E-INFRASTRUCTURES: MAKING EUROPE THE BEST PLACE FOR RESEARCH AND INNOVATION.
Today the European Union (EU) has 1.7 million researchers and 70 million science and technology professionals engaged in the creation of new knowledge, products, services and processes. Underpinning their efforts, digital infrastructures foster innovation and scientific progress across disciplines and between the private and public sectors. Digital infrastructures are the virtual backbone of European research and a vital driver for innovation. Without them there would be no sharing of data, no exchange of know-how and no collaboration.

EU-funded e-infrastructures address the needs of European researchers and innovators for transnational communication networks, high performance and high throughput computing, multidisciplinary data management and collaborative scientific software. They empower researchers with easy and secure online access to facilities and resources and enable them to deliver reusable and reproducible research and innovation outputs. During the period 2014-2020, the EU will have invested more than 850 million euros in digital infrastructures through its Horizon 2020 e-infrastructure programme.

But e-infrastructures are set to play a much larger role in society.

The European Commission adopted on 19 April 2016 the communication “European Cloud Initiative – Building a competitive data and knowledge economy in Europe”[1] that sets out a strategy to strengthen Europe’s position in data-driven innovation, improve its cohesion and help create a Digital Single Market[2] in Europe. This is a fundamental step towards the reinforcement of the EU’s competitiveness in digital technologies and in innovation. The European Cloud Initiative will provide European science, industry and public authorities with world-class data infrastructures, high-speed connectivity and increasingly powerful high performance computers. It will make it easier for researchers, businesses and public services to fully exploit the benefits of big data by making it possible to move, share, re-use and process data seamlessly across global markets and borders, and among institutions and research disciplines.

The initiative will establish the European Open Science Cloud (EOSC)[3] as a virtual environment to store and process large volumes of information generated by the big data revolution. This will be supported by the European Data Infrastructure (EDI)[4], deploying the high-bandwidth networks and the supercomputing capacity necessary to access and process large datasets stored in the cloud.

Both the EOSC and the EDI will build on existing EU-funded e-infrastructures and will bring networking, data and computing services closer to European researchers and innovators.

Today, EU-funded e-infrastructures play a fundamental role in the life of European researchers and innovators. This booklet illustrates that by showing how e-Infrastructure projects funded under the Horizon 2020 work programme 2014-2015 are helping researchers tackle the challenges posed by one specific societal challenge: climate change.

We believe that after having read the whole story, you will not only have a better understanding of how European e-infrastructures help researchers and innovators do their work, but you will also see how they will contribute to making the EOSC and the EDI a success.

If you do research or innovate in Europe, you are in the right place! Happy reading!
Table of content

1. Europe 2016: Bringing climate change research to the next level ..........2

   Seamless access & connectivity..........................3
   Storing data in a secure and trusted environment ..................................4
   Equipping researchers with the computational resources they need .........6
   Collaborative research enhances results ........................................8
   Providing access control of results ........................................11
   Open access as a key for innovation and for boosting researchers’ careers 11

2. In closing .....................................................14

3. Editorial Board .............................................16

4. e-infrastructure projects WP 14-15 ....17
Global climate change has had noticeable effects on the environment. The rise of global temperatures is impacting plant growth stages, affecting migration patterns, threatening the survival of the species and affecting water quality. Many effects predicted by scientists are now commonplace: loss of sea ice, accelerated sea level rise and longer, more intense heat waves. We need to take immediate action to preserve the planet.

European researchers can play a big role in monitoring, predicting and assessing the impact of climate change.
Seamless access & connectivity

John works in a university in Ireland as a Research Associate focussing on the impacts of climate change on the environment and how to mitigate them. He is a member of the European climate change community. For John, access to Earth Observation data is fundamental.

Through GÉANT, Europe’s leading collaboration on e-infrastructure and services for research and education, John can use his university’s network to connect to data facilities around the world. Using the Virtual Private Network services operated by GÉANT and the National Research and Education Networks (NRENs), he can securely transfer extremely large files.

GÉANT’s eduGAIN service allows him to use his campus identity to access data and services using one secure password, making it easier to work and collaborate with international projects. And when he visits other universities, John can use his eduroam® account to connect to the local campus Wi-Fi network.

In February 2016 GÉANT and the European Space Agency (ESA) finalised an agreement to distribute data from Copernicus, the EU earth observation and monitoring programme, to research and education users worldwide.

Thanks to this agreement, John is now able to access a vast amount of global data from satellites and other systems and can transfer extremely large files such as satellite images over large distances in real time.

Did you know?
Copernicus services will be enhanced by EarthServer2. The vision of the EarthServer2 project is to offer researchers ‘Big Earth data at your fingertips’ so that they can access and manipulate enormous data sets with just a few mouse clicks. EarthServer2 will provide services to analyse multi-petabyte scale datasets spread over multiple data providers.

Storing data in a secure and trusted environment

John’s home institution does not have sufficient storage resources to host the Copernicus data, but his research community is federated to EUDAT, the Collaborative Data Infrastructure. Through EUDAT, John is able to access a series of integrated research data management services via a reliable collaborative data infrastructure that brings together 35 European data and HPC centres.

Given the huge amount of data that John is now able to access and use, John needs a vast amount of computing resources to run his earth observation model.

Keep an eye on LEARN

Researchers are the generators and owners of significant amounts of research data; they require support in planning and management at all stages of the research project life-cycle.

The LEARN project will deliver a model for Research Data Management policy, a best practice Toolkit and an Executive Briefing.
Equipping researchers with the computational resources they need

John can run thousands of compute-intensive tasks for producing and analysing large datasets through the high-throughput data analysis service provided by EGI-ENGAGE. If he looks for flexibility, John can deploy his computing platform using the EGI cloud services.

In addition, John can access the high-performance computer systems provided by PRACE, the Partnership for Advanced Computing in Europe. Scientists and researchers from around the world can access PRACE resources free of charge after successful peer review.

Computer simulations speed up the process of acquiring knowledge by shortening or even replacing complex, time-consuming laboratory experiments.

Fernando has recently joined John’s team as a postdoc funded through a Marie Curie fellowship. He advises John to use INDIGO - DataCloud platform solutions to develop integrated applications combining resources from GÉANT, EGI, EUDAT and PRACE, as well as commercial cloud providers if and as needed.

INDIGO - DataCloud

INDIGO - DataCloud develops an Open Source data and computing platform provisioned over private, public, or hybrid e-infrastructures. By filling gaps of current Cloud technologies, INDIGO - DataCloud helps scientists, software developers, resource providers and e-infrastructures to efficiently exploit computing, data and network technologies: Better Software for Better Science.

Is computing only for the research sector?

SESAME Net establishes a network of HPC competency centres for SMEs. The aim is to build a pan-European network involving HPC service providers and SMEs. The network facilitates the exchange of knowledge as well as the provision of technical support and HPC infrastructure for SMEs.
Collaborative research enhances results

To improve the accuracy of his results, John must combine his predictive models with ways to assess risk to near-coastal species and habitats. This is exactly what Juliette, a French biologist, does at her research institute.

John met Juliette when both attended a Research Data Alliance (RDA) Geospatial Interest Group meeting. The RDA working and interest groups are comprised of experts from the community that are committed to directly or indirectly enabling data sharing, exchange, or interoperability. During a coffee break, Juliette told John how Virtual Research Environments (VREs) have improved her daily work and her collaboration with Sven, the CEO of a small Finnish aquaculture enterprise. Sven wanted to expand his aquafarming business to the north of Finland; but to identify a good location for his farm he first needed to understand the environmental impact on the area.

Juliette started analysing the scientific literature by browsing the registry of the OpenMinTeD platform and using its text mining services and workflows. She parsed the publications and extracted the ones indicating a relation between aquafarming and the environment.

At that point, Juliette asked Sven to join the BlueBRIDGE VRE she set up to run her environmental model on the D4Science infrastructure. Thanks to BlueBRIDGE’s statistical manager service, offering over 100 algorithms and the ability to integrate external algorithms, Juliette was able to provide Sven with an answer to his problem in a couple of days.

John was so impressed by this success story that, together with his colleague Anja, a computer scientist in Latvia, he started using the VREs provided by the EVER-EST project, specifically tailored to the needs of the earth science domain.

Being both aware of the advantages of VREs, John and Juliette decide to set up a new VRE to combine their models and start collaborating on this new challenge by sharing data and models and accessing additional datasets.

VREs are powerful instruments to answer questions that are not restricted to a specific sector (e.g., government & intergovernmental organizations, public authorities, industry, SMEs, universities and research institutes) or discipline, but span across them.

The current global research data landscape is highly fragmented, either by disciplines or by domains. The Research Data Alliance (RDA) is an international organization focused on the development of infrastructure and community activities, recommendations designed to reduce barriers to data sharing, and the acceleration of data-driven innovation worldwide.

RDA builds social and technical bridges to achieve its vision of researchers and innovators openly sharing data across technologies, disciplines and countries to address the grand challenges of society.

OpenMinTeD

The project initiates an infrastructural approach to open up research outputs for text and data mining, to foster knowledge discovery, and advance research and innovation within the Open Science ecosystem. OpenMinTeD provides an interoperability layer and services to enable uniform access to openly available research literature and related content, as well as discovery, deployment, and use of interoperable text and data mining resources, tools, services and workflows.
VREs for biology

West-Life (World-wide E-infrastructure for structural biology) brings the world of complex data analysis in structural biology to a simple, web-browser based VRE. West-Life provides a framework for storing and processing data for any laboratory involved in the experimental structural characterization of biomolecules and their chemical make-up.

VREs for blue growth

BlueBridge (Building Research environments fostering Innovation, Decision making, Governance and Education for Blue growth) delivers tailor made data management services to different communities (aquaculture, ecosystem approach to fisheries and research sector) and stakeholders (International organisations, research centres, enterprises, etc.). The services are operated through VREs built on top of an hybrid-data infrastructure.

VREs for earth science

EVER-EST will develop a generic VRE tailored to the needs of the Earth Science domain. It will create an innovative framework which enhances the ability of the Earth Science communities and, observational scientific disciplines in general, to interoperate and share knowledge and expertise. The EVER-EST VRE will incorporate selected innovative and state-of-the-art technologies, systems and tools developed by previous relevant FP7 projects.

VREs for genomics

MuG (Multi-scale complex Genomics) responds to the latest computational challenges of 3D/4D genomics by bringing this community closer to the HPC world and providing a suitable set of tools and infrastructure to integrate the navigation in genomics data from sequence to 3D/4D chromatin dynamics data.

VREs for mathematics

OpenDreamKit (Open Digital Research Environment Toolkit for the Advancement of Mathematics) aims to deliver a flexible toolkit to set up VREs tailored to the needs of mathematics and supporting the full research life cycle from exploration, through proof and publication, to archiving and sharing of data and code. It also contributes to the development of a collection of easy to deploy packages, containers or virtual machines for the major open source computational mathematics software systems.

VREs for archival documents

READ (Recognition and Enrichment of Archival Documents) implements a VRE 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.

VREs for ad hoc research communities

VRE4EIC (A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration) provides an architectural model for a VRE implemented on two large ESFRI e-Research Infrastructures namely EPOS and ENVRI+ and a set of generic software services, to be used to build new or improve existing VREs.

VREs for blue growth

VREs for earth science

VREs for specific geographic areas

VI-SEEM (VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean) provides user-friendly integrated e-infrastructure platform for scientific communities in climatology, life sciences, and cultural heritage for the South East Europe and Eastern Mediterranean region; by linking compute, data, and visualization resources, as well as services, software and tools.

In focus - What is a VRE?

A Virtual Research Environment is a web-based environment that can be created on-demand through a simple user interface. It provides researchers and research teams, educators, SMEs, and any other type of user, from different disciplines, institutions or even countries, with controlled access to collaborative tools, services, data and computational facilities meeting their specific needs. Hardware setup and software deployment required to operate these facilities are completely transparent to the VRE creator.
John and Juliette’s experiments generated new knowledge essential to make progress on their climate change research and on which their colleagues can build upon. John stored some of the results on his organisation’s in-house resources and made the results directly available for Juliette and her colleagues through the single identity services provided by the Authentication and Authorisation for Research and Collaboration project (AARC).

Providing access control of results

John and Juliette are now ready to share the results of their joint research with the external world by writing an article and publishing it in an open access journal. On submission, Juliette provides her ORCID ID - a permanent identifier that distinguishes her from every other researcher. The journal encourages them to cite their sources of data with a DataCite DOI - a permanent identifier that connects their article directly to the data. Juliette’s and John’s ORCID record will be updated every time they publish and obtain a DOI. It helps them to maintain their CV, receive appropriate recognition and even apply for funding.

To follow the latest open science trends and to meet the funder’s Open Access mandates, John and Juliette take the necessary steps to ensure a proper and permanent storage, citation and linking of all their results so that others can replicate and validate their research. For that they:
- Deposit the publication in their institutional repository, where it is discoverable by OpenAIRE, automatically connected to its funding, and reported to the funder information system;
- Identify via OpenAIRE and re3data.org the most fitting repository to deposit, curate and get help on how to link the accompanying research data;
- Place other supplementary data or a snapshot of their models from Github to Zenodo, OpenAIRE’s catch-all repository hosted by CERN.

Because today’s research is increasingly data-driven, the key skills needed for the use and management of e-infrastructures are changing as well. E-skills are becoming crucial for researchers and scientists that must widen their knowledge to take full advantage of the current data revolution.

Open access as a key for innovation and for boosting researchers’ careers

The story of John, Juliette, Sven and Anja is just one example of what the European services and digital infrastructures developed by the different e-infrastructure projects funded under the Horizon 2020 Work Programme 2014-2015 can do to effectively and concretely support researchers.

Skills & Policy

The EDISON project has been established to support universities, research centres, and industry and research infrastructures in coping with the potential shortfall of data scientists and to define the framework of competences as well as the body of knowledge for this profession.

The e-infrastructure Reflection Group (e-IRG), supported by the e-IRGSP4 project, produces white papers describing the state-of-the-art and future developments in e-infrastructures, guideline documents to guide specific research communities about how to set-up and manage e-infrastructure services and roadmaps forecasting (desired) developments in e-infrastructures.

PIDs matter?

The THOR Project implements persistent and interoperable identifiers (PIDs) to researchers and data that enable information exchange between every European ICT tool, facilitate reliable publication workflows and ensure that every researcher receives credit for his or her work.

OpenAIRE

OpenAIRE is the European network of Open Access repositories, journals and CRIS’s. It provides a public support helpdesk in 33 countries, with experts helping out researchers and policy makers in making the appropriate OA choice. It currently provides access to 15 million publications and 137,000 datasets from 710 validated data sources. Monitoring services are continuously identifying research outputs and trends, currently about 250,000 publications and 3,500 datasets linked to 10 funders.

OpenAIRE is the European network of Open Access repositories, journals and CRIS’s. It provides a public support helpdesk in 33 countries, with experts helping out researchers and policy makers in making the appropriate OA choice. It currently provides access to 15 million publications and 137,000 datasets from 710 validated data sources. Monitoring services are continuously identifying research outputs and trends, currently about 250,000 publications and 3,500 datasets linked to 10 funders.

OpenAIRE is the European network of Open Access repositories, journals and CRIS’s. It provides a public support helpdesk in 33 countries, with experts helping out researchers and policy makers in making the appropriate OA choice. It currently provides access to 15 million publications and 137,000 datasets from 710 validated data sources. Monitoring services are continuously identifying research outputs and trends, currently about 250,000 publications and 3,500 datasets linked to 10 funders.

OpenAIRE is the European network of Open Access repositories, journals and CRIS’s. It provides a public support helpdesk in 33 countries, with experts helping out researchers and policy makers in making the appropriate OA choice. It currently provides access to 15 million publications and 137,000 datasets from 710 validated data sources. Monitoring services are continuously identifying research outputs and trends, currently about 250,000 publications and 3,500 datasets linked to 10 funders.

OpenAIRE is the European network of Open Access repositories, journals and CRIS’s. It provides a public support helpdesk in 33 countries, with experts helping out researchers and policy makers in making the appropriate OA choice. It currently provides access to 15 million publications and 137,000 datasets from 710 validated data sources. Monitoring services are continuously identifying research outputs and trends, currently about 250,000 publications and 3,500 datasets linked to 10 funders.

OpenAIRE is the European network of Open Access repositories, journals and CRIS’s. It provides a public support helpdesk in 33 countries, with experts helping out researchers and policy makers in making the appropriate OA choice. It currently provides access to 15 million publications and 137,000 datasets from 710 validated data sources. Monitoring services are continuously identifying research outputs and trends, currently about 250,000 publications and 3,500 datasets linked to 10 funders.

OpenAIRE is the European network of Open Access repositories, journals and CRIS’s. It provides a public support helpdesk in 33 countries, with experts helping out researchers and policy makers in making the appropriate OA choice. It currently provides access to 15 million publications and 137,000 datasets from 710 validated data sources. Monitoring services are continuously identifying research outputs and trends, currently about 250,000 publications and 3,500 datasets linked to 10 funders.
John, Juliette, Sven and Anja illustrate the daily life of the researchers around us. They have shown us how important e-infrastructures are for them. But their needs are evolving: data-driven science is changing the way they do research, and e-infrastructures must adapt to this revolution.

Big data needs cloud, bandwidth and powerful computers to store, manage, transport, share, re-use and process data across markets, borders, institutions and research disciplines.

We will all have to ensure that the e-infrastructures that researchers and innovators will use in the coming years are operational, secure and accessible. We will also have to support researchers and innovators with training, allowing them to exploit the full potential of e-infrastructures.

Through the European Cloud Initiative and its ambitious European Open Science Cloud (EOSC) and European Data Infrastructure (EDI) plans, the European Commission has proposed a unifying vision with the potential to align regional, national and European efforts in the area of e-infrastructures for the benefit of the European researchers and innovators. But for the EOSC and the EDI to be a success, all actors must play their role.

Funders must put in place adequate governance mechanisms; e-infrastructures providers and operators must integrate and consolidate their services and platforms and foster their innovation efforts. Finally, users, both researchers and innovators, must engage with e-infrastructures providers to co-design the services of the future.

As this booklet has shown, much good work has already been done, but there is more to do. The stakes are high. It is now up to us to meet the challenge.
Acknowledgements

We wish to thank Sara Garavelli for her fine editing skills, Angela Dappert for her ability to graphically represent complex concepts, Tony Ross-Hellauer and Sara Coelho for valuable contribution to the concept, Riccardo Giuntini for decorating our words, Stéphanie Matt for her ineffable role, and all the EU-funded e-infrastructure projects and their respective Project Officers at the European Commission.
e-infrastructure projects WP 14-15

AARC: Authentication and Authorisation for Research and Collaboration

BlueBridge: Building Research environments fostering Innovation, Decision making, Governance and Education to support blue growth

EarthServer: BigEarth Data at your fingertips

EDISON: Building the data science profession

OpenAIRE: Science set free

OpenDreamKit: Open Digital Research Environment Toolkit for the Advancement of Mathematics

OpenMined: Open Mining Infrastructure for Text and Data

PhenoMeNal: Phenome and Metabolome analysis

www.aarc-project.eu
www.blueridge-vres.eu
www.earthserver.eu
www.edison-project.eu
www.openaire.eu
www.opendreamkit.org
www.openminted.eu
www.phenomenal-h2020.eu

www.e-irg.eu
www.eudat.eu
www.everest.eu
www.prace-ri.eu
www.rd-alliance.org
www.read.transkribus.eu
www.sesamenetwork.eu

Indigo: DataCloud - Integrating Distributed data Infrastructures for Global Exploitation

LEARN: Leaders Activating Research Networks

MuG: Multi-scale complex Genomics

THOR: Technical and Human infrastructure for Open Research

VI-SEEM: Virtual Research Environment (VRE) for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean

VRE4EIC: A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration

www.geant.org/geantproject
www.indigo-datacloud.eu
www.learn-rdm.eu
www.multiscalegenomics.eu
www.indigo-datacloud.eu
www.learn-rdm.eu
www.multiscalegenomics.eu

GÉANT Project (GN4-1): Accelerating research, driving innovation and enriching education

INDIGO - DataCloud

LEARN - Leaders Activating Research Networks

MuG - Multi-scale complex Genomics

THOR - Technical and Human infrastructure for Open Research

VI-SEEM - Virtual Research Environment (VRE) for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean

VRE4EIC - A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration

www.geant.org/geantproject
www.indigo-datacloud.eu
www.learn-rdm.eu
www.multiscalegenomics.eu
www.indigo-datacloud.eu
www.learn-rdm.eu
www.multiscalegenomics.eu

West-Life: World-wide E-infrastructure for structural biology

www.multiscalegenomics.eu

04