floating point accelerators, FPGA components, tensor microprocessors, and neuromorphing chips, but an extremely high ratio of energy efficiency to performance if properly exploited at the level of the algorithms and programming models. Such components have to be integrated in server nodes that combine general-purpose microprocessor sockets and floating point accelerators in hybrid configurations that needs to be co-designed in light of specific application domains. The physics of the materials and matter, the environment and climate change, energy savings and energy production from renewable sources, the health and life sciences, artificial intelligence, as well as cyber security, are and will be the main drivers for the CINECA co-design approach, and co-design will be the main driver for the development roadmap of the CINECA supercomputing infrastructure towards

performances on the order of hundreds of petaflops and at the exascale. Regarding the large hyper scale cluster, the overlap between the low-power multi-core sockets and the software technology for virtualisation and containers pushes the limits of the very large resilient configurations towards multimillion-core clusters. This has led the scientific community and industry to develop new processing methods based on high performance throughput computing for filtering and gaining insight into the huge amount of data produced by extreme experiments or observations, such as the High Energy Lumen LHC experiments at CERN, or the Square Kilometre Array multi-radio telescope project, or the Internet of Things, in the context of industrial innovation processes and public administration decision support work flows.

From this point of view, as the INFN is one of the world's largest tier 1 LHC computing grids, the target is to maintain and improve this leading position and enabling the provisioning of services to consolidate the scientific competitiveness of the national and European community. The implementation of this strategy foresees a multimillion-core cluster at CNAF, complemented with a data repository capacity at the near-exabyte level. The sharing of data and the anywhere-anytime access model requires the availability of large bandwidth geographical networks to federate large

scale facilities producing and processing the increasing availability of big data. In the future, the LEPIDA physical fiber optic network integrated in the GARR-X service will guarantee state of the art high-speed network links, moving from the current tens of Gigabits per second to the scale of hundreds of Gigabits per second, till the next scale of up to 400 Gigabits a second backbone foreseen for wide availability in the timeframe 2021/2022.

In order to deploy on time such a breakthrough in all the dimensions of the digital infrastructure technologies, it will be necessary, starting now, to develop progressively the consolidation of the main hubs around which will be federated the large resources distributed in the different core centres of the national and European public and private research systems and the main core centres of the public administrations.

The aim of the SUPER regional infrastructural project will be to proceed in the direction of coordinating the resources distributed at the national level around the Bologna high performance computing and big data world class hub as a model for open science in Italy and in Europe and as an enabling facility to support the development of the proofs of concepts for innovation in the industrial sector and the optimisation of the decision processes of the public administrations.

4

SCIENTIFIC AND APPLICATION DOMAINS



4.1. DIGITAL ENABLING TECHNOLOGIES

Trends in digital innovation are progressively providing the ability to globally extend connectivity to any person and any device, to seamlessly access any information and data from anywhere and at any time, with on demand processing and analysis of the data for modelling scenarios and supporting the decisions (ATAWAD: anytime, anywhere, any device).

The leading edge components of such a pervasive ecosystem of digitally enabling technologies are the high speed networks to interconnect large scale facilities; large volume repositories to preserve and curate the big data, and high end (HPC and HPTC) computing systems. These technologies deal with simulation and modeling of scientific and engineering problems as well as data analysis. The scientific capabilities, industrial competitiveness and sovereignty depend critically on access to world-leading HPC computing and data infrastructures to keep pace with the growing demand and complexity of the problems to be solved.

As the problems modeled in computer simulations and decision support systems grow in size and complexity (to enable more detailed predictions, to cope with ever larger amounts of data, or both), so do the demands on digital resources. In many areas, spanning from health, biology and climate change to automotive, aerospace, energy and economics, digital technology provides a practical way to address complexity, and access to it becomes essential.

To totally harvest the benefits of digital technology, it is necessary to support a full ecosystem comprising hardware and software infrastructures, applications, skills, services and their interconnections. Digital innovation research has a direct and major impact on the whole ICT and HPC value

chains, both for large national, European and international players and, at the level of regional ecosystems, for SMEs and innovative industries and creative startups, with big data modelling and processing market in hardware, software and ICT services a fast growing multibillion-euro business sector of industry.

In order to unleash the potential arising from the exploitation of big data value, it is essential to research new technologies, including HPC, data processing, data science and artificial intelligence.

Multimedia, data storage, data movement and networking, communication and visualisation tools are complementary enabling technologies for the development of the big-data market. Fairness in the access to public data and at the same time strong safeguards for privacy and cyber security are also essential.

Furthermore, a consistent long-term strategy is essential for educating and training professional data scientists to favor an holistic and integrated multidisciplinary and interdisciplinary problem solving approach and to manage policies and complexity to support the research infrastructures and industries in properly handling and gaining insight from big data.

Major European Projects: AI4EU, AIDA, AIDA-2020, ANTAREX, BD2DECIDE, DeepHealth, EnABLES, ENABLE-S3, EOSC-hub, EPEEC, EUDAT2020, EUROfusion, ExaNeSt, EXDCI, XDC, HERCULES, HPC-EUROPA3, ICEI, OPRECOMP, PRACE-4IP, SeaDataCloud, SECREDAS, Re-Search Alps, SEADATANET-II.

Active members: CMCC, CNR, ENEA, INAF, INFN, INGV, LEPIDA, UNIBO, UNIMORE, UNIPR, CINECA.



4.2. HEALTH AND LIFE SCIENCES

Big data are attracting a great deal of attention for the potential in healthcare and, more widely, in all biomedical scenarios. In the last few years, hospitals, universities and research centres have started fruitful and data-productive analyses at the European, national and regional levels. For example, omics studies and multidimensional imaging scans producing large volumes of data and knowledge are helpful in areas from neuroscience to orthopaedics, from cardiovascular diseases to aging. By this means, an increasing number of case studies in healthcare are well suited for a big data solution. In addition, big data allow solving clues in common conditions but also discovering new developments for treating rare diseases, steering healthcare towards personalised and "precision" medicine. Nowadays, there is a pressing need for fast and preferential connections for data transfer to transform potential into reality and instruments for information aggregation and extrapolation are urgent, such as rapid and high-performance techniques for machine learning and data mining. Despite this strong impulse, the exploitation of big data information to generate new health knowledge is much delayed/hindered by the inherent heterogeneity of big data and the lack of broadly accepted standards, as well as by legal issues surrounding the use of personal data.

At the regional level, a proper and coordinated technology framework, taking care of context and metadata, will activate a data-driven improvement for a meaningful use of big data in healthcare development, boosted by top-level HPC infrastructures. A more accurate and defined management of large-scale information and local data exchange and integration will lead physicians, medical doctors and researchers to better treatments capable of answering patients' issues and, at the same

time, reducing costs for the national health services and/or regional health authorities.

Big data in the life sciences and medical practice are still a big challenge, but they will leverage the existing scientific knowledge, improving translational research, develop personalised strategies, and implement innovative technological products. All these efforts will have an impressive social, economic and industrial impact. In this perspective, a regional infrastructure, able to manage and process large-scale data, will allow running prospective population studies that will give an insight in treating epidemiologically significant diseases such as cancer, neurological and cardiovascular conditions, osteoporosis, muscle-skeletal diseases, as well as geriatric and ageing-related pathologies.

Major European Projects: ADOPT-BBMRI-ERIC, AirPROM, ChiLTERN, COMPARE, CORBEL, DeepHealth, ECDP, EDEN ISS, EJPRD, EU-STAND4PM, HPB SGA1, INforFUTURE, LANGELIN, MCDS-Therapy, MYNEWGUT, Neuromics, ORTHOUNION, PAPA-ARTIS, PLATYPUS, PROPAG-AGEING, THALAMOSS, VPH-Share.

Active members: CNR, CINECA, INFN, INGV, IOR, UNIBO, UNIFE, UNIMORE, UNIPR.



4.3. AGRI-FOOD, BIOBASED INDUSTRY AND BLUE GROWTH

The relevance of big data to the agri-food sector, bio-based industry, and blue growth is pivotal, also for the current blooming of innovative technologies and omic tools applied to the most diverse industrial sectors (food production and safety, primary production and animal/plant breeding, industrial biotechnology, enzyme and microbial discovery, etc.). Managing these (big) data is a formidable challenge. Improving consumer health by monitoring food-related data is one of the areas that may benefit most from a radically innovative use of big data to provide more personal recommendations, via various technological platforms, which can improve the quality of life. Beyond the typical use of data analysis for food safety, big data is also related to predictive analytics, with an impact on economy and logistics, as well as to metagenomics for the characterisation of food spoilage. EFSA is interested in future strategic approaches to risk assessment in areas of food and feed safety that benefit from the acquisition, processing, and sharing of large quantities of data and evidence. Therefore, EFSA supports this initiative and is open to supporting solutions that reduce the costs and improve the effectiveness of data acquisition by sharing resources and capabilities.

Big data has also an increasing importance in agricultural practices, since the integration of data from sensors, Internet of Things (IoT) applications and genomics can lead farmers to increase their productivity and sustainability. They can also improve efficiency and sustainability as well as flexibility and safety in the food and biobased industry.

Finally, IoT, cloud computing and big data and data analytics are required for sharing, advanced processing and

analysis of key data of interest for maritime security (including migration phenomena), maritime navigation and transportation safety and security, sustainable fisheries and aquaculture, biotic and abiotic sea resources exploitation, etc.

Major European Projects: CIRCLES, CYBELE, INMARE, LEGVALUE, MedAID, PANINI, PerformFISH, Prolific, RES URBIS, SMARTCHAIN, TREASURE, VALUMICS.

Active members: CINECA, CNR, INGV, UNIBO, UNIPR.



4.4. SMART CITIES AND SECURITY

Big data is a growing area of interest for public policy makers, for cities and urban management: it is related with the enormous stream of data coming from administrative and social data, from city energy, mobility and transport infrastructures, from large and increasing sensor networks, comprehending large nodes connected under the Internet-of-Things paradigm, video surveillance and environmental cameras, and so on.

The analysis of data for purposes of social innovation and human-centric services (e.g. personal safety and physical security, crowd-sensing/participatory sensing in an urban environment), for smart mobility and smart logistics services, for critical infrastructure protection, for emergency management, etc. is becoming crucial and challenging due to the magnitude of the data and the requirement of timeliness. Connected with smart city and social big data, cyber security (including privacy, access control management, biometrics, etc.) is one of the most critical areas in big data analytics and management. There's a growing demand for security information and event management technologies and services, which gather and analyse security event big data that is used to manage threats.

In transport, the volume of data has increased because of the growth in the amount of traffic (all modes) and detectors. Also, travellers, goods and vehicles are generating more data from mobile devices and tracking transponders (including trains, ships and aircraft). Infrastructure, environmental and meteorological monitoring also produce data that is related to transport operations and users. New ways to collect, manage and analyse vast quantities of data are important both for governments and private companies

offering transport services.

A smart use of big data supports governments in optimising multimodal transport and managing traffic flows, making cities smarter. Real-time transportation planning and safety, environmentally sustainable and resource-efficient transport, socio-economic and behavioural research, and forward looking activities for policy making are fundamental topics which greatly benefit from big data. As well, the huge amount of data coming from GPS systems, real-time traffic monitoring, parking availability, electric charge station availability, etc., can provide competitive advantages in order to optimise vehicle design, maintenance and energy management, improve on-board automation and safety, and reduce CO2 emissions.

Moreover, the widespread use of modern monitoring systems in electricity transmission, distribution systems and smart meters monitored by the consumer enables the acquisition of real-time, high-resolution data. This coupled with the addition of other data sources, such as weather data, usage patterns and market data, dramatically increases the need for appropriate big data handling techniques and machine learning approaches. Existing grid capacities could be better used, and renewable energy resources could be better integrated.

Major European Projects: Domino, FLEXMETER, FOCAS, ICARUS, MARISA, OSMOSE, PRYSTINE, SocialCar, SOGNO, TRAFAR.

Active members: CINECA, CMCC, ENEA, LEPIDA, UNIBO, UNIMORE.



4.5. MATERIALS SCIENCE

Data and metadata for the development of new materials, their processing, and application life cycles are considered key assets to accelerate discovery and innovation across all design and manufacturing sectors. With its academic and industrial knowledge base and the related production of data, Emilia Romagna competes with the most advanced regions in Europe and worldwide. Here, the main challenges are often related to the construction and maintenance of well-annotated and structured repositories, their accessibility, their specialisation/integration and interactions with similar European and international operations, as well as the development and deployment of data analytics strategies.

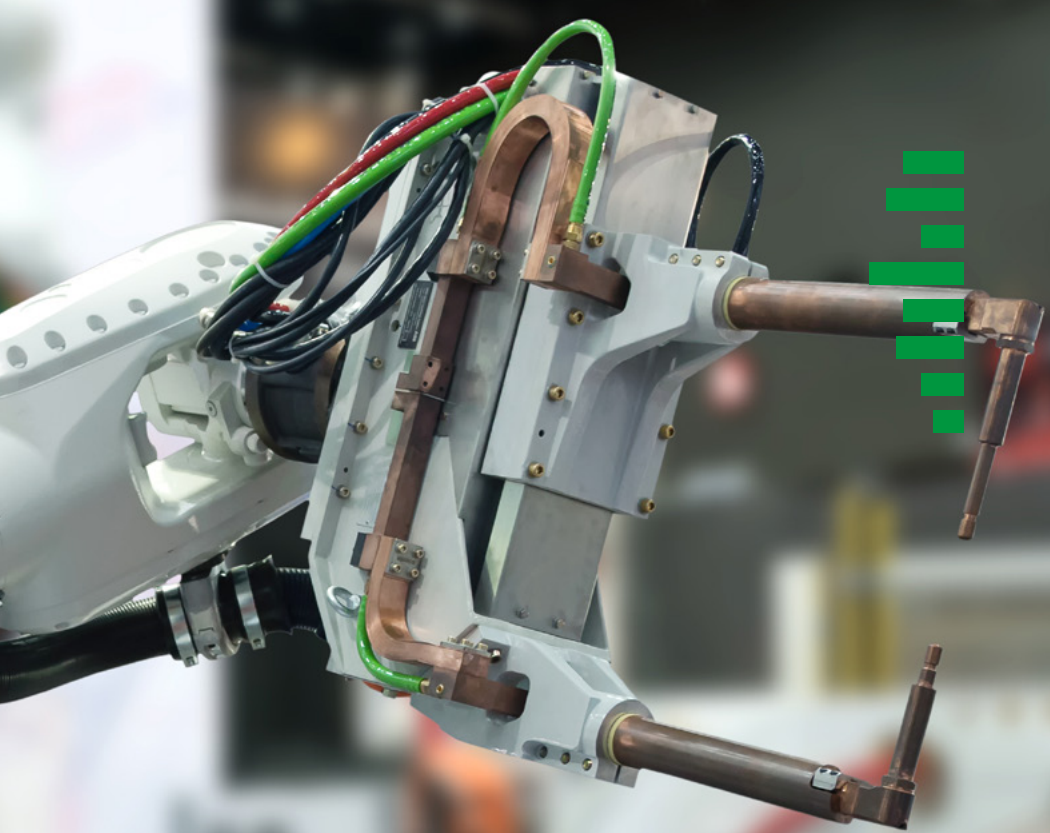
A special feature of the materials domain is the wealth of predictive information that can be obtained from simulations, now extending to quantum and multiscale approaches thanks to HPC and HTC. In the silico design of materials and (bio)molecular systems this is no longer limited to their structures and stability: it includes a huge range of functionalities, e.g. from friction and wear to biocompatibility, from colour and optical appearance to hydrophobicity, from electron transport to thermal properties within devices. Italy hosts an important community of advanced users of such applications, but also leads European efforts of code developers who work at the frontiers of the current and future HPC technologies, invest in software/hardware co-design, and are creating an ecosystem of capabilities, applications, data workflows and analysis, and user-oriented services. Importantly, Italy coordinates the European Centre of Excellence for HPC on materials design at the exascale, MaX.

Data acquisition from the advanced spectroscopy and imaging of materials

is increasingly important, and requires sharing and analysing the resulting highly distributed, heterogeneous data-sets. An example comes from the electron microscopy community: they produce multidimensional data that need to be stored, transferred, shared, managed and processed, both online (i.e. during acquisition) and offline. Efforts in this direction are going to involve broader communities working on automated materials and systems analysis at large scale facilities (e.g. synchrotron and neutron facilities), research laboratory equipment, or distributed sensing systems in academic and industrial environments.

Major European Projects: EMCC-CSA, EoCoE, EXT MOS, GEMMA, GRAPHENE FLAGSHIP, GRIDABLE, INTERSECT, IQubits, MaX, NEXTOWER, SimDOME, NANODOME.

Active members: CINECA, CNR, ENEA, UNIBO, UNIMORE.



4.6. INDUSTRY 4.0

Manufacturing industrial sectors are the most relevant and impactful area in Italian national research, innovation, and company activities, in terms of employees, production, and business. The related applied research activity is carried out in a set of coordinated research & innovation organisations (university labs as well as interdepartmental, technopoles, the 8 Competence Centres established in the framework of the Industry 4.0 programme, etc.) and industries working in manufacturing, mechanics, industrial processing, and in many related areas such as mechatronics, automotive, robotics and automation, packaging, ceramics, electronic equipment, textile and garments, machinery and metal production, etc.

The changes in production processes and their management increasingly require the exploitation of knowledge extracted from data, acquired (offline and online) in virtual prototyping, testing, production, use, and maintenance. Industry 4.0 leverages the new paradigms of the IoT and cyber physical systems, exploiting the enormous amount of data coming from simulations, digital twins, and real machines to realise superior performance and energy-efficient manufacturing.

The manufacturing industry is currently in the middle of a data-driven revolution, from traditional manufacturing facilities to highly optimised smart manufacturing facilities to create manufacturing intelligence from machine-to-machine dataflow and to support accurate and timely decision-making. For example, new techniques, instruments, services for data analysis, learning, prediction and statistics enable novel diagnostics and prognostics for predictive maintenance by processing huge amounts of data describing the time-history of working conditions, integrated with open data about the execution environment and

synthetic data from digital-twin simulations. Manufacturing facilities must be capable of meeting the requirements of the exponential increase in data production, as well as possess the analytical techniques needed to extract valuable knowledge from big data. Techniques, services, applications, platforms, and infrastructure needs are often shared with other research and production fields, in terms of cloud/edge computing and HPC, big data analytics and visualisation, machine learning, and online stream processing.

Major European Projects: AEOLIX, CARIM, CLASS, CloudiFacturing, ColRobot, DREAM, ETEKINA, EXCELLERAT, FIRST, Fortissimo, FORTISSIMO2, I-MECH, IMPROVE, INCLUSIVE, IO-Twins, LINCOLN, Plan4Res, ROSSINI, SARAS, SYMBIOPTIMA, SYMPLEXITY.

Active members: ART-ER, CINECA, CNR, UNIBO, UNIFE, UNIMORE.



4.7. EARTH SCIENCE AND CLIMATE CHANGE

Big data in climate change is related to services for: (i) developing, porting and running Earth-system models at high spatial resolution; (ii) developing metrics and diagnostic tools for evaluating the models and for specific end-users' needs; (iii) acquisition, storage and processing of data on the order of petabytes from model simulations and observations.

High performance computing, data repository, data sharing and staging services are crucial to allowing a wide user base to produce and have access to a set of climate variables at high temporal resolutions and at extremely high spatial resolutions, and to perform probability predictions from sub-seasonal to multi-annual time scales. This is necessary in order to provide reliable climate information and maintain competitiveness at the international level with the major international meteo-climate services.

Users include both climate scientists and researchers from a wide range of fields, but climate data is of crucial interest for a large number of public/private actors in the energy, agriculture, water, health, tourism, urban management and transport sectors. Overall, the output of high-resolution climate simulations is a valuable asset at both the national and regional levels.

The term 'Earth data' refers both to atmosphere, ocean and sea state simulations (uncoupled or coupled in earth system models), environmental, geophysical and oceanographic observing systems, seafloor mapping and forecast services for institutional agencies (Civil Protection, Coast Guard and Ministry of the Environment) and to large databases of energy uses in industry, emission factors, basic data for life cycle assessment (LCA) and carbon footprint calculation and to databases related to the uses of the sea under a blue growth strategy and the BLUEMED Initiative.

Atmospheric models have increased their spatial and temporal resolution and produced a large

amount of output data both in diagnostic and in forecast mode. Archives of high resolution (about 1 km) atmospheric surface parameters (wind, temperature, solar radiation, etc.) predicted, on a daily basis, up to 48 hours ahead, are available for many environmental and renewable energy applications. Computing services for numerical weather predictions are provided daily and operational ensemble forecasts on the monthly timescale are performed on a weekly basis. Chemical transport models produce both 5-day air pollution forecasts for Italy and Europe and annual diagnostic simulations for the country. Hourly data for air pollutant concentrations and meteorological fields are calculated over the national domain with horizontal spatial resolution ranging from 4 to 1 km for 12 vertical levels from ground level to an altitude of 12 km.

Environmental change is particularly rapid in the coastal zone and offshore under the pressure of climate change (e.g. relative sea level rise) having impacts on drainage basins, coastal plains (e.g. man-induced subsidence), and offshore (e.g. hydrocarbon exploitation, fish trawling, energy plants, maritime traffic). Ocean observing systems combined with HPC provide scenarios of storm surges and associated coastal flooding, sea state, coastal erosion and long term ecological trends. Quantitative 4D geophysical pictures of the seafloor and sub-seafloor after major events (blooms, river floods, coastal erosion or anthropic interventions) help determine habitat modifications, coastal dynamics and subsidence trends.

Major European Projects: AtlantOS, BE-OI, ChEESE, CLARA, CRESCENDO, ENVRI PLUS, HIGHLANDER, IMMERSE, Iscape, LISTEN MED-GOLD, MEDSCOPE, MISTRAL, MOSES, ODYSSEA, OPERANDUM, PRIMAVERA, ROCK, RURITAGE, SeaDataCloud, SECLI-FIRM, SOCLIMPACT, SYSTEM-RISK.

Active members: ART-ER, CINECA, CMCC, CNR, ENEA, INGV, UNIBO, UNIMORE.



4.8. GLOBAL SOCIAL SCIENCES AND ECONOMICS

Big data in the social sciences and in economics are increasingly important for research, business, finance, policy-making and the society at large, creating social innovation, unique opportunities for job creation and worldwide leadership in R&I and the market.

These data (e.g. videos, photos and micro-messages shared on social networks, GIS, networks of relations, as well as behavioural and psychological data) promote interdisciplinary activities with regard to economic aspects, legal and ethical issues, and privacy-by-design principles. Behavioural data refers to multidimensional data-sets on human and social behaviour, actions, and interactions that allow understanding individual as well as community preferences, attitudes and habits, while psychological data-sets collect the results of psychological tests, surveys, attitude exams, and similar sources. Behavioural data analysis contributes to designing focused and personalised services in different fields (e.g. transport, healthcare, the environment); social network and GIS data analysis is focused on understanding inner aspects of human relations, and it can be applied to many fields (e.g. commercial advertising, study of the propensity to buy, sentiment analysis); psychological data analysis, on the other hand, helps researchers in proving or disproving theories, and/or supporting conclusions with statistical evidence. The collection and analysis of large amounts of behavioural data has also important consequences for the formulation and application of legal rules.

Moreover, the increasing use of large-scale administrative data-sets, private sector data and social media data can contribute to impactful social science research that transcends traditional disciplinary

boundaries, including the fields of economics, actuarial and financial sciences, mathematics, statistics, programming and computer science. The emergence of a multidisciplinary big data approach can provide a solid evidence-base for policy-making, as well as a quantitative knowledge base for solving problems in finance, insurance and risk-management. Legal aspects can also play a big role in the success of big data in finance and economy, especially as regards competition law, intellectual property rights, privacy law and consumer law.

In a digital global marketplace, leveraging big data is an unparalleled opportunity for companies including retail, finance, banking, advertising and insurance to gain a competitive edge. The real-time and fast processing of large volumes and varieties of data has a strong potential impact on key areas such as: customer data analytics; innovative marketing strategies; market trends and expectations; access to credit and finance for households and enterprises; e-commerce and e-paying systems for an easier access of consumers and citizens to services; risk management; business insights, organisational intelligence and operational efficiency.

Major European Projects: CoSIE, MICADO, MIREL.

Active members: UNIBO.



4.9. HUMANITIES, CULTURAL HERITAGE

Big data in the humanities and the field of cultural heritage have a tremendous potential for the transmission of the European cultural heritage, history, culture and identity, and for enhancing creativity and boosting economic development.

These data not only deal with large and networked cultural data-sets and with massive digital objects including large-scale digital corpora, but also call for new methods of study and interpretation in the social sciences and humanities (SSH), including the field of Digital Humanities. At least three sides can be recognised in this emerging world: digital objects, interpretations and interfaces.

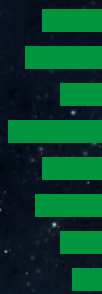
The development of new interpretive theories relates to understanding the technical complexity of the data processing pipelines: digitisation, transcription, pattern recognition, text, image and video analytics, simulation and inferences, preservation, and curation. The interpretation of SSH big data in the era of Digital Culture implies dealing with large-scale digital communities, collective discourses, global players, and evolving software. Physical and virtual interfaces, such as websites and virtual reality devices, make big cultural data accessible to scholars and to the general public. Interpretation is a critical step especially in the context of social networks, where trustable and fake messages are mixed together; improving the detection of malicious, fake or violent messages will improve the way consumers and users approach online communication. This requires prompt and coherent answers by the legislator, in order to find an adequate balance between the circulation and protection of personal and non-personal data.

Interfaces can be immersive, or linguistic, or provide synthetic data interpretation and

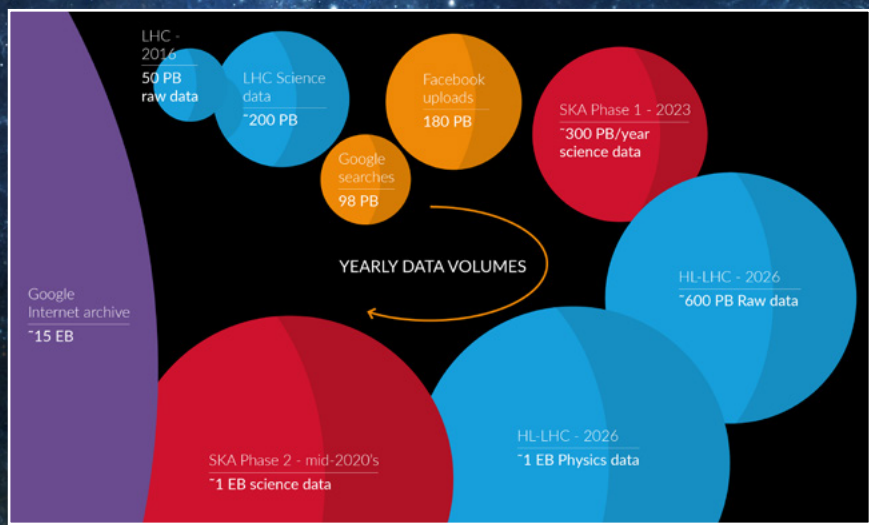
representation, or they can have a hybrid nature (e.g. augmented reality), challenging both the scientific and the industrial worlds. Nationwide-leading visualisation equipment is available in Italy, such as the CINECA Virtual Environment.

Major European Projects: I-Media-Cities, INSCRIBE, IPERION CH.

Active members: CINECA, CNR, UNIBO, UNIMORE.



FUTURE NEEDS IN DIFFERENT PHYSICS FIELDS



4.10. PHYSICS, ASTROPHYSICS AND SPACE SCIENCE

Many the physics, astrophysics and space science research areas that see our community as a global player in the big data domain. The key areas can be summarised as follows:

- The study of the large-scale structure of the universe, cosmology and the nature of black holes and the hot and energetic universe, as well as the theoretical formulation of the formation of galaxies and their evolution
- Mapping the geometry of the dark universe
- Observation and study of the Universe with next generation world's largest telescopes in the optical-infrared, radio, and high energy electromagnetic spectral bands
- Simulation of gravitational lensing data and simulation of signals from (and in connection with the search for) gravitational waves
- Quantum Chromo-Dynamics simulations: particle production and properties
- Studies of the properties of the Higgs boson with the LHC experiments at CERN Geneva
- Search for New Physics to study the nature and origin of dark matter and dark energy in the Universe and to investigate the possibility of the existence of extra dimensions
- Study of the properties of the neutrino and dark matter in underground laboratories, in space or underwater conditions
- Large-scale non-perturbative simulations of QCD in the lattice regularisation
- Multi-wavelength observations and study of gravitational waves through their electromagnetic counterparts

Accelerator particle physics, in particular LHC experiments, next generation ground-based telescopes (e.g. CTA, the world's largest and most sensitive high-energy observatory, or

SKA, the world's largest radiotelescope, or the ELT, the world's largest optical-infrared telescope), and astrophysics and space science missions (e.g. the ESA GAIA space mission for a multidimensional mapping of one billion stars in the Galaxy, the ESA Euclid space mission to map the geometry of the dark universe) have already entered the so called big data regime. Some examples: by 2030, the CTA is expected to produce more than 100 PB of data, while SKA 1 total raw data is expected to exceed by a factor of 5 the global internet traffic generated in the year 2015; Euclid (launch in 2022) will collect several millions of images and about 30 Petabytes of data that will be combined with other gigantic data-sets obtained with ground-based telescopes; every year the LHC experiments collect a few Petabytes (PB) of data that are copied to the national computing infra structures, in Italy to the CNAF Tier 1. An almost equal amount of simulated data is needed to finalise these studies. These topics naturally imply the need for big data, in terms of:

- Data archive handling, accessibility, and interoperability;
- High throughput and high performance computing, high speed network;
- Image processing and modelling tools.

In the future the computing resources needed by LHC science, ESA space missions (Euclid, ATHEnA), and large ground-based telescopes (ELT, CTA, SKA) will grow significantly: in 2023 LHC will require an increase by a factor of about 60 in CPU and 40 in disk space. As it is shown in the figure alongside for SKA alone the data volume will be comparable with the present day world-wide global Internet traffic.

Major European Projects: AMICI, ARIES, ENSAR2, ESCAPE, EuroPLEx, FELLINI, INSPYRE, INTENSE, KM3Net 2.0, MAGCOW, RADIATE, SINE2020, TEQ, XLS.

Active members: ENEA, INAF, INFN, INGv, UNIBO, UNIPR.

5

EUROPEAN PROJECTS LIBRARY

Projects funded by H2020, FP7 or CEF programs that are still active or started since 2015 and that involve at least one member of the association.

<u>ADOPT-BBMRI-ERIC</u>	ImplementAtion and OPeration of the gateway for healTh into BBMRI-ERIC
<u>AEOLIX</u>	Architecture for EurOpean Logistics Information eXchange
<u>AI4EU</u>	A European AI On Demand Platform and Ecosystem
<u>AIDA</u>	Artificial Intelligence Data Analysis
<u>AIDA-2020</u>	Advanced European Infrastructures for Detectors at Accelerators
<u>AirPROM</u>	Airway Disease PRedicting Outcomes through Patient Specific Computational Modelling
<u>AMICI</u>	Accelerator and Magnet Infrastructure for Cooperation and Innovation
<u>ANTAREX</u>	AutoTuning and Adaptivity appRoach for Energy efficient eXascale HPC systems
<u>ARIES</u>	Accelerator Research and Innovation for European Science and Society
<u>AtlantOS</u>	Optimizing and Enhancing the Integrated Atlantic Ocean Observing System
<u>BD2Decide</u>	Big Data and models for personalized Head and Neck Cancer Decision support
<u>BE-OI</u>	Beyond EPICA - Oldest Ice
<u>CARIM</u>	Commercialization of a full carbon wheel manufactured with an automated high-volume process for the automotive market
<u>ChEESE</u>	Centre of Excellence for Exascale in Solid Earth
<u>ChiLTERN</u>	Children's Liver Tumour European Research Network
<u>CIRCLES</u>	Controlling micRobiomes CircuLations for bEtter food Systems
<u>CLARA</u>	Climate forecast enabled knowledge services
<u>CLASS</u>	Edge and CLoud Computation: A Highly Distributed Software Architecture for Big Data AnalyticS
<u>CloudiFacturing</u>	Cloudification of Production Engineering for Predictive Digital Manufacturing
<u>ColRobot</u>	Collaborative Robotics for Assembly and Kitting in Smart Manufacturing
<u>COMPARE</u>	COLlaborative Management Platform for detection and Analyses of (Re-) emerging and foodborne outbreaks in Europe
<u>CORBEL</u>	Coordinated Research Infrastructures Building Enduring Life-science services
<u>CoSIE</u>	Co-creation of service innovation in Europe
<u>CRESCENDO</u>	Coordinated Research in Earth Systems and Climate: Experiments, kNowledge, Dissemination and Outreach
<u>CYBELE</u>	Fostering precision agriculture and livestock farming through secure access to large-scale-HPC-enabled virtual industrial environment empowering scalable BIG DATA analytics
<u>DeepHealth</u>	Deep-Learning and HPC to Boost Biomedical Applications for Health
<u>Domino</u>	Novel tools to evaluate ATM systems coupling under future deployment scenarios

DREAM	Design for Resource and Energy efficiency in cerAMic kilns
ECDP	European Cohort Development Project
EDEN ISS	Ground Demonstration of Plant Cultivation Technologies and Operation in Space for Safe Food Production on-board ISS and Future Human Space Exploration Vehicles and Planetary Outposts
EJPRD	European Joint Programme on Rare Diseases
EMMC-CSA	European Materials Modelling Council
EnABLES	European Infrastructure Powering the Internet of Things
ENABLE-S3	European Initiative to Enable Validation for Highly Automated Safe and Secure Systems
ENSAR2	European Nuclear Science and Application Research 2
ENVRI PLUS	Environmental Research Infrastructures Providing Shared Solutions for Science and Society
EoCoE	Energy oriented Centre of Excellence for computer applications
EOSC-hub	Integrating and managing services for the European Open Science Cloud
EPEEC	European joint Effort toward a Highly Productive Programming Environment for Heterogeneous Exascale Computing
ESCAPE	European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures
ETEKINA	Heat pipe technology for thermal energy recovery in industrial applications
EUDAT2020	EUDAT 2020: Collaborative Data Infrastructure
EUROfusion	Implementation of activities described in the Roadmap to Fusion during Horizon 2020 through a Joint programme of the members of the EUROfusion consortium
EuroPLEx	European network for Particle physics, Lattice field theory and Extreme computing
EU-STANDS4PM	A European standardization framework for data integration and data-driven in silico models for personalized medicine
ExaNeSt	European Exascale System Interconnect and Storage
EXCELLERAT	The European Centre of Excellence for Engineering Applications
EXDCI	European eXtreme Data and Computing Initiative
EXTMOS	EXTended Model of Organic Semiconductors
FELLINI	FELLowship for Innovation at INFN
FIRST	virtual Factories: Interoperation supporting buSiness innovaTION
FLEXMETER	Flexible smart metering for multiple energy vectors with active prosumers
FOCAS	FOCAS Organisation, Coordination And Support
Fortissimo	Factories of the Future Resources, Technology, Infrastructure and Services for Simulation and Modelling
Fortissimo 2	Factories of the Future Resources, Technology, Infrastructure and Services for Simulation and Modelling 2
GEMMA	GEneration iv Materials MAaturity
GRAPHENE FLAGSHIP	Graphene Flagship
GRIDABLE	Plastic nanocomposite insulation material enabling reliable integration of renewables and DC storage technologies in the AC energy grid

<u>HBP SGA1</u>	Human Brain Project Specific Grant Agreement 1
<u>HERCULES</u>	Towards geoHazards rEsilient infRastruCtUre under changing cLimatES
HIGHLANDER	HIGH performance computing to support smart LAND sERVICES
<u>HPC-EUROPA3</u>	Transnational Access Programme for a Pan-European Network of HPC Research Infrastructures and Laboratories for scientific computing
<u>ICARUS</u>	Innovative Coarsening-resistant Alloys with enhanced Radiation tolerance and Ultra-fine -grainedStructure for aerospace application
<u>ICEI</u>	Interactive Computing E-Infrastructure for the Human Brain Project
<u>I-MECH</u>	Intelligent Motion Control Platform for Smart Mechatronic Systems
<u>I-Media-Cities</u>	Innovative e-environment for Research on Cities and the Media
<u>IMforFUTURE</u>	Innovative training in methods for future data
<u>IMMERSE</u>	Improving Models for Marine EnviRonment SERVICES
<u>IMPROVE</u>	Innovative Modeling Approaches for Production Systems to raise validatable efficiency
<u>INCLUSIVE</u>	Smart and adaptive interfaces for INCLUSIVE work environment
<u>INMARE</u>	Industrial Applications of Marine Enzymes: Innovative screening and expression platforms to discover and use the functional protein diversity from the sea
<u>INSCRIBE</u>	INvention of SCRipts and their BEginnings
<u>INSPYRE</u>	Investigations Supporting MOX Fuel Licensing in ESNII Prototype Reactors
<u>INTENSE</u>	Particle physics experiments at the high intensity frontier, from new physics to spin-offs. A cooperative Europe - United States - Japan effort.
<u>INTERSECT</u>	Interoperable Material-to-Device simulation box for disruptive electronics
IO-Twins	Distributed Digital Twins for industrial SMEs: a big-data platform
<u>IPERION CH</u>	Integrated Platform for the European Research Infrastructure ON Cultural Heritage
<u>IQubits</u>	Integrated Qubits Towards Future High-Temperature Silicon Quantum Computing Hardware Technologies
<u>ISCAPE</u>	Improving the Smart Control of Air Pollution in Europe
<u>KM3NeT 2.0</u>	Astroparticle and Oscillations Research with Cosmics in the Abyss (ARCA and ORCA)
<u>LANGELIN</u>	Meeting Darwin's last challenge: toward a global tree of human languages and genes
<u>LEGVALUE</u>	Egstering sustainable legume-based farming systems and agri-feed and food chains in the
<u>LINCOLN</u>	Lean innovative connected vessels
<u>LISTEN</u>	Lost In translation: Strengthening communication skills between real world and climaTe modELs for seasonal to decadal prediction
<u>MAGCOW</u>	The Magnetised Cosmic Web
<u>MARISA</u>	Maritime Integrated Surveillance Awareness
<u>MaX</u>	MAterials design at the eXascale. European Centre of Excellence in materials modelling, simulations, and design
<u>MCDS-Therapy</u>	Repurposing of carbamazepine for treatment of skeletal dysplasia
<u>MedAID</u>	Mediterranean Aquaculture Integrated Development

MED-GOLD	Turning climate-related information into added value for traditional MEDiterranean Grape, OLive and Durum wheat food systems
MEDSCOPE	MEDiterranean Services Chain based On Climate PrEdictions
MICADO	Migrant Integration Cockpits and Dashboards
MIREL	Mining and REasoning with Legal texts
MISTRAL	Meteo Italian Supercomputing Portal
MOSES	Managing crOp water Saving with Enterprise Services
MYNEWGUT	Microbiome Influence on Energy balance and Brain Development-Function Put into Action to Tackle Diet-related Diseases and Behavior
NanoDome	Nanomaterials via Gas-Phase Synthesis: A Design-Oriented Modelling and Engineering Approach
Neuromics	Integrated European -omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases
NEXTOWER	Advanced materials solutions for next generation high efficiency concentrated solar power (CSP) tower systems
ODYSSEA	OPERATING A NETWORK OF INTEGRATED OBSERVATORY SYSTEMS IN THE MEDITERRANEAN SEA
OPERANDUM	OPEn-air laboRAatories for Nature baseD solUtions to Manage environmental risks
OPRECOMP	Open transPREcision COMPUting
ORTHOUNION	ORTHOpedic randomized clinical trial with expanded bone marrow MSC and bioceramics versus autograft in long bone nonUNIONS
OSMOSE	Optimal System-Mix Of flexibility Solutions for European electricity
PANINI	Physical Activity and Nutrition Influences In ageing
PAPA-ARTIS	Paraplegia Prevention in Aortic Aneurysm Repair by Thoracoabdominal Staging with 'Minimally-Invasive Segmental Artery Coil-Embolization': A Randomized Controlled Multicentre Trial
PerformFISH	Consumer driven Production: Integrating Innovative Approaches for Competitive and Sustainable Performance across the Mediterranean Aquaculture Value Chain
Plan4Res	Synergistic approach of multi-energy models for an european optimal energy system management tool
PLATYPUS	PLAsticiTY of Perceptual space Under Sensorimotor interactions
PRACE-4IP	PRACE 4th Implementation Phase Project
PRIMAVERA	PRocess-based climate sIMulation: AdVances in high resolution modelling and European climate Risk Assessment
Prolific	Integrated cascades of PRocesses for the extraction and valorisation of proteins and bioactive molecules from Legumes, Fungi and Coffee agro-industrial side streams
PROPAG-AGEING	The continuum between healthy ageing and idiopathic Parkinson Disease within a propagation perspective of inflammation and damage: the search for new diagnostic, prognostic and therapeutic targets
PRYSTINE	Programmable Systems for Intelligence in Automobiles
RADIATE	Research And Development with Ion Beams – Advancing Technology in Europe
RES_URBIS	REsources from URban Blo-waSte
Re-Search Alps	Research Laboratories in the Alpine Area

<u>ROCK</u>	Regeneration and Optimisation of Cultural heritage in creative and Knowledge cities
<u>ROSSINI</u>	RObot enhanced SenSing, INtelligence and actuation to Improve job quality in manufacturing
<u>RURITAGE</u>	Rural regeneration through systemic heritage-led strategies
<u>SARAS</u>	Smart Autonomous Robotic Assistant Surgeon
<u>SeaDataCloud</u>	SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management
<u>SEADATANET-II</u>	SeaDataNet II: Pan-European infrastructure for ocean and marine data management
<u>SECLI-FIRM</u>	The Added Value of Seasonal Climate Forecasts for Integrated Risk Management Decisions
<u>SECREDAS</u>	Cyber Security for Cross Domain Reliable Dependable Automated Systems
<u>SimDOME</u>	Digital Ontology-based Modelling Environment for Simulation of materials
<u>SINE2020</u>	World class Science and Innovation with Neutrons in Europe 2020 – SINE2020
<u>SMARTCHAIN</u>	Towards Innovation - driven and smart solutions in short food supply chains
<u>SocialCar</u>	Open social transport network for urban approach to carpooling
<u>SOCLIMPACT</u>	DownScaling CLimate IMPACTs and decarbonisation pathways in EU islands, and enhancing socioeconomic and non-market evaluation of Climate Change for Europe, for 2050 and beyond.
<u>SOGNO</u>	Service Oriented Grid for the Network of the Future
<u>SYMBIOPTIMA</u>	Human-mimetic approach to the integrated monitoring, management and optimization of a symbiotic cluster of smart production units
<u>SYMPLEXITY</u>	Symbiotic Human-Robot Solutions for Complex Surface Finishing Operations
<u>SYSTEM-RISK</u>	A Large-Scale Systems Approach to Flood Risk Assessment and Management
<u>TEQ</u>	Testing the Large-Scale Limit of Quantum Mechanics
<u>THALAMOSS</u>	THALAssaemia MOdular Stratification System for personalized therapy of beta-thalassemia
<u>TRAFAIR</u>	Understanding traffic flows to improve air quality
<u>TREASURE</u>	Diversity of local pig breeds and production systems for high quality traditional products and sustainable pork chains
<u>VALUMICS</u>	Understanding food value chains and network dynamics
<u>VPH-Share</u>	Virtual Physiological Human: Sharing for Healthcare - A Research Environment
<u>XDC</u>	Extreme-Datacloud
<u>XLS</u>	CompactLight



6

CONTACTS

Fabio Fava

fabio.fava@unibo.it

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA**Alma Mater Studiorum – Università di Bologna**

The Alma Mater Studiorum – Università di Bologna, founded in 1088, is the oldest university in the Western world. Organised in 5 campuses, with 32 Departments and 12 R&I Centers, Alma Mater has a diversified course catalogue with more than 200 degree programmes, offered to over 82,000 students, 3,600 graduates are enrolled in PhDs and other 3rd-cycle programmes. University of Bologna paves the way for innovation through cutting-edge research and strong alliances with industry and public/private organizations. Beyond its close European links, it enjoys multiple connections with America, Africa, Asia and Australia.

Leda Bologni

leda.bologni@art-er.it

**ART-ER**

ART-ER Attractiveness Research Territory is the Emilia-Romagna Joint Stock Consortium Company that was born in May 2019 from the merger of ASTER and ERVET, with the purpose of fostering the region's sustainable growth by developing innovation and knowledge, attractiveness and internationalization of the territorial system.

Sanzio Bassini

s.bassini@cineca.it

**CINECA**

Established in 1969, CINECA is a non-profit consortium of 70 Italian Universities, 6 National Research Institutes and the Ministry of Education, University and Research (MIUR). CINECA is the largest Italian supercomputing centre, represents Italy in PRACE and is founding members of the European Technology Platform for HPC (ETP4HPC) as well as full member of the BDVA (Big Data Value Association), where is also recognized as i-Space. CINECA SCAL is the Italian representative in the pan-European EUDAT Collaborative Data e-Infrastructure and core partner in the EU flagship project Human Brain.

Antonio Navarra

antonio.navarra@cmcc.it

**CMCC**

CMCC Foundation (Centro Euro-Mediterraneo sui Cambiamenti Climatici) conducts and promotes scientific and applied activities within the scope of international climate change research. CMCC aims to gain in-depth knowledge on climate variability, its causes, its consequences, and its interactions with the socio-economic systems.

Elisa Molinari

molinari@unimore.it

**CNR**

The National Research Council (CNR) is the largest scientific public organisation in Italy: about 5000 researchers and 4,000 support staff members, organised in 7 scientific Departments and 102 Institutes across the country, in sectors ranging from biotechnology and medicine to materials, advanced systems of production, information and communication, climate and environment, social sciences and arts. It is beneficiary of over 460 European-funded H2020 projects and coordinates 114 of those, and leads a large number of relevant national and regional projects.

Simonetta Pagnutti

simonetta.pagnutti@enea.it

**ENEA**

ENEA (National Agency for New Technologies, Energy and Sustainable Economic Development) is a public body aimed at research and technological innovation.

Its focus sectors are energy technologies, nuclear fusion and safety, raw materials, cultural heritage, seismic protection, pollution, climate change with a strong tradition in computer modeling

Giuseppe Malaguti

malaguti@inaf.it

**INAF**

The Universe "speaks to us" at all wavelengths from ground-based observations with new-generation telescopes and radio telescopes, to satellite observations, to the "close" exploration of the solar system. In this context, the National Institute of Astrophysics (INAF), has over 1500 employees and contract staff and it is one of the flagships of the country, both for the high level of the scientific production of its researchers, and for its ability to design, implement and manage large national and international infrastructures.

Antonio Zoccoli

antonio.zoccoli@bo.infn.it

**INFN**

The National Institute for Nuclear Physics (INFN) is the Italian research agency dedicated to the study of the fundamental constituents of matter and the laws that govern them. Through the contributions of about 5,000 scientists, it conducts theoretical and experimental research in the fields of subnuclear, nuclear and astroparticle physics.

Francesca Quareni

francesca.quareni@ingv.it

**INGV**

The National Institute of Geophysics and Volcanology (INGV) is a research institute established in 1999 for geophysics and volcanology in Italy. Its main mission is the study and investigation of geophysical processes, as well as the monitoring of the phenomena which can be observed in our planet. Within the Italian Civil Protection system INGV is in charge of the maintenance and monitoring of the national networks for seismic and volcanic phenomena. The institute employs around 1000 people distributed between its sections located in Rome, Milan, Bologna, Pisa, Naples, Catania, and Palermo and several observatories all over the country.

Mariapaola Landini

mariapaola.landini@ior.it

**Istituti Ortopedici Rizzoli**

The Rizzoli Orthopedic Institute is one of the 51 Italian "Scientific Research Hospitals" (IRCCS) and it is dedicated to healthcare and research in the field of musculoskeletal diseases. The institute's strength lies in the close integration between healthcare and scientific research, which is carried out in 21 clinics (350 beds) and 15 laboratories, employing a staff of >1000 people including doctors, nurses, researchers, technicians and administrative officers.

Beatrice Nepoti

beatrice.nepoti@lepidaitalia.it

**LepidaScpA**

LepidaScpA is a subsidiary of the Emilia-Romagna Region and is the main operational instrument as regards implementation of the Regional ICT Plan that defines the strategies of broadband networks, ensures and optimizes the delivery of ICT services, develops cloud infrastructure, implements and manages innovative solutions for the modernization of healthcare paths to improve the relationship between citizens and the Regional Health Service in accordance with the provisions of the European, National and Regional Digital Agendas.

Fabrizio Riguzzi

fabrizio.riguzzi@unife.it

**Università di Ferrara**

The University of Ferrara, founded in 1391, is one of the oldest universities in Italy. Its identity is rooted in the historical and cultural tradition that welcomed and formed illustrious figures such as Copernico and Paracelso. In terms of its size, facilities, quality and quantity of education and research, the university is a point of excellence within Italy, and it is split into 12 departments, 1 Faculty of medicine, Pharmacy and Prevention and 31 centres, offering 62 degree courses and numerous research and post-graduate courses.

Rita Cucchiara

rita.cucchiara@unimore.it

**Università di Modena**

UNIMORE has a longstanding tradition (it was founded in 1175) and is considered one of the best universities in Italy for teaching and research. It is ranked 2nd among public universities according to Italy's leading financial daily, and among the top 8 medium-sized Italian universities by the Times Higher Education Ranking 2011-2012.

Stefano Cagnoni

stefano.cagnoni@unipr.it

**Università di Parma**

The University of Parma is a state university, holding 9 departments, 39 first-cycle, 6 single-cycle, 41 second-cycle degree courses, as well as many postgraduate schools, teacher-training courses, Master programmes and PhD courses. It has 25,000 students and about 1,700 faculty and staff members.

associazione
BigData
from volume to value

 Regione Emilia-Romagna