EMILIA ROMAGNA BIG DATA COMMUNITY

FROM VOLUME TO VALUE

RegioneEmilia-Romagna

January 25, 2016

FONDI STRUTTURALI E DI INVESTIMENTO EUROPEI

I fondi strutturali e di investimento europei (Fondi SIE) sono il principale strumento di intervento della politica di investimenti dell'Unione europea per promuovere lo sviluppo e la coesione all'interno dei paesi dell'Unione, che gestiscono i fondi in modo decentralizzato.

I fondi SIE sono quattro: il Fondo europeo di sviluppo regionale (FESR), il Fondo sociale europeo (FSE), il Fondo europeo agricolo per lo sviluppo rurale (FEASR - PSR), il Fondo europeo per gli affari marittimi e la pesca (FEAMP).

EMILIA ROMAGNA BIG DATA COMMUNITY

FROM VOLUME TO VALUE

POR FESR EMILIA-ROMAGNA 2014/2020

ER educazione ricerca Emilia-Romagna

Programma di Sviluppo Rurale dell'Emilia-Romagna 2014-2020

SUMMARY

1
THE EMILIA ROMAGNA BIG DATA

1.1 THE EMILIA ROMAGNA BIG DATA COMMUNITY IN FAVOR OF REGIONAL GROWTH, COMPETITIVENESS AND PROSPERITY

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

The Emilia Romagna region is investing remarkably on Research and technology transfer in all major domains of innovation proving local sustainable growth and jobs.
The traditional attention to innovation comes mainly from a widespread interdisciplinary and dynamic research system (4 public universities, prominent research infrastructures and facilities and the territorial branches of national research centres) well integrated into the local industrial landscape mainly composed by SMEs.
During the last ten years great attention has been paid to:

- build on, sharing and exploiting existing results, knowledge, capacities, and research and innovation initiatives and frameworks;
- fostering cooperation between public and private sectors to maximize the leverage effects of research investments both commercially and with respect to public policy at regional, national and EU levels
- promoting joint actions including coordination, planning and programming of relevant research and innovation activities;
- supporting researcher careers, training and mobility, and development of skills in relevant sectors to ensure the necessary highly qualified workforce needed to underpin a prosperous and sustainable growth.

The funds invested in R&D within the SF programming period 2007-2013, both from research supply and demand results perspective, created an effective local innovation ecosystem which includes all the regional stakeholders and strengthens their collaboration along a common path by significantly contributing to the economic and social growth.
To concretely carry out such activities, it seems more and more important to integrate and match different research groups in order to enlarge the critical mass and the capacity to answer businesses requirements and technical needs.
Supercomputing and big data are examples of convergent skills, facilities and technologies offering new opportunities of economic growth and scientific progress.

In fact, the added value coming from managing large amounts of data is a skill whose importance will grow exponentially in the near future. It requires a great computing capacity in terms of performance and memory available. Possible applications range go from several research areas (eg. particle physics, space exploration, etc..) to applicative domains (eg. financial analysis, health, environmental monitoring and geophysical simulations, cultural heritage management, precision farming, multimedia and analysis of images and video, etc.).
Impacts of big data on regional Smart Specialization Strategy might be remarkable if properly applied in areas of specialization going from knowledge-intensive regional systems (health and well-being field and cultural and creative industries), to the consolidated production systems. Moreover, the issue is closely connected to the services sector innovations.
Many world-class institutions in supercomputing and big data are located in the regional territory. These players are able to manage and analyze large amounts of data. In addition, the territory counts on a quite effective big data infrastructures (ultra-wideband).
Therefore, the potential in exploiting the available expertises, public and private actors as well as the facilities of the sector to generate better local economic and societal opportunities is very relevant.
This document is a result of a cooperative work carried out by all regional stakeholders relevant for supercomputing and big data production and management and collects figures, expertises, technologies and facilities available in the Emilia-Romagna region in each of the knowledge and innovation domains of major relevance for the Region.

1.2 INTRODUCTION

High Performance Computing, big data and high-speed networks (a.k.a e-Infrastructures) are the technological foundations of modern society. In all developed countries they support economy, business, scientific and industrial research, education, healthcare and are increasingly strategic for homeland security.

The recently issued "A Digital Single Market Strategy for Europe Digital Market" Communication from the Commission to the EU Parliament identifies as one of the necessary priorities for Europe to "Maximize the growth potential of the digital economy" through "actions with far-reaching effects on European industrial competitiveness, investment in ICT infrastructures and technologies such as Cloud Computing and Big Data, research and innovation as well as inclusiveness and skills".

Italy needs to provide SMEs with an easy access to big data and the related analytic tools, high bandwidth networking and high performance and high throughput computing to increase their competitiveness and stimulate job creation and economic growth.
Central and regional administrations need a well-developed data and computing-based infrastructure to provide services to the citizen in the education, health, transport sectors and also e-Government at national and regional levels. The ability to analyze big data provides unique opportunities for Universities and Research Centers involved in advanced research projects. Instead of being limited to sampling large data sets, you can now use much more detailed and complete data to perform the required assessment or interpret the investigated phenomena.

In Italy, the research infrastructures INFN, CNR, GARR and CINECA have already implemented big data, high performance computing and high performance national network e-infrastructures to support major research and academic communities. These efforts are coordinated at EU level and, in the majority of case, connected to worldwide initiatives.
The majority of these Italian high performance and high throughput computing resources are concentrated in the Emilia-Romagna region where are also located prominent Universities, Research Centers and companies which are other essential players of the regional big data platform.

The expansion and integration of these e-Infrastructures, will make it possible to create a more powerful local open science platform to support not only the institutional scientific communities covered by these and other agencies, but also stimulate regional growth through the exploitation of scientific data and knowledge made widely openly available. The strong regional economic structure with many industries and SMEs may highly benefit from such an advanced and integrated e-infrastructure.

2
EMILIA ROMAGNA BIG DATA PROFILE

2.1 EMILIA ROMAGNA BIG DATA COMMUNITY IN FIGURES (2013-2015)

1791

RESEARCHERS INVOLVED

230

FOREIGN RESEARCHERS HOSTED

94

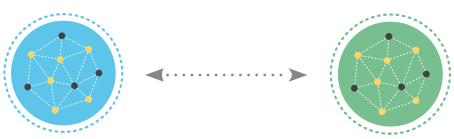
INTERNATIONAL EVENTS

60

HIGHER EDUCATION INITIATIVES INCLUDING

- PhD courses
- Laurea magistrale
- Master
- Summer schools

2.2 THE CONNECTIVITY



GARR-X

RESEARCH NETWORK

- Up to **100 gbps**
- Capacity **4 point of presences** (pop)

LEPIDA

THE EMILIA-ROMAGNA REGIONAL NETWORK OF THE PUBLIC ADMINISTRATIONS

- Towards **100 gbps** capacity
- More than **140,000 km optical fibers** and **2,500 access nodes**
- **42 Pop nodes + 4 integrated** regional data centers
- Current coverage of fast internet availability in emilia-romagna: **40% of households**

2.3 | THE HARDWARE COMPUTING INFRASTRUCTURES

An High Performance Computing facility (HPC) hosting a Tier0 and Tier1 and operating within PRACE (Partnership for Advanced Computing in Europe) at CINECA, in Bologna, with the following capabilities (*):

CPU: ~16 PETAFLOPS / 350.000 COMPUTING CORES

STORAGE: ~20 PB OF NET DISK SPACE

(*) available by end of 2016 with the installation of the MARCONI Supercomputer.

A High Throughput Computing facility (HTC) which hosts the WLCG Tier1 at the CNAF-INFN in Bologna, with the following capabilities:

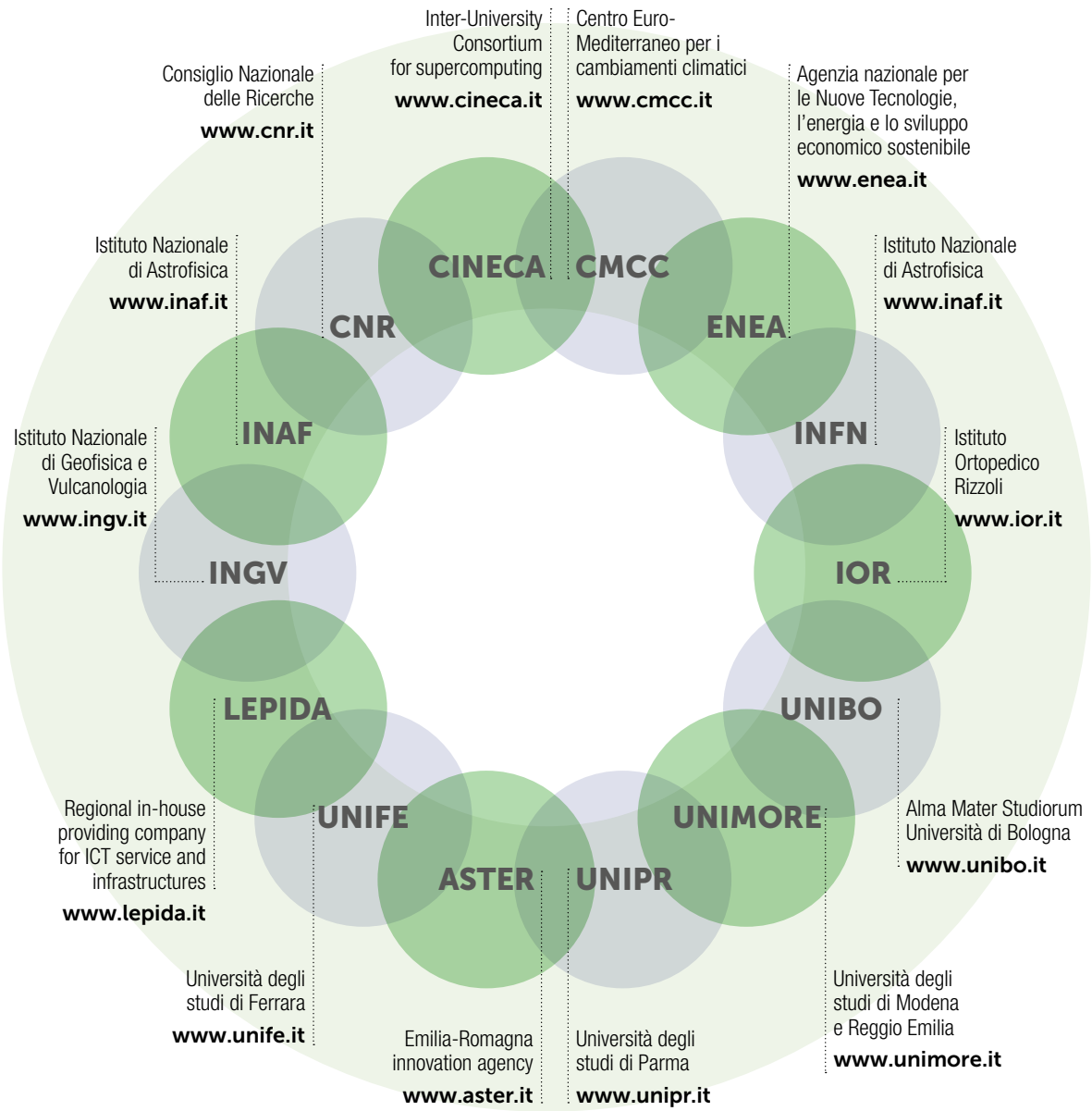
CPU: ~193 KHS06 / ~15600 COMPUTING CORES

STORAGE: ~17 PB OF NET DISK SPACE

LIBRARY: ~22PB OF TAPE SPACE.

GARR-X/LEPIDA
Fast and effective nation-wide network connection, mainly provided by GARR and Lepida

2.4 | THE STAKEHOLDERS



2.5 | HIGH PERFORMANCE COMPUTING FACILITY

The **HPC** facility hosted by CINECA is currently based on a Tier0 System (200,000 cores for a peak performance in excess of 2 Petaflops) and a Tier1 Big Data System (Linux cluster for high performance computing data analytics – HPDA).
From next June 2016 the Tier0 system will be replaced by a new Tier0 system, logical name MARCONI, with a configuration of more than **350,000 computing cores**. At the end of the deployment process the **MARCONI** system will provide a peak computing performance in excess of 16 Petaflops, which presumably will rank CINECA in the TOP 10 HPC supercomputing centers at worldwide level, being the procurement process already completed and fixed. The Tier1 Big Data system will also be renewed introducing a computing platform specializing in big data processing and analytics. The storage capacity will also be tailored sized according to the computing capability with the introduction of new storage servers for a total storage capacity in excess of **20 Petabyte** of on line storage and more of **25 Petabyte** of archiving and long time preservation and curation service for big data. The CINECA HPC facility enables a wide range of scientific research through open access granted by independent international peer reviewed processes.

Three main actions lead this access model:

PRACE

Worldwide scientists having an affiliation in a European entity;

ISCRA

European scientists having an affiliation in an Italian entity;

LISA

Italian and European scientists having an affiliation with an entity located in Lombardy Region.

From next June 2016 CINECA will provide service to the Worldwide Eurofusion community, being the contractor of the European tender for that specific service. CINECA also provide operational computing service for weather forecast for National Civil Protection under the supervision of Emilia Romagna ARPA-SMR. Moreover, CINECA is part of the European digital infrastructure for many ESFRI RI facilities and initiatives, among others, EPOS: European Plate Observing System, lead by INGV, ELIXIR – Infrastructure for Life Science, and HPB European Human Brain Flagship Project. At national level CINECA signed a partnership agreement with Telethon Foundation for the national repository of genomic data and with ISTAT (Italian Central Statistics Office) for web crawling and cognitive computing research and development. CINECA has entered formal partnerships for added value services and R&D activities with ENI and UNIPOL.

2.6 | HIGH THROUGHPUT COMPUTING FACILITY

The **HTC** facility is hosted in Bologna by CNAF, which is one of the INFN National Centers local structures defined in the INFN Statute. CNAF has been charged with the primary task of setting up and run the so-called **Tier-1 data center** for the Large Hadron Collider (LHC) experiments at CERN in Geneva. Now it hosts computing for many other experiments ranging from high energy physics to astroparticles. CNAF participated as a primary contributor in the development of Grid middleware and in the operation of the Italian Grid infrastructure.
This facility is operating in the framework of a national INFN HTC infrastructure consisting of the CNAF Tier1 and 10 smaller facilities, called Tier2, placed over the whole Italian territory.

CNAF HTC Tier1 Data Center operates about 1,000 computing servers providing ~15,000 computing cores allowing the concurrent run of 20,000 jobs. 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 power for about 95% of time.

CNAF operates a very large storage infrastructure based on industry standards, for connections (all disk servers and disk enclosures are interconnected through a dedicated Storage Area Network) and for data access (data is hosted on GPFS file systems, typically one per major experiment). This solution allows the implementation of a completely redundant data access system.

The main features of the storage system are:

~17 PB OF NET DISK SPACE IN 15 FILE SYSTEMS

~22PB OF TAPE SPACE.

AGGREGATE BANDWIDTH BETWEEN THE FARM AND THE STORAGE: 60 GBYTES/S.

SUPPORT OF STANDARD ACCESS PROTOCOLS AND INTERFACES, AS DEFINED BY THE WLCG/EGI PROJECTS (GRIDFTP, XROOTD, WEBDAV, SRM).

The users community of the CNAF Tier1 facility is composed mainly by research groups from INFN and most Italian Universities (including UNIBO, UNIFE, UNIPR), working on nuclear, subnuclear, astroparticles and theoretical physics.

In the next five years the HTC computing resources of INFN CNAF will be increased by about a factor 5 in order to match the computing and storage requirements of the foreseen experiments.

2.7 OTHER RELEVANT FACILITIES

UNIBO, UNIMORE, UNIPR and UNIFE have many hardware resources at their disposal in terms of multi-core and many-core clusters of workstations and servers, used by research groups at Departments and Research Centers for big data analysis, modeling and engineering. Some of them have been co-funded by Emilia Romagna Region for the Technopole's labs under the EU 2007-2013 ERDF program.

Computing facilities also include state-of-the-art platforms and software tools for data analysis.

In addition, Emilia-Romagna Regional Government has entrusted Lepida SpA with the design, implementation and provision of four data centers, geographically distributed, for the use of Public Administrations. The data centers (natively connected to the Lepida network) offer advanced computing services, storage, disaster recovery, backup, business continuity. These data centers will meet at least Tier 3 specifications, with native disaster recovery functions and a specific focus on energy management.

CONTACTS

SANZIO BASSINI
(CINECA)
s.bassini@cineca.it

ANTONIO ZOCCOLI
(UNIBO AND INFN)
antonio.zoccoli@unibo.it

KUSSAI SHAHIN
(LEPIDA)
kussai.shahin@lepida.it

3
BIG DATA IN RESEARCH AND INNOVATION DOMAINS

3.1 BIG DATA IN ICT AND DIGITAL CONTENT

As enormous volumes of data are being created every day, an enormous research effort is focused on Information and Communication Technologies for the theoretical and practical aspects of managing content and extracting knowledge from Big Data .

This domain is devoted to cover the entire data value chain and for any type of digital content, both digital-by-design content and digitalized content by dematerialization processes. They embrace business data and user-generated-content, content created in an ATAWAD (anytime, anyway, anydevices) seamless way, data coming from societal domains (administrative data, broadcast, web multimedia content, social networks, educational/cultural/social archives, financial and production data...) and open data.

The research has a direct impact on the ICT industry, for the large International and European players and, especially in Emilia Romagna, for innovative startups and cultural and creative industries. The Big Data Value market in hardware, software and ICT services is a fast growing multi-billion-euro business. In order to unleash the potential rising from the exploitation of Big Data, it is essential to research into new technologies, defined when the size of the data becomes part of the problem itself, together with the tolerable elapsed time: Data Management, including data indexing and query processing, integration and interoperability, data mining and knowledge discovery; Data Storage and Data Processing hardware and software infrastructures, including HPCs, cloud and middleware related aspects; Data and Business Analytics and Business Intelligence with novel algorithms for predictive and prescriptive analytics, and deep use of artificial intelligence, automatic reasoning, machine learning and pattern recognition models; Multimedia Data Processing, dealing with images, videos, graphics and virtual/augmented reality data; Data Protection solutions dealing with privacy and security; Data visualization tools as well as Data Transmission and Communication, in ultra-fast and broadband networks and in mobile nodes.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

- Related e-Infrastructures projects :
  - EUDAT, RDA, EGI-ENGAGE, INDIGO-DATACLOUD, EXANETS
- Major Projects :
  - H2020/FP7 related initiatives: ETP4HPC, NEM, NESSI
  - H2020 projects: ANTAREX, HNSciCloud, R2RAM
  - FP7 projects: ERC MULTITHERMAN, BONE, MIBISOC, E-Policy, JPI AAL: HELICOPTER, PEPPOL, CONNECARE, SAPERE, FORTISSIMO
  - Other relevant projects and initiatives: NESUS , KEYSTONE, CERN experiments (LHCb Data Preservation, LHCbDIRAC), National Clustering Initiative for Smart Community "EDUCATING CITY"
- Datasets: several datasets owned and generated by the major projects listed above.

REGIONAL STAKEHOLDERS

UNIBO, UNIMORE, CINECA, INFN, UNIFE, INAF, CNR, UNIPR, LEPIDA, ENEA

CONTACTS

Paola Salomoni (UNIBO): paola.salomoni@unibo.it
Rita Cucchiara (UNIMORE): rita.cucchiara@unimore.it
Sanzio Bassini (CINECA): s.bassini@cineca.it
Leda Bologni (ASTER): leda.bologni@aster.it

3.2 BIG DATA IN LIFE SCIENCE

Big Data is producing a big hype in healthcare and biomedical scenarios. In the last few years , hospitals, universities and research centres started fruitful and data-productive analyses at European, National and Regional levels such as 'omics studies and multidimensional imaging scans producing zettabytes of data and knowledge. As a matter of fact, a number of case studies in healthcare are well suited for a big data solution. In addition, Big Data allows us to solve clues in common conditions but also to disclose new developments for treating rare diseases steering healthcare towards personalized and "precision" treatments. Now, there is pressing need for fast and preferential connections for data transfer to transform hype in reality and instruments for info aggregation and extrapolation are urgent such as rapid and high performance techniques for machine learning and data mining.

At a regional level a proper and coordinated technology framework (that takes care of context and metadata) will activate a data-driven improvement for a meaningful use of Big Data in healthcare development. 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 answer patients issues and at the same time, at the same time, reduce costs for National and Regional Authorities. Big Data can bring large difficulties and huge problems, but it will leverage the existing scientific knowledge; in fact, a regional infrastructure able to manage and process large-scale data will allow us to run prospective large scale population studies that will give an insight in treating epidemiologically relevant diseases like cardiovascular conditions or osteoporosis.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

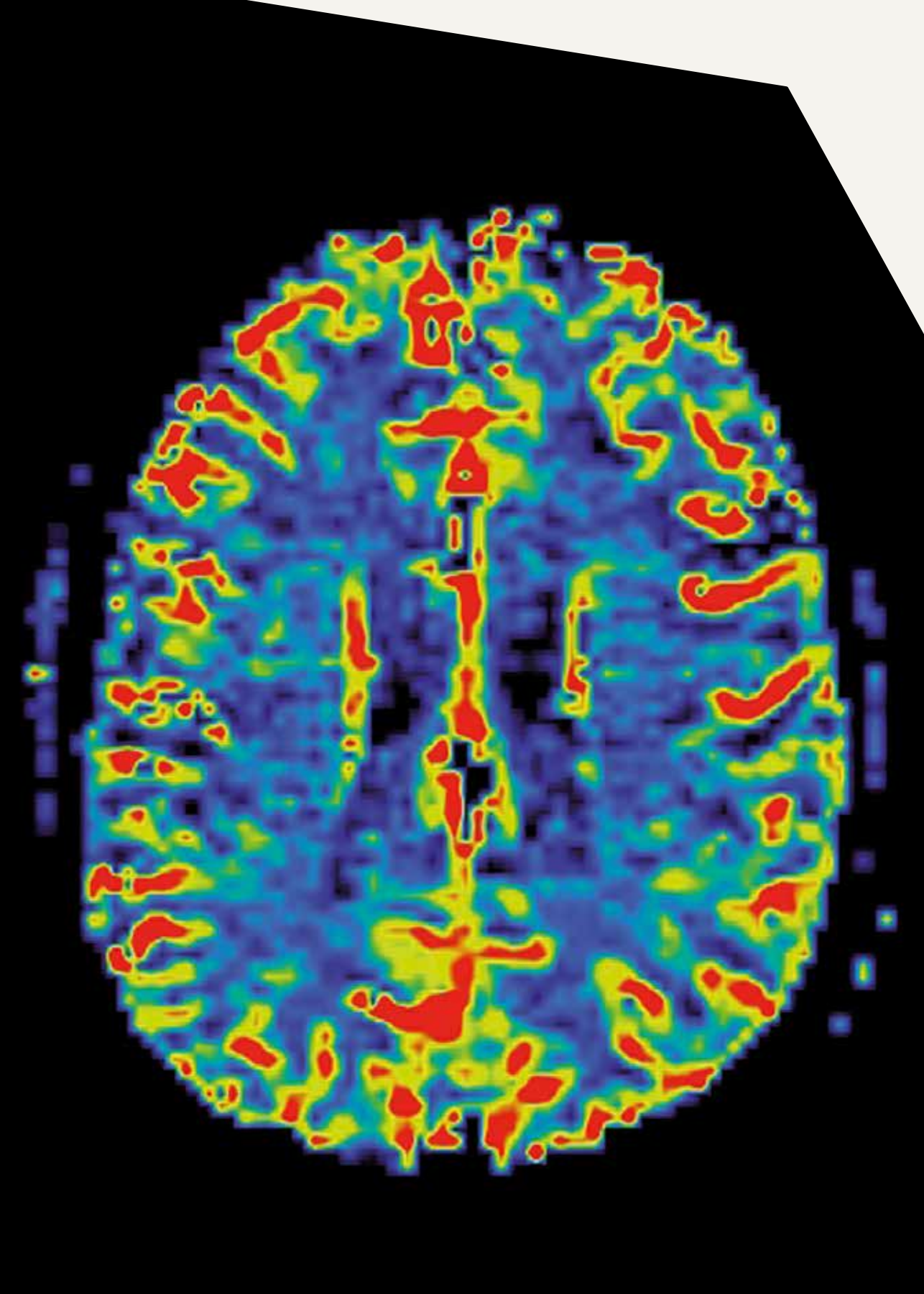
- **Infrastructures**: ELIXIR, BBMRI, EATRIS,
- **Major Projects**:
  - H2020 and FP7: VPH-Share, ADOPT BBMRI-ERIC, CORBEL, PROPAG-AGEING, AFIB2ROTIC, MARK-AGE, FLIP; MEDIGENE, NGS-PTL, MIMOmics, LANGELIN, MISSION T2D, CLEVER, BIO INSPIRED BONE REGENERATION, THALAMOSS, AirPROM, DECERPH, NEUROMICS, B2D2Decide, DIAGNOPTIC, ERC GRANT Program H2020
  - ERA-NET on TRANSCAN: Ma.Tr.Oc. ;
  - ERASMUS+: BIOTECH MA;
  - National funded projects:
    - OPLON - OPportunities for active and healthy LONgevity, IN-BDNF - Physiology and pathophysiology of BDNF,
    - TRAIL - Role of TNF Related Apoptosis Inducing Ligand in diabetes mellitus and metabolic syndrome
  - Regional funded projects:
    - RARER - Next-generation sequencing and gene therapy to diagnose and cure rare diseases in Emilia Romagna region,
    - NGS RARE - Diagnostic accuracy and cost-effectiveness of Next Generation Sequencing strategies in the genetic testing of Rare Orthopedic Diseases ;
    - BioPharmanet – Innovation Lab for Life Sciences.
- **Datasets:**
  - Disease Registries,
  - RIPO Registry of prosthesis,
  - Regional Biobanking Network
  - Several datasets owned and generated by the major projects listed above and by clinical trials and patients cohorts.

REGIONAL STAKEHOLDERS

IOR, UNIBO, UNIMORE, CINECA, INFN, UNIFE, CNR, UNIPR

CONTACTS

Luca Sangiorgi (IOR): luca.sangiorgi@ior.it
Fabio Fava (UNIBO): fabio.fava@unibo.it
Leda Bologni (ASTER): leda.bologni@aster.it



3.3 | BIG DATA IN HUMAN BRAIN AND NEUROSCIENCE COMPUTING

Understanding the human brain is one of the greatest challenges facing 21st century science. High-tech, revolutionary computing and new methods are required for innovative neuroscience and to develop new treatments for neurological and psychiatric diseases that have an increasing burden in an ageing population.

Neuroscience needs effective strategies to better understand brain structures and functions.

Genomics, brain imaging, multielectrode (hundreds) recordings, *in silico* neuroscience generate exponentially growing volumes of data (hundreds of Exabytes) which must be analyzed and integrated in a multidisciplinary way to reveal the mechanisms leading from genes to cells and circuits, to cognition and behaviour, to disease.

The leading role of Emilia-Romagna (ER), where two IIT centres for Neuroscience are located, is witnessed by some of the most important discoveries of the last twenty years and is boosted by CINECA's top-level HPC infrastructure, main partner of the EU Flagship Human Brain Project (HBP) and hosting the HBP massive data analytics supercomputer.

Unprecedented business opportunities, involving a much wider area than the so-called **"biomedical valley",** will be opened. Translational research, development of personalized therapeutic strategies (precision medicine), and implementation of new technological products (such as neuromorphic interfaces or brain-inspired computing systems) will have an impressive social, economic and industrial impact, not only in the so-called "neuromarket", but also in many other contexts, such as robotics, communication media and law (the "neuroethics").

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

- Major Projects:
  - H2020: HBP Flagship SGA 1, ECOMODE,
  - FP7: EYESHOTS, MIPforAction, HBP; POETICON++, EPITARGET, EPIXCHANGE, NEUROFAST, CREAM, ARTEMIS JTI - HIGH PROFILE, ENIAC JTI – CSI, MYNEWGUT, WIRELESS, Parietalaction.
- Datasets:
  - Mouse brain Database produced by LENS (CNR-FI) with light sheet microscopy and hosted in CINECA (100 TeraByte)
  - Several more datasets generated by the major projects listed above.

REGIONAL STAKEHOLDERS

CINECA, UNIFE, UNIBO, UNIPR.

CONTACTS

Giovanni Erbacci (CINECA): g.erbacci@cineca.it
Gaetano Zanghirati (UNIFE): g.zanghirati@unife.it
Leda Bologni (ASTER): leda.bologni@aster.it

3.4 BIG DATA IN AGRI-FOOD AND BIOINDUSTRY

The relevance of Big Data in the agri-food and bio-industry is pivotal, and might become enormous due to the current blooming of innovative technologies and omic tools applied to the most diverse industrial sectors (food production, food safety, agriculture, primary production and animal/plant breeding, industrial biotech, enzyme and microbial discovery). 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 characterization of food spoilage.

Big data has also an increasing relevance in agricultural practices, since the integration of data from sensors, Internet of Things applications and genomics can lead farmers to increase their productivity and sustainability.

Given the relevance of this topic, 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 costs and effectiveness of data acquisition by sharing resources and capabilities.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

- Major Projects
  - JPI: AAL: FOOD, HDHL: ENPADASI
  - H2020 and FP7: PRIME-FISH, WATBIO, TRITICEAGENOME, Nu-AGE, MYNEWGUT, TREASURE, INMARE, BIGCHEM
- National projects: National Clusters Initiative CLAN: PROS.IT - Promotion of consumer's health, nutritional enhancement of Italian traditional agri-food products
- Datasets
  - Datasets owned by UNIBO and UNIPR regarding results from the major projects listed above

REGIONAL STAKEHOLDERS

UNIBO, UNIPR, UNIMORE.

CONTACTS

Stefano Cagnoni (UNIPR): cagnoni@ce.unipr.it
Fabio Fava (UNIBO): fabio.fava@unibo.it
Leda Bologni (ASTER): leda.bologni@aster.it

3.5 BIG DATA IN TRANSPORT

The use of big data in the transport sector is relevant for governments and private companies providing public transportation services (traffic control and congestion management, fleet planning and maintenance, route planning for car/e-car sharing services, train/bus/crew scheduling, etc.), for the private sector (travel industry, logistics, automotive, intelligent transportation systems, etc.) and for individuals (route and travel planning). A smart use of big data supports governments in optimizing multimodal transport and managing traffic flows, making our cities smarter. Citizens can save time and money through the use of route planning support systems, taking into account the huge amount of data coming from GPS systems, real-time traffic monitoring, parking availability, electric charge station availability, fuel cost in different fuel stations, etc. Companies can get competitive advantages from the same data in order to optimize vehicle design, maintenance and energy management, improve on-board automation and safety, and reduce CO_2 emissions. Integrated multimodal fare systems which allow passengers to access multiple forms of transit (such as trains or buses) using a single ticket or card reduce the burden on users, and the design of these systems requires a very thorough analysis of the urban transport modes being integrated. Real-time transportation planning and safety, environmentally sustainable and resource-efficient transport, socio-economic and behavioral research, and forward looking activities for policy making are fundamental topics which greatly benefit from big data.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

- Major Projects
  - FP7: ONTIME, COLOMBO, ARTEMIS JTI-IoE; ARTEMIS JTI ARROWHEAD
  - H2020: SUCCESS
- Datasets
  - Traffic data of regional public transport (buses and trains)

REGIONAL STAKEHOLDERS

Lepida, UNIBO

CONTACTS

Kussai Shahin (LEPIDA): kussai.shahin@lepida.it
Fabio Fava (UNIBO): fabio.fava@unibo.it
Leda Bologni (ASTER): leda.bologni@aster.it

3.6 BIG DATA IN MATERIALS

Data and metadata for materials development, processing, and application life cycles are considered key assets to accelerate discovery and innovation across all design and manufacturing sectors. For its academic and industrial knowledge base and the related production of data, Emilia Romagna competes with the most advanced regions in Europe and worldwide. In this context, the main challenges are often related to the construction and maintenance of well-annotated and structured repositories, their accessibility, their specialization/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 silico* design of materials and (bio)molecular systems 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 color and optical appearance to hydrophobicity, from electron transport to thermal properties within devices. Emilia Romagna 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.

Data acquisition from advanced spectroscopies and imaging of materials is quickly increasing its relevance, and requires sharing and analyzing the resulting highly distributed, heterogeneous data sets. An example in Emilia Romagna 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 with large-scale facilities (e.g. synchrotrons, neutrons), research laboratories equipment, or distributed sensing systems in academic and industrial environments.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

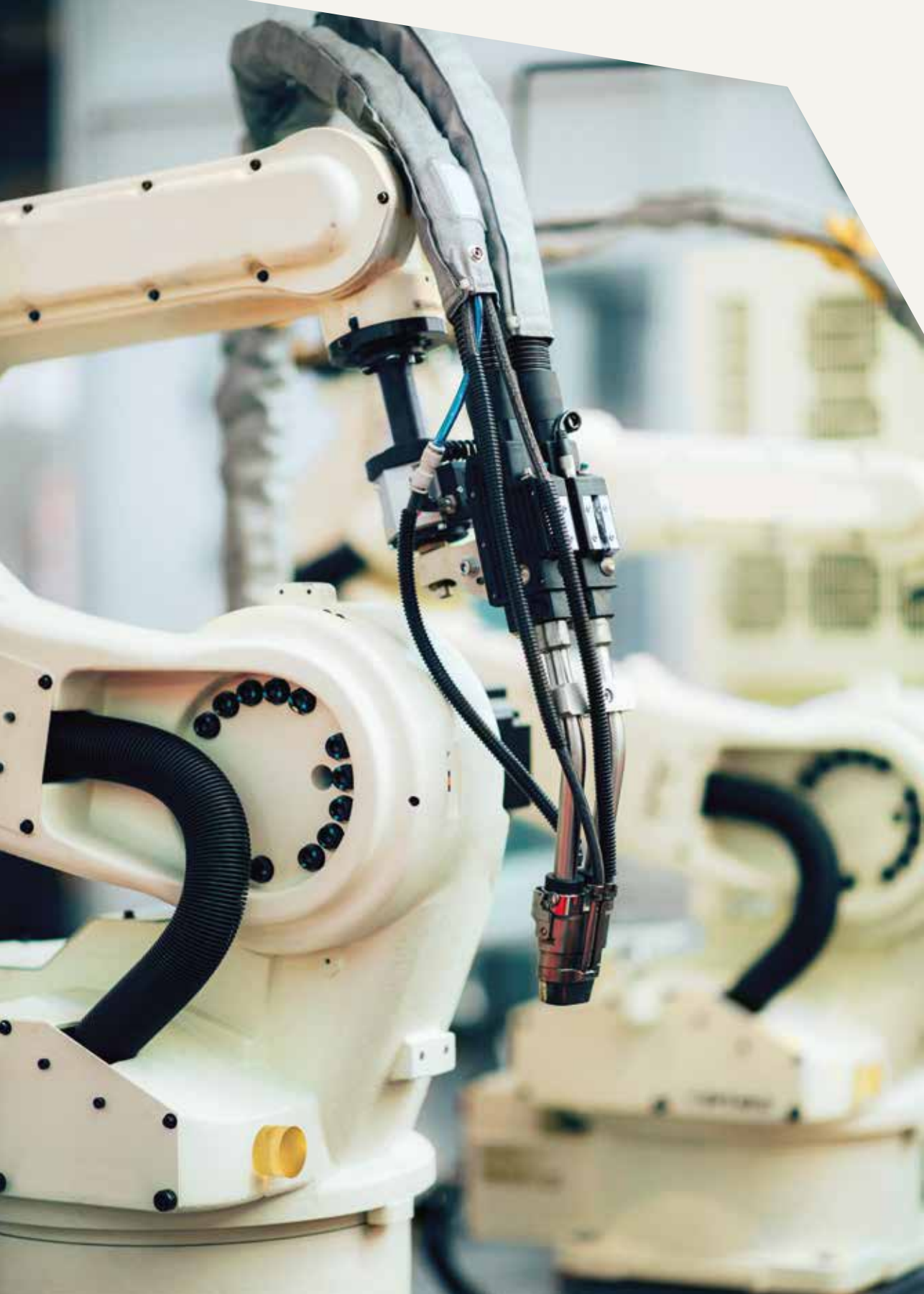
- Related e-infrastructures projects: MaX
- Major projects:
  - H2020/FP7 projects: NANODOME, EXTMOS, GRAPHENE FLAGSHIP, MINOTOR, GRAND, GRADE, III-V-MOS, INACMA, RAINBOW, SUPRABARRIER, BION;, MAGNONMAG, BIGCHEM, UPGRADE, SUN, CRONOS, MODYNA, MOQUAS, TYGRE, TAME-PLASMONS
  - H2020/FP7 Initiatives: KIC-EIT RAW MATERIALS, EERA-JPNM, SINE2020, ERANETS: NANOSCI-ERA+, M-ERANET
- Other relevant projects and initiatives: COST: REDUCIBLE OXIDES, NANOFRICTION, MOPROSURF

REGIONAL STAKEHOLDERS

CNR, UNIBO, UNIMORE, UNIPR, CINECA, INFN, ENEA.

CONTACTS

Elisa Molinari (CNR and UNIMORE): elisa.molinari@unimore.it
Fabio Fava (UNIBO): fabio.fava@unibo.it
Leda Bologni (ASTER): leda.bologni@aster.it



3.7 BIG DATA IN MECHANICS AND INDUSTRIAL PROCESSING

Mechanics and Industrial sectors represent the highest potential efforts in Emilia Romagna activities, in terms of employees, production and business. The research activity is carried out in Research Centres and Industries working in Mechanics, in Industrial Processing and in many related areas such as Mechatronics, Automotive, Manufacturing, Robotics and Automation, Packaging, Electronic Equipment, Textile and Garments, Machinery and Metal production, Ceramics etc.

The changes in production processes and in the production management increasingly require exploitation of knowledge extracted by data, acquired in prototyping, production and testing (i.e. in thermo/fluid-dynamic simulations). The frameworks underlined in Industry 4.0 initiatives, the availability of big sensory data following new paradigms of Internet-of-Things and Cyber Physical Systems, the enormous amount of data coming from simulation before production, need new techniques, instruments, services for data analysis, learning, prediction and statistics.

The manufacturing industry is currently in the midst of a data-driven revolution, from traditional manufacturing facilities to highly optimised smart manufacturing facilities to create manufacturing intelligence from real-time data and support accurate and timely decision-making.

Manufacturing facilities must be capable of meeting the requirements of exponential increase in data production, as well as possessing the analytical techniques needed to extract meaning from big data. Techniques, services and applications, platforms and infrastructure needs are often shared with other research and production fields, in terms of cloud computing and HPC, Big data Analytics and Visualization, Machine Learning and data processing. However, the exploitation of big data and created knowledge in industrial processing is at its infancy: the Emilia Romagna initiatives represent an excellence point in the European panorama of research.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

- Infrastructures: CICLOPE (Centre for International Cooperation in Long Pipe Experiments)
- Major projects:
  - H2020: SYMPLEXITY, SESAME, IPERION CH, CARIM
  - FP7; THERMACO, SHERPA
  - National projects: SILK
  - National cluster Initiative “Fabbrica intelligente”: ADAPTIVE MANUFACTURING, HIGH PERFORMANCE MANUFACTURING and SMART MANUFACTURING
  - Regional projects: DIAMANTE
- Dataset: EUIT

REGIONAL STAKEHOLDERS

(i.e., gli autori dei template usati per preparare la presente scheda)

UNIMORE, UNIBO, CINECA, CNR, UNIPR

CONTACTS

Marcello Pellicciari (UNIMORE): marcello.pellicciari@unimore.it

Fabio Fava (UNIBO): fabio.fava@unibo.it

Leda Bologni (ASTER): leda.bologni@aster.it

3.8 BIG DATA IN ENVIRONMENT AND ENERGY

Big Data in environment and energy refers both to atmospheric model simulations, environmental, geophysical and oceanographic modelling, survey 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.

In the last few years atmospheric models have increased their spatial and temporal resolution and produced a large amount of output data both in diagnostic and in forecast mode. Even the need to store the outputs for further statistics and ensemble modelling increased dramatically the amount of data to be managed and preserved.

Operational ensemble forecasts on the monthly timescale are performed on a weekly basis. Archives of high resolution (about 1 km) atmospheric surface parameters (wind, temperature, solar radiation, etc.) predicted, on a daily basis, up to 48 hours are available for many environmental and renewable energy applications.

Provision of computing services for numerical weather predictions are daily provided.

Chemical Transport Models provide both five days air pollution forecast over Italy and Europe and annual diagnostic simulations over 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 to 12 km.

Future activities will require computing capabilities at least 4 time bigger than the present ones (meteorological modelling is presently performed on an in-house cluster with 12 nodes for a total of 192 cores), with several petabyte storage, in order to double the current spatial resolution and to perform probability prediction of precipitation and increase the resolution of air quality simulations to assess impacts on health, vegetation and cultural heritage. This is necessary in order to maintain the capability to compete at an international level with major international meteorological centres.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

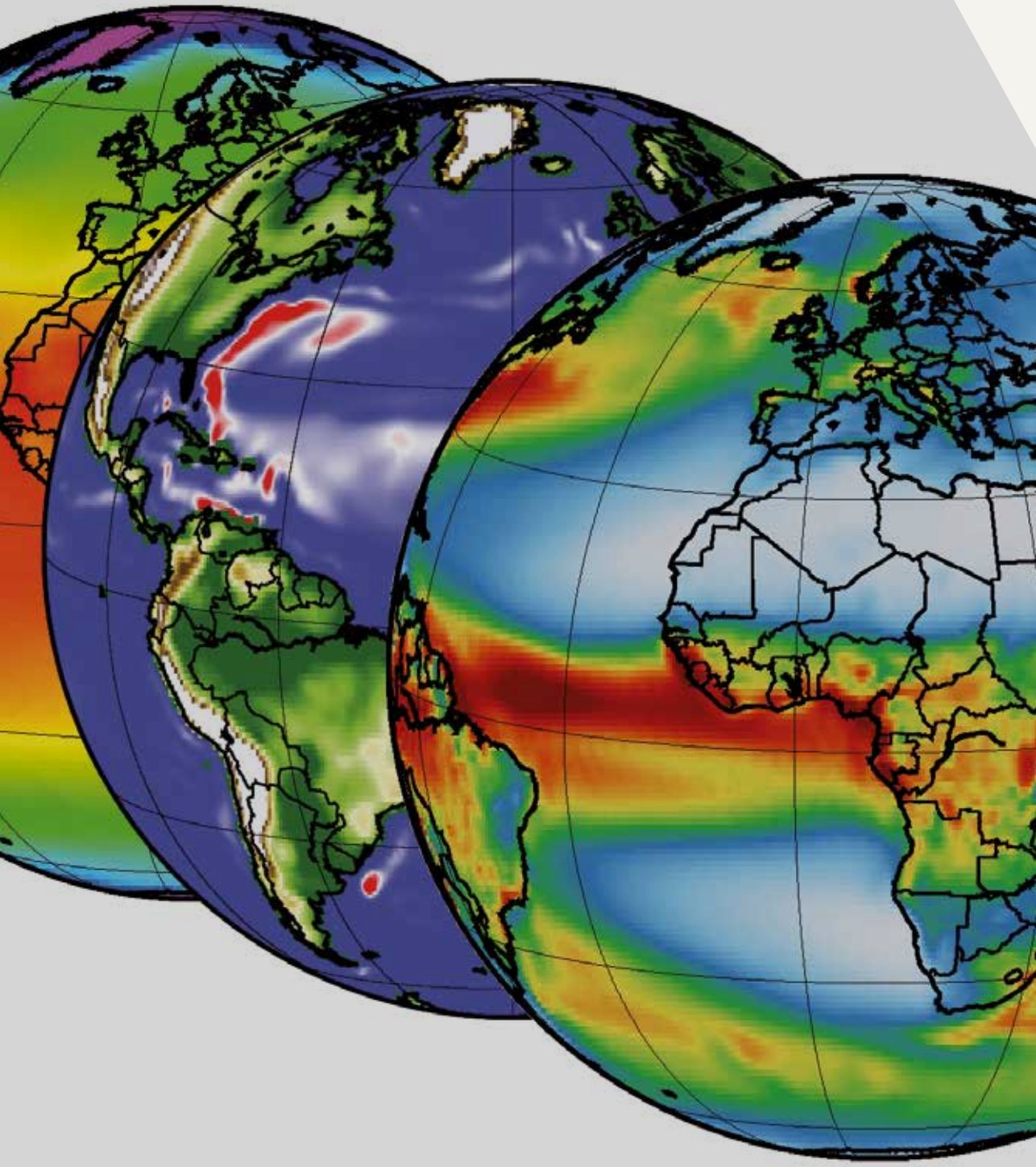
- Infrastructures: EPOS
- Major projects:
  - FP7/H2020: Subseasonal to Seasonal Prediction Project (a WCRP/WWRP project); HYMEX, EU- MED-HISS, EU EcoAdd, GENESI, Resource, NanoReg, TORUS , SYSTEM-RISK, THESEUS, MELODIES, E-AIMS, SEADATANET II, CMEMS, EUCISE2020, MEDSUV,
  - COPERNICUS programme: MARINE ENVIRONMENT MONITORING SERVICE
  - National projects: MINNI - Italian National Integrated Model, ByMuR
- Dataset: EMMA, RCMT, PANDA, GAT@BO

REGIONAL STAKEHOLDERS

ENEA, CINECA, UNIBO, CNR, UNIPR, INGV, UNIFE

CONTACTS

Cristina Sabbioni (CNR): c.sabbioni@isac.cnr.it
Gabriele Zanini (ENEA): gabriele.zanini@enea.it
Leda Bologni (ASTER): leda.bologni@aster.it



3.9 BIG DATA IN CLIMATE CHANGE

Big Data in Climate Change is related to: (i) developing, porting and running Earth-System Models at high spatial resolution; (ii) developing metrics and diagnostic tools for models' evaluation and for specific end-users' needs; (iii) acquisition, storage and processing of data in the order of petabytes from model simulations and observations.

High performance computing, data repository, data sharing and staging services provided are crucial to allow 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. Users include both climate scientists and researchers from a wide range of fields, studying the impacts of climate change and of extremes on topics such as ecosystems, floods, landslides, fires.

In addition, Climate Data is of crucial interest for a number of public/private sectors such as water, energy, agriculture, health, tourism, urban management, transport, impact studies on ecosystems. Overall the output of high-resolution climate simulations is a most valuable dataset at national and regional levels.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

- Major projects:
  - EU-PRACE : Climate SPHINX, DATA SPHINX (EUDAT Data Pilot Project)
  - H2020: PRIMAVERA, IS-ENES2, EUCISE,CRESCENDO, ENVRIPLUS PROJECTS, EARH2OBSERVE
  - FP7: ESFRI SIOS, ETC/CCA, CEOP-AEGIS, ISCAPE, ATLANTOS
  - EU-COPERNICUS (European Programme for the establishment of a European capacity for Earth Observation): Seasonal Prediction Service; Global Climate Projections: Data Access, product generation and impact of front-line developments.
  - National Project of Interest NextDATA
- Datasets: Climate variables at different resolutions up to 16 km for past climate and future scenarios (in production; order 1-2 PB).
- Regional Innovation Community of the EIT Climate KIC

REGIONAL STAKEHOLDERS

CMCC, CNR, CINECA, LEPIDA, UNIBO, UNIMORE, UNIPR, UNIFE

CONTACTS

Cristina Sabbioni (CNR): c.sabbioni@isac.cnr.it
Antonio Navarra (CMCC): antonio.navarra@cmcc.it
Leda Bologni (ASTER): leda.bologni@aster.it


data
social sciences
BIG
computing
web
international
European
LHC
CINECA
research
future
national
CNAF
brain
challenge
relevant
solution
infrastructure
preservation
support
transport
ICT
innovation
opportunities
economic
SCIENCE
Flagship
MECHANICS
climate
Petaflops
curation
terms
processing
storage
service
production
Emilia
Emilia-Romagna
modelling
regional
urban
Istituto
sector
supercomputing
capability
facilities
network
Romagna
genomic
interfaces
content
information
industry
HPC
COMMUNITY
performance
energy
analytic
management
INFN
technologies
introduction
tools
planning
resources
partnership
LEPIDA
amount
Tier
level
digital
UNIBO
facility
field
UNIPR
experiment
capacity
system
industrial
physics
access
safety
require
working
cluster
simulation
analysis
security
cultural
neuroscience
available
strategies
center
knowledge
large
social
scientific
processes
space
smart
disk
Petabyte
model
development
deployment
e-infrastructure
MARCONI
platform
cloud
entered
researchers
Nazionale
Bologna
advanced
healthcare
environment
operational
software
learning
repository
Italy
UNIFE
Europe
forecast
private
initiative
strategies
Fondo
throughput
Central
Foundation
Università
reduce
techniques
Human
Life
weather
Project
CNR
studies
hosts
study
Civil
process
visualization
factor
sharing
hardware
effort
scale
defined
current
order
user
value
integration
reality
Protection
potential
addition
line
increase
peak
manufacturing
topic
CHANGE
spatial
lead
cores
market
sviluppo
competitiveness
centres
connection
currently
integrated
growth
properties
connected
business
Region
Humanities
exceeds
materials
event
acquisition
essential
Universities
equipment
virtual
requirements
benefit
Communication
enormous
structure
capabilities
health
availability
cities
design
fast
volume
physical
ASTER
machine
scientists
increasing
located
europeo
INGV
June
interpretation
application
activities
domain
area
government
servers

3.10 BIG DATA IN SOCIAL SCIENCES AND HUMANITIES

Big data in Social Sciences and Humanities (SSH) has tremendous potential for social and economic impacts, with unprecedented investment opportunities and room for worldwide leadership in research and market. It is in fact strategic for policy-development supporting social innovation and inclusion, for the transmission of European cultural heritage, history, culture and identity, and for enhancing creativity.

This data does not only deal with large and networked cultural datasets, but also calls for new study and interpretation methods in the SSH, including the Digital Humanities field. At least three sides can be recognized in this emerging world: digital objects, interpretations and interfaces.

Massive cultural **digital objects** include large-scale digital corpora; videos, photos and micro-messages shared on social networks; GIS; networks of relations. They strongly promote interdisciplinary activities.

The development of **new interpretive theories** relates to understanding the technical complexity of the data processing pipelines: digitization, transcription, pattern recognition, text, image and video analytics, simulation and inferences, preservation, and curation. Interpretation of SSH big data in the era of Digital Culture implies to deal 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. Interfaces can be essentially immersive, or linguistic, or provide synthetic data interpretation and representation, or they can be of hybrid nature (e.g. augmented reality), challenging both the scientific and the industrial worlds. Nation-wide-leading visualization equipments are available in the Region, such as the VisitLab (CINECA).

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

- Major Projects:
  - H2020: MIREL, INCEPTION, ComEDIA, IMediaCities Reflective,
  - National projects:"Social museum & smart tourism", V-MusT.net
- Datasets
  - REDIGe: network of infrastructures providing access to legal documents
  - Corpora developed at the Department of Classical Philology and Italian Studies of UNIBO: CORIS/CODIS, DiaCORIS, Bononia Legal Corpus
  - Corpora developed at the Department of Interpreting and Translation of UNIBO: EPIC - European Parliament Interpreting Corpus FORLIXT, Multimedia database, WaCky project
  - Multimedia Center CCR-MM: ASFE, Vespasiano da Bisticci letters; Corago; Classical Reception; La Dama Boba
  - Venice and Ravenna Chronicles digital catalogue
  - Photographic archive on academic life and construction of academic buildings
  - "Rodrigo Pais" Photographic Archive
  - The Catalogue of the Federico Zeri Foundation
  - DanteLab - digital library of Dante's 600 MSS (ComEDIA project)
  - MuVi - Museo Virtuale della vita quotidiana
  - Virtual Archeology

REGIONAL STAKEHOLDERS

UNIBO, UNIFE, UNIMORE, CINECA, UNIPR

CONTACTS

Antonino Rotolo (UNIBO): antonino.rotolo@unibo.it
Gaetano Zanghirati (UNIFE): g.zanghirati@unife.it
Leda Bologni (ASTER): leda.bologni@aster.it

3.11 BIG DATA IN SMART CITIES, SAFETY & SECURITY

Big Data is a growing area of interest for public policy makers, for borders, 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 Internet-of-Things paradigm, video-surveillance and environmental cameras, and so on. Smart cities are places where digital technologies translate into better services for citizens and businesses and big data presents great opportunities and it is an essential component that is driving the Smart Cities movement. The analysis of data for social innovation purposes and human-centric services (e.g. personal safety and physical security, crowd-sensing/participatory sensing in 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 data size and timeliness requirements. Connected with smart city and social big data, cyber security (including privacy, access control management, biometrics, etc.) is one of the most critical area in big data analytics and management.

There's a growing demand for security information and event management technologies and services, which gather and analyze security event big data that is used to manage threats.

Big data analytics tools will be the first line of defense, combining machine learning, text mining and abnormal event detection to provide holistic and integrated security threat prediction and recognition, and deterrence and prevention programs . As well, security for cloud services is becoming essential to guarantee confidentiality, integrity and availability of users data of private and public domains.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

- Major Projects:
  - FP7: SNAPSHOT EUCHIPS 2012, BESAFE (NATO Science), PPDR-TC, E-SPONDER, THIS (JLS/CHIP), ePOLICY, DAREED, , FIDELITY, INGRESS, ASTARTE, SPARTACUS, ECOSSIAN, EDEN,
  - H2020: EUO2CEN (EU agr OME), SPYCIT (GOme 2013), NESUS (EU Cost), FLEXMETER
  - Other European/International projects: Central Europe ENERGYCITY, EERA JPs
  - National projects DTCHE; VISERAS; RIGERS; OPEN CITY Platform (OCP), Marche cloud
- Datasets
  - Environment monitoring data (traffic, rain gauge, hydrometer, etc)

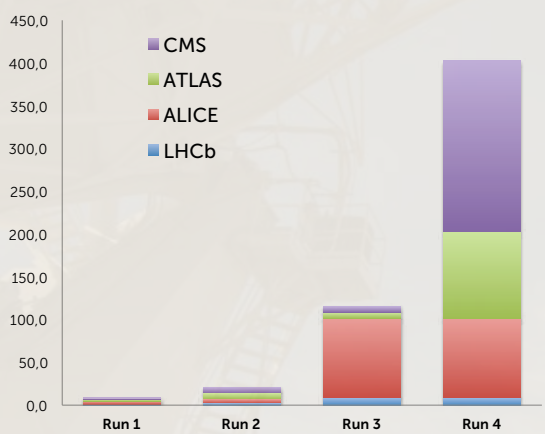
REGIONAL STAKEHOLDERS

LEPIDA, UNIMORE, CNR, UNIBO, UNIFE, INFN, ENEA

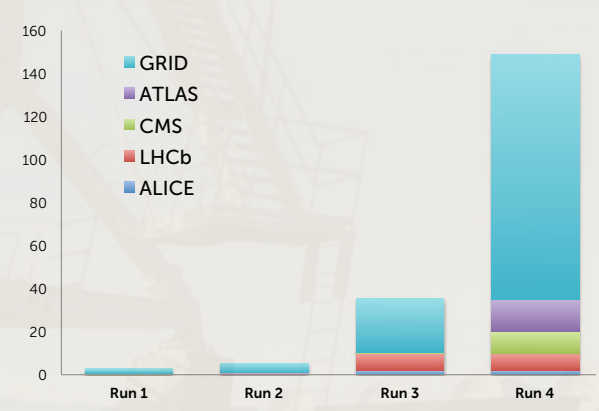
CONTACTS

Kussai Shahin (LEPIDA): kussai.shahin@lepida.it
Rita Cucchiara (UNIMORE): rita.cucchiara@unmore.it
Leda Bologni (ASTER): leda.bologni@aster.it

RAW DISK/TAPE



MHS06



3.12 | BIG DATA IN PHYSICS, ASTRO-PHYSICS AND SPACE SCIENCE

The Emilia Romagna scientific community is active in many research areas both at a National and an International level. The more relevant ones for the Big Data domain are:

- Study of large-scale structure of the universe, cosmology and the nature of black holes
- cosmology and galaxy formation/evolution theoretical modelling
- Simulation of gravitational lensing data also in connection withsearch for gravitational waves
- Quantum Chromo-Dynamics Physics simulations: particles production and properties
- Studies of the Higgs boson properties with 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 Neutrino properties and dark matter characteristics in underground laboratories, in space or underwater conditions

Accelerator particle physics, in particular LHC experiments, astrophysics and space science (e.g. the ESA GAIA space mission for a multidimensional mapping of 1 billion stars in the Galaxy) have already entered in the so called Big Data regime. For example, every year the LHC experiments collect few Petabytes (PB) of data that are copied to the national computing infrastructures, in Italy to the CNAF Tier1. An almost equal amount of simulated data is needed to finalize 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 modeling tools.

In the future the computing resources needed by LHC science, ESA space missions (Euclid, ATHEnA), and large ground-based telescopes (E-ELT, CTA, SKA) will grow significantly: in 2023 LHC will require an increase by a factor about 60 in CPU and 40 in disk space (see Figs 1 and 2). This challenge will force the development of new Big Data technologies and strategies and Emilia Romagna Region will participate in this frontier activities.

INITIATIVES CARRIED OUT IN EMILIA ROMAGNA

- Infrastructures: SKA, CTA, E-ELT, SLHC-PP, KM3NET, SPIRAL2,
- Major Projects:
  - FP7: ERC Cosmic-Lab, ERC GLENCO, CIG eEASy, IEF SIDUN, IOF MWSPEC, ITN GREAT, HADRON PHYSICS 3, Egi-inspire, EMI
  - H2020: EGI-Engage, INDIGO-DataCloud, HNSciCloud, ASTERICS, The ExaNeSt project, HPC-LEAP
  - ESA: EUCLID , “Cosmic Vision”, Planck, GAIA; ATHEnA
  - CERN: ATLAS, CMS, ALICE, LHCb and AMS
  - Other international projects: ALMA, VLBI, space-VLBI
  - National projects: SUMA, COKA, COSA
- Datasets
  - All the datasets of the CERN experiments, CTA and KM3Net experiments.

REGIONAL STAKEHOLDERS

INAF, INFN, UNIBO, UNIFE, UNIPR, UNIMORE

CONTACTS

Giuseppe Malaguti (INAF): malaguti@iasfbo.inaf.it
Antonio Zoccoli (UNIBO and INFN): antonio.zoccoli@unibo.it
Leda Bologni (ASTER): leda.bologni@aster.it

ASTER
innovazione attiva

ASTER is the Consortium among the Emilia-Romagna regional government, the Universities located in the regional territory, the National Research Council (CNR), the Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), the National Institute for Nuclear Physics (INFN), and the Chambers of Commerce regional system working in collaboration with the industrial associations.

ASTER makes the Emilia-Romagna Region innovative and competitive, inclusive and sustainable, creative and open to the world, by promoting innovation for development of the territory and its businesses, enhancing its excellent research resources, the qualified employment of its talents, and the well-being of its residents.


ASTER
innovazione attiva


RegioneEmilia-Romagna


Consiglio Nazionale delle Ricerche


CINECA


ENEA

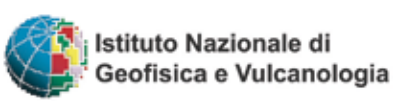

cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici


INAF


SERVIZIO SANITARIO REGIONALE
EMILIA - ROMAGNA


INFN
Istituto Nazionale
di Fisica Nucleare


lepidaspa


Istituto Nazionale di
Geofisica e Vulcanologia


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA


UNIVERSITÀ
DEGLI STUDI
DI FERRARA


UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA


UNIVERSITÀ DEGLI STUDI DI PARMA