EUROPEAN COMMISSION
Directorate-General for Communications Networks, Content and Technology

Excellence in Science
eInfrastructure

11th e-Concertation meeting on e-Infrastructures

Brussels, 9th November 2015

Participant Booklet

Contents

Agenda

Timetable	Topic
09:00 – 09:15	**Registration and coffee**
09:15 – 10:00	**Welcome & Update from EC – WP2016-2017**
10:00 – 11:45	**AM Group Breakout Sessions**
	a. **HPC cluster & Centres of Excellence**
	b. **Biomedical and Life Sciences**
	c. **Earth Sciences, energy and environment**
	d. **Social Sciences and Humanities**
11:00 – 11:15	*Coffee break*
11:45 – 12:30	**Plenary session. Project presentations by AM project groups (focus on synergies)** – (5' presentation + 5' discussion)
12:30 – 13:30	*Networking lunch*
13:30 – 15:15	**PM Group Breakout Sessions**
	e. **Open Scholarships**
	f. **Skills and training**
	g. **Transversal services**
	h. **Integration of data and computing services**
	i. **International Cooperation**
15:15 – 15:45	*Coffee break*
15:45 – 16:45	**Plenary session. Project presentations by PM project groups (focus on synergies)** – (5' presentation + 5' discussion)
16:45 – 17:30	**Discussion, conclusions and wrap-up**
17:30 – 17:45	**Closing of the event**
17:45 - 20:00	*Networking drink*

Welcome

Dear participant,

Thank you for attending the 11th e-concertation meeting for European e-infrastructure projects.

The objective of the meeting is to bring our e-infrastructure projects together to discuss current issues related to the implementation of the first Horizon 2020 projects and future activities in light of the recently adopted Work Programme 2016-2017[1]. We want to foster both cooperation among e-infrastructure projects and sharing of experiences and best practices. To get the conversations started, we have set up a number of project groups whose members are expected to work together during the e-concertation meeting to find synergies enabling them to satisfy the H2020 requirements and to efficiently address the challenges ahead.

We would like you to address issues such as KPIs, TRLs and catalogue of services. These are fundamental novelties under the Work Programme2016-2017 and this is the right time to start preparing the ground for the near future.

We will address issues that require a collective approach, whether related to your project's current objectives or to what will come after its end. Whatever is the path you choose, your involvement is essential to integrate and consolidate our innovative European e-Infrastructures through a robust service offering to support excellent science.

A number of discussions have already started on the Digital4Science platform[2]. We are grateful to those who are already expressing their ideas and points of view and want to encourage all of you to participate in the areas of your interest.

We have prepared this booklet for your convenience. Our objective is to stimulate collaboration and explore synergies beyond boundaries.

Looking forward to meeting you all in Brussels

Augusto Burgueño Arjona

---

[1] *http://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-infrastructures_en.pdf*

[2] *https://ec.europa.eu/futurium/en/digital4science/ideas/e-infrastructures*

Project Groups

HPC cluster with CoEs	Biomedical and life sciences	Earth sciences, energy and environment	Social sciences and humanities
EoCoE	BioExcel	EoCoE	READ
BioExcel	VI-SEEM	ESiWACE	VI-SEEM
NoMaD	MuG	VI-SEEM	VRE4EIC
MaX	VRE4EIC	EVER-EST	OpenAIRE2020
ESiWACE	West-Life	BlueBRIDGE	THOR
E-CAM	OpenAIRE2020	VRE4EIC	Eudat2020
POP	OpenMinTed	OpenMinTed	GEANT FPA
COEGSS	THOR	THOR	EGI-Engage
PRACE	Phenomenal	Earthserver2	
SESAMENET	Eudat2020	Eudat2020	
OpenDreamKit	GEANT FPA	GEANT FPA	
GEANT FPA	EGI-Engage	EGI-Engage	
	PRACE	PRACE	

Open scholarship	Skills and training	Transversal services	Integration of data and computing services	International cooperation
OpenDreamKit	OpenDreamKit	VI-SEEM	OpenAIRE2020	VI-SEEM
West-Life	EDISON	MuG	OpenMinTed	GEANT FPA
EDISON	OpenAIRE2020	VRE4EIC	IndigoDataCloud	TANDEM
OpenAIRE2020	GEANT FPA	AARC	Earthserver2	SciGaia
OpenMinTed	LEARN	GEANT FPA	Phenomenal	MAGIC
THOR	PRACE	PRACE	Eudat2020	EGI-Engage
LEARN		SESAMENET	EGI-Engage	RDA
RDA		EGI-Engage	RDA	PRACE
GEANT FPA			GEANT FPA	
			PRACE	

Group – I: HPC cluster with CoE

Description:

All projects under this group are part of the contractual Public-Private Partnership on HPC; the aim of the cPPP is to address the development of the next generation of HPC technologies, applications and systems towards exascale, and to achieve excellence in HPC application delivery and use. The FET-HPC projects are covering the whole spectrum from processors and system architectures to high-level software and tools to delivering prototype exascale systems and associated applications according to specifications. The establishment of Centres of Excellence (CoEs) will focus and coordinate support to the application of HPC in scientific or industrial domains that are most important for Europe. The overall objective of this group is to coordinate how the projects will most efficiently cooperate and contribute to the HPC strategy. Looking solely at the scientific domains, additional synergies and cooperation could be envisaged for the following areas:

Modeling of materials/molecules/atoms:
- NLAFET, ComPat, MB3, ExaNoDe, ExaNest (FETHPC)
- NoMaD, MaX, ECAM (CoE)

Numerical Weather Forecasting:
- ESCAPE (FETHPC): extreme-scale computing capabilities for European operational numerical weather prediction & future climate models
- ESiWACE (CoE): efficiency and productivity of numerical weather & climate simulation on HPC platforms

Group –II: Biomedical and Life sciences

Description:

The objective of the "Biomedical and life sciences" group is to foster the collaboration of e-infrastructure EC funded projects for the aforementioned scientific domains. The group should improve the way e-infrastructures funded by the EC are used by biomedical and life scientists. Tools and e-infrastructures developed by those focussed projects should be promoted and publicised so that are re-used first by the rest of the cluster scientific domain and second through the most transversal projects by other scientific fields.

Biomedical and life sciences have different limiting factors for the use of ICT technologies, e.g. while the limiting factor for HEP has more to do with the computing power, for this cluster the limiting factor has to do with the storage capacity and analysis. However, this does not mean that each of those constraints affect all sciences. The fact is that there are sciences that have to break specific barriers and when these barriers would become a problem for other sciences they are already open because another domain has done its job.

This cluster should concentrate in breaking tis specific ICT constraints and share the use of that know-how with all scientific disciplines.

Group –III: Earth sciences, energy and environment

Description:

The project group on 'Earth sciences, energy and environment' brings together Virtual Research Environment and Centre of Excellence projects which address diverse topics including: an energy-related centre of excellence (EoCoE); weather and climate simulations in high performance computing (ESiWACE); analytics on big earth data cubes (EarthServer-2); earth science data monitoring (EVER-EST); a data e-infrastructure for marine and fisheries (BlueBRIDGE); use cases on climatology (VI-SEEM) and on life sciences, food and agriculture (OpenMinTed); and support to multi-disciplinary data-driven sciences (VRE4EIC). This thematic project portfolio is complemented by support from e-infrastructure service projects including: a collaborative data e-infrastructure (Eudat2020); the European research and education networking commons (GÉANT); data knowledge and digital services for Open Science (EGI ENGAGE); the European partnership for high performance computing services (PRACE) and digital identifiers for scientific artefacts and researchers (THOR).

 The idea behind the group is to explore synergies, remove overlaps and identify gaps, as well as to share knowledge and increase cooperation among the projects in the group by seeking interoperable solutions, reusing existing applications and technologies, sharing computing facilities, organising joint pilots and demos, and exploring common business or sustainability perspectives.

The members of the 'Earth sciences, energy and environment' group are also expected to engage in reflections concerning integrated e- infrastructure services (IT as a service to research) to enable data and compute intensive science in multiple research fields, the transformation of science (digital, data-driven, open), the evolution towards an open science cloud, etc.

Group –IV: Social Sciences and Humanities

Description:

We would like to facilitate collaboration on Social Sciences and Humanities (SSH) among EC funded e-Infrastructure projects with respect to their role and/or (potential) value for the Social Sciences and Humanities Research.

The idea is to exploit synergies, remove overlaps and identify gaps, as well as to share knowledge among the projects. As the majority of the projects in this project group are not explicitly addressing SSH (only) from a discipline perspective, it allows the focus to be on information exchange, on tools, services and platforms for the SSH community.

Some topics of reflection:

- Are the current tools services and platforms convenient and useable for the SSH community?
- What is your tool/service developing that makes the 'difference'?
- What are main constraints for the uptake of (your) e-Infrastructure in the SSH?
- How will you tailor your support to SSH scientists to help them understand the importance of your service?

Group –V: Open Scholarship

Description:

The objective of the "Open Scholarship Infrastructures" group is to foster the collaboration of EC funded projects in creating an infrastructure that is fit for open science compliant research methods. This infrastructure can include access to information, tools and other resources, but also novel digital services and platforms. It should also acknowledge the rapidly changing scholarly communication landscape and the support that practitioners need in this evolving situation.

Examples of the components of "Open Scholarship Infrastructures" include Infrastructures that will integrate digital scholarly record with the methods of e-Science and Open Science. Users, sponsors and service providers require access to information, from articles to datasets, from software to protocols, from projects to impacts, that needs to be identified across geographical, temporal, disciplinary, cultural, organisational and technological boundaries. The main projects participating in this project group are those working directly with the scholarly record. In addition projects dealing with access technologies, preservation, trust, data production, end-user tools, etc. are also relevant participants in the group activities.

Group –VI: Skills and training

Description:

The main objective of the project group "Skills and training" is to facilitate further collaboration among EC funded e-Infrastructure projects with respect to the new skills and new professions in support of the e-Infrastructure needs and beyond.

The group is expected to map the current training needs of stakeholders and available expertise regarding training development and delivery, as well as to define the existing gaps in matching the needs with the expertise.

The discussion will focus on mechanisms for the awareness raising on the importance of the emerging professions (e-infrastructure operators, research technologists, data scientists, data librarians) in the research process; support building the recognition of these professions; create a reference model which define their competencies; promote adoption of the respective curricula by increasing number of universities and professional training organizations; facilitate networking and sharing of expertise, best practice and trends.

Group –VII: Transversal Services

Description:

The Transversal Services project group is part of the e-Infrastructure project groups (Work programme 2014-15) effort to "exploit synergies, remove overlaps and identify gaps, as well as to share knowledge among the projects in each project group" with a specific emphasis on services that cut across institutions, geographies and communities such as authentication and authorisation services.

Group –VIII: Integration of data and computing services

Description:

Supporting the digitalisation of industry, unlocking the potential of a data and computing intensive science are European priorities to create growth and jobs[3]. The data economy brings also the prospect to offer better and more efficient public services to the citizen.

Europe has been supporting the development of a world class digital infrastructure[4] for research and education spreading across Europe and connecting our continent to the rest of the world. It is the foundation of the finest European tradition in exploring the frontiers of knowledge.

However, preserving European e-infrastructure's excellence is challenged by the convergence of connectivity, data and high-performance computing accessible as integrated services. The development of high-capacity digital infrastructure misses a user-driven service orientation. This is an obstacle to build critical mass and to efficient investments.

The objective of this concertation activity is to brainstorm on the best strategies and concrete actions for Europe to address the above challenges by improving information and communication infrastructures for data and computing offered as a service.

---

[3] Data Harvest report - http://bit.ly/1E6UP8J
[4] "e-Infrastructure"

Group –IX: International Cooperation

Description:

It is well recognised that research is global, and our aim is for every researcher anywhere in the world to have access to research facilities. We want to explore methods and tools to reach this objective. Skill development, training, capacity building are also essential aspects to be taken into consideration.

ICT is now recognised by the UN as contributing to poverty reduction. Numerous e-infrastructures projects focus their work on Africa and Latin America and on how to maximise co-operation with Europe. Key challenges for researchers in developing countries are access, connectivity, computation, data collection and processing; but the potential for innovation and application of ICT for development is high.

Some of the questions to be explored:

- How to achieve synergies and co-operation of projects in this group and beyond?
- How to involve local players and stakeholders including decision makers?
- How ICT skill development and training can be applied locally and transferred to local trainers for sustainability?
- What are the common methods and tools that can be applied by all or most of the projects?
- How to maximise the use of lessons learned in the past?
- How to collect and disseminate local lessons learned effectively and measure KPIs?
- How to give visibility to local/regional innovation and explore wider take-up?

Project Abstracts:

AARC

The goal of AARC is to address technical and functional gaps that prevent the interoperability of existing R&E AAIs.AARC objectives are:

1. Deliver the design of an integrated cross-discipline AAI framework, built on federated access production services (eduGAIN)

2. Increase the uptake of federated access within different research communities

3. Pilot critical components of the proposed integrated AAI where existing production services do not address user needs

4. Validate the results of both the JRA and SA by engaging with the research communities

AARC goals will be achieved by:

• Researching main technical (attribute providers, guest IdPs, support for non-web SSO) and policy aspects (level of assurance, security incidents, etc)

• Supporting (commercial) services relevant to the R&E community

• Delivering training to targeted user communities (e.g. libraries, biomedical, arts and humanities) on both technical and legal aspects

Relevance to the work programme:

1. Facilitate the deployment and promotion of a pan-European identity federation – Met by:

• Designing an AAI framework that builds on federated access and on the evolution of eduGAIN's interfederation approach

• Extensive dissemination of AARC results

2. Lower barriers to entry for organisations – Met by:

• Delivering tools to enable federated access for institutions, services and guest users (also meets obj. Allow for public access at large)

• Delivering technical training tailored to institutions and services following the train-the-trainer model (also meets obj. Offer training and outreach for data professionals)

3. Overcome technical, organisational and legal obstacles – Met by:

• Policies and best practices for operational and security aspects for the integrated AAI

• Training on legal aspects, built on experience with eduGAIN

4. Enable the interoperability of different AAIs – Met by:

• Models to offer collective services and to aggregate their accounting information

• Pilots on security token

BioExcel

Life Science research has become increasingly digital, and this development is accelerating rapidly. Biomolecular modelling techniques such as homology modelling, docking, and molecular simulation have advanced tremendously due to world leading European research, resulting in extreme demands for better computational performance and throughput as these tools are used in applied research and industrial development.

This research has direct influence on our daily life in areas such as health and medical applications, the development of new drugs, efficient drug delivery, biotechnology, environment, agriculture and food industry. Life Science is one of the largest and fastest growing communities in need of high-end computing, and it is a critically important industrial sector for Europe. However, compared to some other disciplines, the use of e-Infrastructure is still relatively new - many advanced techniques are not applied commercially due to limited experience.

It requires significant support to:

- Make e-Infrastructure useable by researchers, who are not computing experts.
- Improve the performance and applicability of key life science applications.
- Handle large amounts of data in computational workflows.

BioExcel proposes to tackle these challenges by establishing a dedicated CoE for Biomolecular Research, covering structural and functional studies of the building blocks of living organisms - proteins, DNA, saccharides, membranes, solvents and small molecules like drug compounds - all areas where with large academic and industrial users bases in Europe. Specifically, BioExcel will:

- Improve the efficiency and scalability of important software packages for biomolecular research.
- Improve the usability of ICT technologies for biomolecular researchers in academia and industry.
- Promote best practices and train end users in making good use of both software and e-Infrastructure.
- Develop appropriate governance structures and business plans for a sustainable CoE.

BlueBRIDGE

BlueBRIDGE responds to the Juncker Investment Plan opening funding opportunities for e-Infrastructures, innovating current practices in producing & delivering scientific knowledge advice to competent authorities & enlarges the spectrum of growth opportunities in distinctive Blue Growth areas.

BlueBRIDGE builds on existing EU and International e-Infrastructures providing capacity building in interdisciplinary research communities of scientists, data managers & educators in academic institutions & industries focusing on 4 major challenges:

1) stock assessment

2) socio-economic performance analysis in aquaculture

3) fisheries & habitat degradation monitoring

4) education & knowledge bridging on protection & management of marine resources.

BlueBRIDGE capitalizes on past investments and uses a proven e-Infrastructure connecting 1500+ scientists, integrating +50 repositories, executing +13,000 models & algorithms/month; providing access to over a billion quality records in repositories worldwide, with 99,7% service availability.

BlueBRIDGE focuses on user needs, opening services & data to actors & liaising with competent agencies & SME Innovation Clusters. Major results include service-driven research environments addressing concrete challenges, data and a wide range of user defined Blue Growth indicators.

BlueBRIDGE leverages a set of common services that together foster the realization of an innovative infrastructure-based approach for collaborative knowledge and data sharing, publishing, citation, traceability & trust concretely contributing to the e-Infrastructure Commons.

BlueBRIDGE will be deployed in 30 months by an authoritative & complementary consortium with expertise in multiple scientific domains. It bundles forces from International Government Organizations, research institutes, industry, SMEs, education and computer science domains, establishing a network with a proven track in VREs & e-Infrastructures, marine, environmental & fisheries science & economy.

COEGSS

Global Systems Science – GSS – is an emerging research field focused on the risks and opportunities involved in global coordination problems. Examples of global systems include the internet, financial markets, intellectual property rights, global energy use and others.

Developing evidence and understanding in view of such systems and of related policies is rapidly becoming a vital challenge for modern societies. It requires capabilities for transdisciplinary work that cannot be mastered without massive use of ICT.

By the nature of the problem, the relevant datasets are mostly very big, including data streams from social media. To make things more complicated, the relevant algorithms do require the power of high-performance computing. High Performance Data Analysis (HPDA) is the key to success for GSS!

A key contribution of the Centre of Excellence for Global Systems Science – COEGSS – will be the development of an HPC-based framework to generate customized synthetic populations for GSS applications. By blending GSS and HPC, we will be able to provide decision makers and civil society with real-time assessments of global risks and opportunities as well as with essential background knowledge about them. This will enable the HPC industry to supply hard- and software for applications well beyond the issues to which HPC has been dedicated so far.

Earthserver2

EarthServer-2 makes Agile Analytics on Big Earth Data Cubes of sensor, image, simulation, and statistics data a commodity for non-experts and experts alike through:

- navigation, extraction, aggregation, and recombining of any-size space/time data cubes;
- easy to install & maintain value-adding services extending the existing portfolio of data and compute centers;
- based on open standards, in particular: the OGC Big Data standards and the forthcoming ISO SQL/MDA ("Multi- Dimensional Arrays") standard.

In the Joint Research Activity, the project will advance the existing, world-leading rasdaman Array Database technology wrt. query functionality, inter-federation data processing with automatic data and query distribution, tape archive integration, and 3D/4D visualization based on NASA's virtual globe technology.

In the Services Activity, large data centers (ECMWF, PML, MEEO/ESA, GeoScience Australia, JacobsUni) will set up water, air, weather, and planetary services on 3D & 4D data cubes up to Petabyte-size with user-tailored clients for both visual and textual ad-hoc mix&match.

In the Networking Activity, the project will advance open Big Data standards in OGC, RDA, and ISO (in particular: write ISO SQL/MDA). Further, all adequate channels will be used for strong dissemination & exploitation, specifically: writing a monograph explaining OGC Big Geo Data standards; scientific publications & active conference organization; Earth science data user workshops for each domain addressed; actively contributing technology & experience to GEO / GEOSS and further bodies; establish standardized Big Geo Data benchmark and run it against EarthServer-2 and further relevant systems.

Altogether, EarthServer-2 will maintain and extend the lead in Big Earth Data services established in the highly successful EarthServer-1 project. Being already supported by ESA, rasdaman will form an enabling building block for COPERNICUS / Sentinel.

E-CAM

E-CAM will create, develop and sustain a European infrastructure for computational science applied to simulation and modelling of materials and of biological processes of industrial and societal importance. Building on the already significant network of 15 CECAM centres across Europe and the PRACE initiative, it will create a distributed, sustainable centre for simulation and modelling at and across the atomic, molecular and continuum scales.

The ambitious goals of E-CAM will be achieved through three complementary instruments:

1. development, testing, maintenance, and dissemination of robust software modules targeted at end-user needs;

2. advanced training of current and future academic and industrial researchers able to exploit these capabilities;

3. multidisciplinary, coordinated, top-level applied consultancy to industrial end-users (both large multinationals and SMEs).

The creation and development of this infrastructure will also impact academic research by creating a training opportunity for over 300 researchers in computational science as applied to their domain expertise.

It will also provide a structure for the optimisation and long-term maintenance of important codes and provide a route for their exploitation. Based on the requests from its industrial end-users, E-CAM will deliver new software in a broad field by creating over 200 new, robust software modules. The modules will be written to run with maximum efficiency on hardware with different architectures, available at four PRACE centres and at the Hartree Centre for HPC in Industry. The modules will form the core of a software library (the E-CAM library) that will continue to grow and provide benefit well beyond the funding period of the project.

E-CAM has a 60-month duration, involves 48 staff years of effort, has a total budget of €5,836,897 and is requesting funding from the EC of €4,836,897, commensurate with achieving its ambitious goals.

EDISON

The EDISON project will focus on activities to establish the new profession of ' Data Scientist', following the emergence of Data Science technologies (also referred to as Data Intensive or Big Data technologies) which changes the way research is done, how scientists think and how the research data are used and shared. This includes definition of the required skills, competences framework/profile, corresponding Body of Knowledge and model curriculum. It will develop a sustainability/business model to ensure a sustainable increase of Data Scientists, graduated from universities and trained by other professional education and training institutions in Europe.

To achieve this, EDISON will work with the major Data Science stakeholders from academic, research communities and industry, as well as with the professional community to help them to obtain proper education and training and/or formal certification for already practicing self-made

Data Scientists, grown from the advanced research projects who want to build a new career in Data Science.

Consistent Data Science education and professional training requires besides theoretical knowledge access to real scientific data infrastructure and real large data sets to acquire practical experience and develop data centric thinking. For this, EDISON will leverage on EGI infrastructure and community/activities, as well as products from the APARSEN project, to create a supporting infrastructure for Data Science education and training that will include both example datasets and virtual labs which will allow the students or trainees to work with real data sets, infrastructure and tools.

EDISON will facilitate the establishment of a Data Science education and training infrastructure at major European universities by promoting experience of 'champion' universities involving them into coordinated development and implementation of the model curriculum and creation of cooperative educational and training infrastructure.

EGI-Engage

Over the last decade, the European Grid Infrastructure (EGI) has built a distributed computing and data infrastructure to support over 21,000 researchers from many disciplines with unprecedented data analysis capabilities. EGI builds on the European and national investments and relies on the expertise of EGI.eu - a not-for-profit foundation that provides coordination to the EGI Community, including user groups, EGI.eu participants in the EGI Council, and the other collaborating partners.

The mission of EGI-Engage is to accelerate the implementation of the Open Science Commons vision, where researchers from all disciplines have easy and open access to the innovative digital services, data, knowledge and expertise they need for their work. The Open Science Commons is grounded on three pillars: the e-Infrastructure Commons, an ecosystem of key services; the Open Data Commons, where any researcher can access, use and reuse data; and the Knowledge Commons, in which communities have shared ownership of knowledge and participate in the co-development of software and are technically supported to exploit state-of-the-art digital services.

EGI-Engage will expand the capabilities offered to scientists (e.g. improved cloud or data services) and the spectrum of its user base by engaging with large Research Infrastructures (RIs), the long-tail of science and industry/SMEs. The main engagement instrument will be a network of eight Competence Centres, where National Grid Initiatives (NGIs), user communities, technology and service providers will join forces to collect requirements, integrate community-specific applications into state-of-the-art services, foster interoperability across e-Infrastructures, and evolve services through a user centric development model. The project will also coordinate the NGI efforts to support the long-tail of science by developing ad hoc access policies and by providing services and resources that will lower barriers and learning curves.

EoCoE

The aim of the present proposal is to establish an Energy Oriented Centre of Excellence for computing applications, (EoCoE). EoCoE (pronounce "Echo") will use the prodigious potential offered by the ever-growing computing infrastructure to foster and accelerate the European transition to a reliable and low carbon energy supply.

To achieve this goal, we believe that the present revolution in hardware technology calls for a similar paradigm change in the way application codes are designed. EoCoE will assist the energy transition via targeted support to four renewable energy pillars: Meteo, Materials, Water and Fusion, each with a heavy reliance on numerical modelling. These four pillars will be anchored within a strong transversal multidisciplinary basis providing high-end expertise in applied mathematics and HPC.

EoCoE is structured around a central Franco-German hub coordinating a pan-European network, gathering a total of 8 countries and 23 teams. Its partners are strongly engaged in both the HPC and energy fields; a prerequisite for the long-term sustainability of EoCoE and also ensuring that it is deeply integrated in the overall European strategy for HPC.

The primary goal of EoCoE is to create a new, long lasting and sustainable community around computational energy science. At the same time, EoCoE is committed to deliver high-impact results within the first three years. It will resolve current bottlenecks in application codes, leading to new modelling capabilities and scientific advances among the four user communities; it will develop cutting-edge mathematical and numerical methods, and tools to foster the usage of Exascale computing.

Dedicated services for laboratories and industries will be established to leverage this expertise and to foster an ecosystem around HPC for energy. EoCoE will give birth to new collaborations and working methods and will encourage widely spread best practices.

ESiWACE

ESiWACE will substantially improve efficiency and productivity of numerical weather and climate simulation on high performance computing platforms by supporting the end-to-end workflow of global Earth system modelling in HPC environment.

This will be obtained by improving and supporting (1) scalability of models, tools and data management on state-of-the-art supercomputer systems (2) Usability of models and tools throughout the European HPC eco-system, and (3) the Exploitability of the huge amount of resulting data.

We will develop solutions for cross-cutting HPC challenges particular to the weather and climate domain. This will range from the development of specific software products to the deployment of user facing services for both, computing and storage.

ESiWACE leverages two established European networks, namely (1) the European Network for Earth System modelling, representing the European climate modelling community and (2) the world leading European Centre for Medium-Range Weather Forecasts. The governance structure

that defines the services to be provided will be driven by the European weather and climate science community.

Weather and climate computing have always been one of the key drivers for HPC development, with domain specific scientific and technical requirements that stretch the capability and capacity of existing software and hardware to the limits.

By developing solutions for Europe and at European scale, ESiWACE will directly impact on the competitiveness of the European HPC industry by engendering new products, providing opportunities for exploitation beyond the project itself, and by enhancing the skills base of staff in both industry and academia.

ESiWACE will be at once thematic, as it focuses on the HPC application domain of climate and weather modelling, transversal, as it covers several aspects of computational science, and challenge-driven, as climate and weather predictability represents a major societal issue.

Eudat2020

EUDAT2020 brings together a unique consortium of e-infrastructure providers, research infrastructure operators, and researchers from a wide range of scientific disciplines under several of the ESFRI themes, working together to address the new data challenge.

In most research communities, there is a growing awareness that the "rising tide of data" will require new approaches to data management and that data preservation, access and sharing should be supported in a much better way. Data, and a fortiori Big Data, is a cross-cutting issue touching all research infrastructures.

EUDAT2020's vision is to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment, as part of a Collaborative Data Infrastructure (CDI) conceived as a network of collaborating, cooperating centres, combining the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centres.

EUDAT2020 builds on the foundations laid by the first EUDAT project, strengthening the links between the CDI and expanding its functionalities and remit. Covering both access and deposit, from informal data sharing to long-term archiving, and addressing identification, discoverability and computability of both long-tail and big data, EUDAT2020's services will address the full lifecycle of research data.

One of the main ambitions of EUDAT2020 is to bridge the gap between research infrastructures and e-Infrastructures through an active engagement strategy, using the communities that are in the consortium as EUDAT beacons and integrating others through innovative partnerships.

During its three-year funded life, EUDAT2020 will evolve the CDI into a healthy and vibrant data-infrastructure for Europe, and position EUDAT as a sustainable infrastructure within which the future, changing requirements of a wide range of research communities are addressed.

EVER-EST

The EVER-EST project will enhance Research and Capacity Building in the Earth Science (ES) domain, by providing a generic Service Oriented-based Architecture Virtual Research Environment (VRE) tailored to the needs of the ES community. The full range of both generic and ES data management specific services provided by EVER-EST will facilitate the dynamic approach to collaborative working and research. Scientist will be able to discover, access, process, share and interoperate heterogeneous data, algorithms, results and experiences within and across their communities, including those in domains beyond Earth Science. Data providers will be able to monitor users experience and collect feedback through the VRE, improving their capacity to adapt to end-user needs offering solutions to unlock scientist potentials and creativity.

EVER-EST e-infrastructure will be validated by four Virtual Research Communities (VRC) covering different multidisciplinary working ES domains: ranging from Sea Monitoring, to Natural Hazards (floods, weather, wildfires), Land Monitoring and Risk Management (volcanoes and seismic). Each of these VRC will use the VRE according to its own specific requirements for data, software, best practices and community engagement. The VRE will leverage on the results of several FP7 projects which have produced state-of-the-art technologies for scientific data management and curation.

Such capabilities will be enriched during the EVER-EST project using the output of a second group of FP7 initiatives providing models, techniques and tools for the preservation of scientific methods and their implementation in computational forms such as scientific workflows. EVER-EST will implement solutions for data publication, citation and trust. EVER-EST will provide the means to climb up existing peaks and obstacles to information sharing in Earth Science allowing research teams to collaboratively collect, manipulate and share resources, allowing creation of new knowledge.

GEANT

The overall objective is to provide a stable environment for the implementation of GÉANT as the European Communications Commons for the European Research Area, which will provide the best possible digital infrastructure to ensure that Europe remains in the forefront of research.

GÉANT's extensive and long-standing contacts with large data disciplines such as biology, radio-astronomy and high-energy physics help shape the evolution of the networking facilities required. This is complemented by partner contacts with research domains that are new users of high performance networks and services, such as: digital preservation, real-time art and humanities. The vision is to position the GÉANT partnership optimally to achieve the strategic objectives of the FPA while ensuring the continuity and improvement of the services successfully offered under the GN3plus FP7 project. The extensive experience of the GÉANT partnership in providing high-quality and innovative services has been applied to the preparation phase of this proposal. The following GN4-1 work package objectives deserve to be highlighted:

• Maintain and enhance the production service and achieve cost reductions without negative effects on network service levels.
• Coordinate with the other European e-infrastructure efforts both individually and in the framework of the increased coordination effort.

- Prepare a new round of Open Call projects to start immediately after the end of the GN4-1 project.
- Expand the footprint and the depth of services offered to roaming users with enhanced trust and identity services through Federation as a Service and with group and attribute management added to eduGAIN.
- Develop online real-time services to improve the ease and function of videoconferencing, open learning support and general multimedia use.
- Review the existing service catalogue using the Product Lifecycle Management process to assess the cost/benefits of each carefully to define the future service strategy.

IndigoDataCloud

The INDIGO-DataCloud project (INDIGO for short) aims at developing a data/computing platform targeted at scientific communities, deployable on multiple hardware, and provisioned over hybrid (private or public) e-infrastructures. This platform will be built by leading European developers, resource providers, e-infrastructures and scientific communities in order to ensure its successful exploitation and sustainability.

All members of the consortium share the common interest in developing advanced middleware to sustain the deployment of service models and user tools to tackle the challenges of the Big Data era. INDIGO will exploit the formidable know-how that was built in Europe along the past ten years of collaborations on scientific computing based on different consolidated and emerging paradigms (HPC, Grid and Cloud).

Regarding Cloud computing, both the public and private sectors are already offering IaaS-type Cloud resources. However, numerous areas are of interest to scientific communities where Cloud computing uptake is currently lacking, especially at the PaaS and SaaS levels. The project therefore aims at developing tools and platforms based on open source solutions addressing scientific challenges in the Cloud computing, storage and network areas.

INDIGO will allow application development and execution on Cloud and Grid based infrastructures, as well as on HPC clusters. The project will extend existing PaaS solutions, allowing public and private e-infrastructures, including those provided by EGI, EUDAT, PRACE and HelixNebula, to integrate their existing services, make them available through GEANT-compliant federated and distributed AA policies, guaranteeing transparency and trust in the provisioning of such services.

INDIGO will also address the development of a flexible and modular presentation layer connected to the expanded PaaS and SaaS frameworks developed by the project and allowing innovative user experiences, also from mobile appliances.

LEARN

We live in the age of the 'data deluge', where digital technology enables us to store petabytes of data and to make that available for sharing as open data. Sharing data has the potential to revolutionise the way that researchers work. It avoids costly duplication in the collecting of data and enables research collaborations across the world which otherwise would not be possible.

The purpose of this bid is to take the LERU Roadmap for Research Data produced by the League of European Research Universities (LERU) and to develop this in order to build a coordinated e-infrastructure across Europe and beyond. LEARN will deliver a model for Research Data Management (RDM) policy, a Toolkit to support implementation, and an Executive Briefing in five core languages so as to ensure wide outreach.

LEARN will hold a series of Workshops within four European countries and one international country. The workshops will serve to advocate the Recommendations on RDM and open data made by the LERU Roadmap, and gain feedback from Workshop attendees for a new Toolkit of best practice. Furthermore, guidance to implement the Roadmap will be provided following identification of Best Practices supported by case studies identified through the workshops.

By producing an exemplar RDM policy, which could then be tailored by any university or research institution to meet their needs, LEARN aims to address the challenges of the Work Programme concerning the fragmentation of e-infrastructures and the need to maximize on global research data. Specifically, LEARN will address Stakeholder initiatives; Policy coordination; Take-up of digital infrastructures; and Support cooperation with developing countries. LEARN thus delivers support actions to quicken the take-up of RDM and the move to open data in the emerging world of Science 2.0.

MAGIC

The MAGIC Project seeks to establish a set of agreements for Europe, Latin America and other participating World Regions, aiming at consolidating and completing the building blocks of middleware necessary for the establishment of a marketplace of services and real-time applications for international and inter-continental research groups which facilitates mobility and the work of global science communities. More specifically, the project specific goals are:

- To foster the deployment of the platforms that enable mobility of people and seamless access to services by promoting the establishment of identity federations interconnected via eduGAIN, creating awareness of privacy and security issues and spreading eduroam by: sharing experiences, carrying out training and making available shared infrastructure to ease the process in the partner world regions.

- To develop a model for inter-operation between NREN cloud application markets of participating world regions, based on cloud provisioning and taking advantage of applications developed and run by NRENs across different continents to create a model for a worldwide application market for collaboration tools and services

- To seek consensus among participating world regions on the importance of interoperability of real-time applications and work towards the adoption of standards such as those proposed by the Global CEO Forum to promote the creation of a worldwide environment for these applications.

- To foster the collaborative work of Global Science Communities by actively promoting the participation of Latin American and other regions' researchers in European Commission

Calls and those of other international funding agencies with high impact in the participating regions and other networking activities.

MaX

Materials are crucial to scientific and technological advances and industrial competitiveness, and to tackle key societal challenges - from energy and environment to health care, information and communications, manufacturing, safety and transportation.

The current accuracy and predictive power of materials' simulations allow a paradigm shift for computational design and discovery, in which massive computing efforts can be launched to identify novel materials with improved properties and performance; behaviour of ever-increasing complexity can be addressed; sharing of data and work-flows accelerates synergies and empowers the science of big-data; and services can be provided in the form of data, codes, expertise, turnkey solutions, and a liquid market of computational resources.

Europe has the human resources, track record and infrastructure to be worldwide leader in this field, and we want to create a CoE in materials' modelling, simulations, and design to endow our researchers and innovators with powerful new instruments to address the key scientific, industrial and societal challenges that require novel materials.

This CoE will be a user-focused, thematic effort supporting the needs and the vision of all our core communities: domain scientists, software scientists and vendors, end-users in industry and in academic research, and high-performance computing centres.

The proposal is structured along two core actions:

(1) Community codes, their capabilities and reliability; provenance, preservation and sharing of data and work-flows; the ecosystem that integrates capabilities; and hardware support and transition to exascale architectures.
(2) Integrating, training, and providing services to our core communities, while developing and implementing a model for sustainability, with the core benefit of propelling materials simulations in the practice of scientific research and industrial innovation.

MuG

Genomics is probably the fastest evolving field in current science. A decade ago our main concern was to obtain the sequence (the 1D code) of the genome; but today the big challenges are to determine how genotype information is transferred into phenotype, and how pathological phenotypic changes can be predicted from genome alterations. While investigating these points, we have realized that a part of the regulation of gene expression is implicitly coded in the way in which chromatin is folded.

As technology has advanced and information of the folded state of chromatin has emerged, a new branch of genomics (3D/4D genomics) has emerged. Hundreds of laboratories are now defining a young and active community that, though in the end concerned with the same scientific problem, uses many different approaches to study it that individually target radically different length and timescales. The community faces severe practical problems related to:

> i) how huge, noisy, and diverse data related to widely different size and time scales can be integrated,
> ii) the lack of standardized analysis and simulation tools,
> iii) the complete disconnection of associated informatics databases, and
> iv) the lack of validated and flexible visualization engines.

MuG is born at the critical point in the evolution of the field, in a bottom up approach from the biologist who are suffering severe IT problems. MuG, supported by European leaders in the field, join three different expertises: biologist with interest in chromatin structure, methods developers and HPC facilities with strong history of supporting Bio-computational problems. We believe that MuG will be a steep-forward in approaching the potential of High Performance Computing to the development of 3D/4D genomics, and will contribute to give a structure to this new and exciting field.

NoMaD

Essentially every new commercial product, be they smart phones, solar cells, batteries, transport technology, artificial hips, etc., depends on improved or even novel materials. Computational materials science is increasingly influential as a method to identify such critical materials for both R&D. Enormous amounts of data, precious but heterogeneous and difficult to access or utilise, are already stored in repositories scattered across Europe. The NoMaD CoE will open new HPC opportunities by enabling access to this data and delivering powerful new tools to search, retrieve and manage it.

NoMaD will foster sharing of all relevant data, building on the unique CECAM, Psi-k and ETSF communities, putting Europe ahead of materials science in other continents. Unprecedented, already initialised networking with researchers, with industry, with students and with other stakeholders will guarantee relevance and end-user value. NoMaD will become a crucial tool for atomistic simulations and multi-scale modelling in the physical, materials, and quantum-chemical sciences.

This field is characterised by a healthy but heterogeneous eco-system of many different codes that are used at all HPC centres worldwide, with millions of CPU hours spent every day, some of them at petascale performance. NoMaD will integrate the leading codes and make their results

comparable by converting (and compressing) existing inputs and outputs into a common format, thus making these valuable data accessible to academia and industry.

NoMaD will develop "big-data analytics" for materials science. This will require novel algorithms, e.g., for statistical learning based on the created materials encyclopaedia, offering complex searches and novel visualisations. These challenges exploit the essential resources of our HPC partners. Without the infrastructure and services provided by the NoMaD CoE, much of the information created with the above mentioned petascale (towards exascale) computations would be wasted.

OpenAIRE2020

OpenAIRE2020 represents a pivotal phase in the long-term effort to implement and strengthen the impact of the Open Access (OA) policies of the European Commission (EC), building on the achievements of the OpenAIRE projects. OpenAIRE2020 will expand and leverage its focus from:

(1) the agents and resources of scholarly communication to workflows and processes,
(2) from publications to data, software, and other research outputs, and the links between them, and
(3) strengthen the relationship of European OA infrastructures with other regions of the world, in particular Latin America and the U.S.

Through these efforts OpenAIRE2020 will truly support and accelerate Open Science and Scholarship, of which Open Access is of fundamental importance. OpenAIRE2020 continues and extends OpenAIRE's scholarly communication infrastructure to manage and monitor the outcomes of EC-funded research.

It combines its substantial networking capacities and technical capabilities to deliver a robust infrastructure offering support for the Open Access policies in Horizon 2020, via a range of pan-European outreach activities and a suite of services for key stakeholders. It provides researcher support and services for the Open Data Pilot and investigates its legal ramifications. The project offers to national funders the ability to implement OpenAIRE services to monitor research output, whilst new impact measures for research are investigated.

OpenAIRE2020 engages with innovative publishing and data initiatives via studies and pilots. By liaising with global infrastructures, it ensures international interoperability of repositories and their valuable OA contents.

To ensure sustainability and long-term health for the overall OpenAIRE infrastructure, the proposed OpenAIRE2020 project will establish itself as a legal entity, which will manage the production-level responsibilities securing 24/7 reliability and continuity to all relevant user groups, data providers and other stakeholders.

OpenDreamKit

OpenDreamKit will deliver a flexible toolkit enabling research groups to set up Virtual Research Environments, customised to meet the varied needs of research projects in pure mathematics and applications and supporting the full research life-cycle from exploration, through proof and publication, to archival and sharing of data and code.

OpenDreamKit will be built out of a sustainable ecosystem of community-developed open software, databases, and services, including popular tools such as LinBox, MPIR, Sage(sagemath.org), GAP, PariGP, LMFDB, and Singular. We will extend the Jupyter Notebook environment to provide a flexible UI. By improving and unifying existing building blocks, OpenDreamKit will maximise both sustainability and impact, with beneficiaries extending to scientific computing, physics, chemistry, biology and more and including researchers, teachers, and industrial practitioners.

We will define a novel component-based VRE architecture and adapt existing mathematical software, databases, and UI components to work well within it on varied platforms. Interfaces to standard HPC and grid services will be built in. Our architecture will be informed by recent research into the sociology of mathematical collaboration, so as to properly support actual research practice. The ease of set up, adaptability and global impact will be demonstrated in a variety of demonstrator VREs.

We will ourselves study the social challenges associated with large-scale open source code development and of publications based on executable documents, to ensure sustainability.

OpenMinTeD

Recent years witness an upsurge in the quantities of digital research data, offering new insights and opportunities for improved understanding. Text and data mining is emerging as a powerful tool for harnessing the power of structured and unstructured content and data, by analysing them at multiple levels and in several dimensions to discover hidden and new knowledge. However, text mining solutions are not easy to discover and use, nor are they easily combinable by end users.

OpenMinTeD aspires to enable the creation of an infrastructure that fosters and facilitates the use of text mining technologies in the scientific publications world, builds on existing text mining tools and platforms, and renders them discoverable and interoperable through appropriate registries and a standards-based interoperability layer, respectively. It supports training of text mining users and developers alike and demonstrates the merits of the approach through several use cases identified by scholars and experts from different scientific areas, ranging from generic scholarly communication to literature related to life sciences, food and agriculture, and social sciences and humanities.

Through its infrastructural activities, OpenMinTeD's vision is to make operational a virtuous cycle in which
   a) primary content is accessed through standardised interfaces and access rules
   b) by well-documented and easily discoverable text mining services that process, analyse, and annotate text
   c) to identify patterns and extract new meaningful actionable knowledge, which will be used

d) for structuring, indexing, and searching content and, in tandem,

e) acting as new knowledge useful to draw new relations between content items and firing a new mining cycle.

To achieve its goals, OpenMinTeD brings together different stakeholders, content providers and scientific communities, text mining and infrastructure builders, legal experts, data and computing centres, industrial players, and SMEs.

Phenomenal

In the coming decade a significant number of the 500.000.000 European (EU/EEA) citizens will have their genome determined routinely. This will be complemented with much cheaper (currently ~20 Euro per measurement) acquisition of the metabolome of bio-fluids (e.g. urine, saliva, blood plasma) which will link the genotype with metabolome data that captures the highly dynamic phenome and exposome of patients. Having such low cost solutions will enable, for the first time, the development of a truly personalised and evidence-based medicine founded on hard scientific measurements. The exposome includes the metabolic information resulting from all the external influences on the human organism such as age, behavioural factors like exercise and nutrition or other environmental factors.

Considering that the amount of data generated by molecular phenotyping exceeds the data volume of personal genomes by at least an order of magnitude, the collection of such information will pose dramatic demands on biomedical data management and compute capabilities in Europe. For example, a single typical National Phenome Centre, managing only around 100,000 human samples per year, can generate more than 2 Petabytes of data during this period alone. A scale-up to sizable portions of the European population over time will require data analysis services capable to work on exabyte-scale amounts of biomedical phenotyping data, for which no viable solution exists at the moment.

The PhenoMeNal project will develop and deploy an integrated, secure, permanent, on-demand service-driven, privacy compliant and sustainable e-infrastructure for the processing, analysis and information-mining of the massive amount of medical molecular phenotyping and genotyping data that will be generated by metabolomics applications now entering research and clinic.

POP

High performance Computing is becoming a fundamental tool for the progress of science and engineering and as such for economic competitiveness. The growing complexity of parallel computers is leading to a situation where code owners and users are not aware of the detailed issues affecting the performance of their applications. The result is often an inefficient use of the infrastructures. Even when the need to get further performance and efficiency is perceived, code developers may not have sufficient insight on its detailed causes for addressing the problem properly. This may lead to blind attempts to restructure codes and consequent lack of efficiency.

The objective of POP is to operate a Centre of Excellence in Computing Applications in the area of Performance Optimisation and Productivity. POP will offer the service of precisely assessing the

performance of computing application of any sort, from a few hundred to many thousand processors. Also, POP will show its Customers the issues affecting the performance of their code and the most optimal way to alleviate them.

POP will target code owners and users from all domains, including infrastructure operators, academic and industrial users. The estimated population of such applications in Europe is 1500 and within the project lifetime POP has the ambition of serving over 150 such codes. The Added Value of POP's services is the savings generated in the operation and use of a code, which will result in a significant Return on Investment (fixing a code costs less than running it below its optimal levels) by employing best-in-class services and release capacity for resolving other priority issues.

POP will be a best-in-class centre. By bringing together the European world-class expertise in the area and combining excellent academic resources with a practical, hand-on approach, it will improve the access to computing applications, thus allowing European researchers and industry to be more competitive.

PRACE

PRACE, the Partnership for Advanced Computing, was established in May 2010 as a permanent pan-European High Performance Computing service providing world-class systems for world-class science. Six systems at the highest performance level (Tier-0) are deployed by Germany, France, Italy and Spain providing researchers with over 9 billion core hours of compute time.

HPC experts from twenty-five member states - funded in part in three implementation projects - enabled users from academia and industry to ascertain leadership and remain competitive in the Global Race.

Currently PRACE is preparing for PRACE 2.0, the successor of the initial five year period. The objectives of PRACE-4IP are to build on and seamlessly continue the successes of PRACE and start new innovative and collaborative activities proposed by the consortium. These include: assisting the transition to PRACE 2.0; strengthening the internationally recognised PRACE brand; continuing advanced training which so far provided more than 15.000 person-training days to over 4700 persons, preparing strategies and best practices towards exascale computing, coordinating and enhancing the operation of the multi-tier HPC systems and services, and supporting users to exploit massively parallel systems and novel architectures.

The proven project structure will be used to achieve each of the objectives in six dedicated work packages. The project will continue to be managed by Jülich. The activities are designed to increase Europe's research and innovation potential especially through: seamless and efficient Tier-0 services and a pan-European HPC ecosystem including national capabilities; promoting take-up by industry and special offers to SMEs; analysing new flexible business models for PRACE 2.0; proposing strategies for deployment of leadership systems; collaborating with the ETP4HPC, the coming CoEs and other European and international organisations on future architectures, training, application support and policies.

RDA

The Research Data Alliance (RDA) is rapidly building the social and technical bridges that enable open sharing and re-use of data on a global level. RDA/Europe - the European plug-in to the global Research Data Alliance activities – will ensure that European research, industrial, e-infrastructure and policy stakeholders are all aware of its achievements, engaged with and actively involved in the global RDA activities and will advance the use of its results.

European domain scientists and data scientists are involved in and driving a series of working and interest groups, generating first results of RDA and this momentum should be maintained and even increased. A collaborative web platform, currently under strong development, is widely used by the RDA members and coordinating administration.

RDA Europe will look for innovative ways of facilitating the take-up and implementation of RDA outputs in Europe, both in academia and industry, and will continue to develop and maintain the RDA web-based collaboration framework. Further steps will be taken in steering, planning and implementing the sustainability of RDA through a legal entity, established in Europe, and investigating the most appropriate business models for long term sustainability. This RDA/Europe proposal will mainly build on the successful consortium of the previous years but strengthen its anchoring at policy, scientific and industrial uptake to drive forward RDA Europe objectives that are all grounded in technological excellence and innovation and have clear international ties. RDA Europe partners are very actively involved in the strategic planning, implementation, marketing & communication of RDA through the governance bodies of RDA.

READ

The overall objective of READ is to implement a Virtual Research Environment where archivists, humanities scholars, computer scientists and volunteers are collaborating with the ultimate goal of boosting research, innovation, development and usage of cutting edge technology for the automated recognition, transcription, indexing and enrichment of handwritten archival documents.

This Virtual Research Environment will not be built from the ground up, but will benefit from research, tools, data and resources generated in multiple national and EU funded research and development projects and provide a basis for sustaining the network and the technology in the future. This ICT based e-infrastructure will address the Societal Challenge mentioned in Europe in a Changing World namely the "transmission of European cultural heritage" and the "uses of the past" as one of the core requirements of a reflective society.

Based on research and innovation enabled by the READ Virtual Research Environment we will be able to explore and access hundreds of kilometres of archival documents via full-text search and therefore be able to open up one of the last hidden treasures of Europe's rich cultural heritage.

SciGaia

In African Communities of Practice (CoPs), international collaboration and the pursuit of scientific endeavour has faced a major barrier with the lack of access to e-Infrastructures and high performance network infrastructure enjoyed by European counterparts.

With AfricaConnect, the proposed AfricaConnect2 and regional developments, this situation is changing rapidly. In the project Teaming-up for exploiting e-Infrastructures' potential to boost RTDI in Africa (eI4Africa) it has been demonstrated clearly that it is possible to develop e-Infrastructures in Africa. It has also been demonstrated clearly that, as with the rest of the world, easy to use web portals, or Science Gateways, are needed to help CoPs to easily access e-Infrastructure facilities and through these collaborate with CoPs across the world.

However, a major problem exists: it is very difficult for non-experts to develop Science Gateways and supporting e-Infrastructures. Elements of guides and supporting materials exist but these are either written for different audiences or out of date. This present Coordination and Support Action, called Energising Scientific Endeavour through Science Gateways and e-Infrastructures in Africa (Sci-GaIA), therefore proposes to bring together these materials into clearly structured guides and educational documents that can be used to train and support representatives of NRENs, CoPs and, importantly, Universities to develop Science Gateways and other e-Infrastructures services in Africa.

Sci-GaIA plans to work with new and emerging CoPs to develop these exciting technologies, to strengthen e-Infrastructure service provision, especially in terms of open access linked data, and to deliver training and dissemination workshops. This will give a sustainable foundation on which African e-Infrastructures can be developed and be linked to scientific networks across Africa. Importantly, the results of our project will be usable by CoPs in Europe and the rest of the world.

SESAMENET

The project will contribute to the implementation of the European HPC strategy, in particular to foster the use of HPC by SMEs. By bringing together the proposed partnership it will be possible to address all the activities outlined in the call, i.e.:

1. A network of existing HPC Competence Centres providing HPC services to exchange best practices and pool technical, expertise or business resources
2. Awareness raising and visibility of the benefits of HPC for SMEs, in particular for countries that do not currently have such centres
3. Identification of the pool of SMEs and available expertise in the different business areas at European level, and mechanisms to match SME needs and the available expertise
4. Training (in synergy with the activities carried out by other organisations providing specific training for SMEs in HPC).

TANDEM

The TANDEM project aims to create favourable conditions for WACREN (West and Central African Research and Education Network) to be able to participate in the forthcoming AfricaConnect project extension and to draw maximum benefit from it, and thereby to ensure WACREN's integration into the global Research and Education networking community.

TANDEM long-term goal is to make it possible for researchers and academics to contribute with their peers around the world to the socio-economic development of the West and Central African Region. To this end, TANDEM is structured around three objectives:

1. Enhance the dialogue between WACREN and policy makers, international donors and regulating authorities, in order to promote connectivity and obtain political and financial support.
2. Enhance the dialogue between WACREN, end users, national research and education networks (NRENs) and research and tertiary education institutes, in order to identify the global e-infrastructure services expected by the end users.
3. Build a permanent stakeholder dialogue by creating favourable conditions for the implementation of three main communities: the "policy makers, donors and regulating authorities" community; the NRENs community; the "end users" community.

TANDEM provides for the implementation of a virtual collaboration platform in order to support the continuity of communities beyond the end of the project, and to enable monitoring of the e-infrastructure services. In order to respond to the growing demand from African research communities, TANDEM is focused on three key research thematic topics: health, food security, and environment; and on the specific needs of these areas.

The consortium was carefully selected to provide a balanced mix of research and e-infrastructure organisations, with diverse experiences in academic issues, research and innovation management. The consortium also brings consolidated experience in implementing and managing large research and education networks.

THOR

Five years ago, a global infrastructure to uniquely attribute to researchers their scientific artefacts (articles, data, software…) appeared technically and socially infeasible. Since then, DataCite has minted over 3.5m unique identifiers for data. ORCID has deployed an open solution for identification of contributors with over 850,000 registrants in less than 2 years.

THOR will leverage these emerging global infrastructures to support the H2020 goal to make every researcher 'digital' and increase creativity and efficiency of research, while bridging the R&D divide between developed and less-developed regions. We will establish interoperability between existing resources, linking digital identifiers across platforms and propagating attribution information. We will integrate PID services across the research lifecycle and data publishing workflows in four advanced research communities, and then roll-out core services and service building blocks for the wider community. These open resources will foster an open and sustainable e-infrastructure across stakeholders to avoid duplications, give economies of scale, richness of services and the ability to respond rapidly to opportunities for innovation.

THOR is not just relevant to the EINFRA-7-1024 Call, but will become a pervasive element of the EINFRA family of e- Infrastructure resources over the next 3 years. It will allow data-management and curation services to exploit knowledge of data location and attribution; provide robust and persistent mechanism for linking literature and data; enable search and resolving services and generate incentives for Open Science; deliver provenance and attribution mechanisms to underpin data exchange; and provide minting and resolving services for data citation workflows. Its impact will enable third-party services, no-profit and commercial, to leverage the scholarly record.

VI-SEEM

In the last decade, a number of initiatives were crucial for enabling high-quality research - by providing e-Infrastructure resources, application support and training - in both South East Europe (SEE) and Eastern Mediterranean (EM). They helped reduce the digital divide and brain drain in Europe, by ensuring access to regional e-Infrastructures to new member states, states on path to ascension, and states in European Neighbourhood Policy area – in total 14 countries in SEE and 6 in EM. This VI-SEEM proposal brings together these e-Infrastructures to build capacity and better utilize synergies, for an improved service provision within a unified Virtual Research Environment (VRE) for the inter-disciplinary scientific user communities in the combined SEE and EM regions (SEEM).

The overall objective is to provide user-friendly integrated e-Infrastructure platform for regional cross-border Scientific Communities in Climatology, Life Sciences, and Cultural Heritage for the SEEM region; by linking compute, data, and visualization resources, as well as services, models, software and tools. This VRE will provide the scientists and researchers with the support in full lifecycle of collaborative research: accessing and sharing relevant research data, using it with provided codes and tools to carry out new experiments and simulations on large-scale e-Infrastructures, and producing new knowledge and data - which can be stored and shared in the same VRE. Climatology and Life Science communities are directly relevant for Societal Challenges.

The driving ambition of this proposal is to maintain leadership in enabling e-Infrastructure based research and innovation in the region for the 3 strategic regional user communities: supporting multidisciplinary solutions, advancing their research, and bridging the development gap with the rest of Europe.

The VI-SEEM consortium brings together e-Infrastructure operators and Scientific Communities in a common endeavour.

VRE4EIC

The VRE4EIC project addresses key data and software challenges in supporting multidisciplinary data driven sciences. These include:

1) understanding complex user requirements across domains by closely involving committed user communities;
2) improving the quality of VRE user experience by providing user centred, secure, privacy compliant, sustainable environments for accessing data, composing workflows and tracking data publications;
3) increasing VRE usage in multidisciplinary research domains by abstracting and reusing building blocks and workflows from existing VRE initiatives;
4) improving the interoperability of heterogeneous discovery, contextual and detailed metadata across all layers of the VRE; and
5) promoting the exploitation of VRE4EIC solutions to different research communities and commercially.

The project contributes to the Work Programme through innovations in the following areas:

- support of excellent research through improved VREs and interoperation of heterogeneous VREs leading to new multidisciplinary science;
- increased user acceptance through support for trust, security and privacy throughout the VRE architecture;
- a novel VRE architecture, prototypes and enhanced metadata that can be used for future multidisciplinary VREs;
- a semantic web, linked open data view of VRE information allowing cross-linking to open government data enabling heterogeneous discovery;
- interoperable standard software services retro-fitted to enhance existing VREs;
- solutions for societal challenges demonstrated in the environmental and earth science domains (environmental pollution, climate change, earthquakes), building on 25 real use cases;
- increased VRE adoption to other domains and research communities through a training programme and learning environment to empower researchers to utilise the full potential of VRE4EIC and to enhance collaboration.

VRE4EIC covers all EU member states and EFTA countries, and will affect 70,000 researchers all over Europe.

West-Life

The focus of structural biology is shifting from single macromolecules produced by simpler prokaryotic organisms, to the macromolecular machinery of higher organisms, including systems of central relevance for human health. Structural biologists are expert in one or more techniques. They now often need to use complementary techniques in which they are less expert. Instruct supports them in using multiple experimental techniques, and visiting multiple experimental facilities, within a single project.

The Protein Data Bank is a public repository for the final structure. Journals require deposition as a precondition of publication. However, metadata is often incomplete.

This project will pilot an infrastructure for storing and processing data that supports the growing use of combined techniques.

There are some technique-specific pipelines for data analysis and structure determination. Little is available in terms of automated pipelines to handle integrated datasets. Integrated management of structural biology data from different techniques is lacking altogether.

The proposed activity will integrate the data management facilities that already exist, and enable the provision of new ones.

The resulting integration will provide users with an overview of the experiments performed at the different research infrastructures visited, and links to the different data stores. It will extend existing facilities for processing this data. As processing is performed, it will automatically capture metadata reflecting the history of the project. The effort will use existing metadata standards, and integrate with them new domain-specific metadata terms.

This proposal will develop application level service specific to uses cases in structural biology, enabling structural biologists to get the benefit of the generic services developed by EUDAT and the EGI.