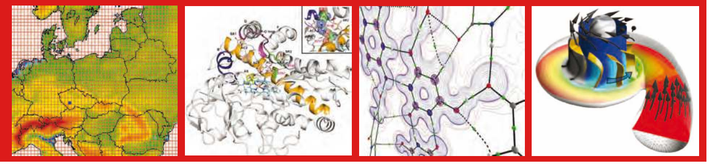


cerit scientific cloud

Research&Development

Research and development activities focus on two areas:

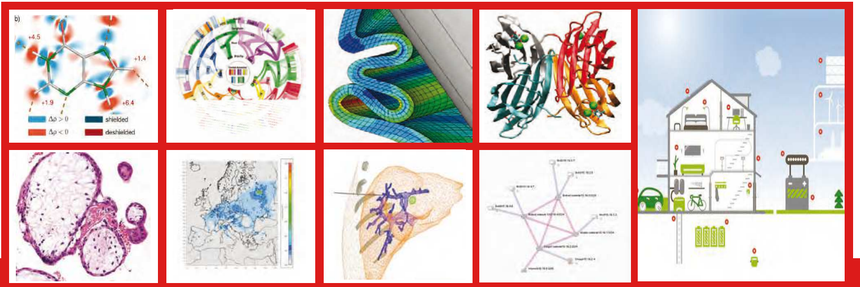
- Interdisciplinary research collaborations in many research areas – CERIT-SC users do not act only as service consumers, but mainly as partners in research and production application of its results. CERIT-SC is usually responsible for the e-infrastructure, data processing and computation optimization related part of the research, thus allowing researchers and their teams to focus on and progress more quickly with their own research.
- Computer science e-Infrastructure research – focuses on the most important aspects of optimized large scale data processing and analytics, efficient and fast computations including use of manycore processors, and e-infrastructure development and optimization driven by the requirements of its flexible advanced use.

Research Infrastructure CERIT-SC tightly cooperates with the Faculty of Informatics as well as other parts of the Masaryk University and other universities. It targets specifically into integrating master and graduate students. PhD study consultants work in the centre, helping students to integrate tightly into relevant research areas, their results immediately applied in the centre itself as well as other places.

Interdisciplinary Collaborations

In Life Sciences data processing and computational simulations are done extensively at various scales, from nano to macro, covering the following areas: **molecular simulations** (drug design), **structural biology, bioinformatics and sequencing data** (understanding genome of living organisms), **processing microscopic images**, **tissue simulations** (surgery planning etc.), **neurological models** (function of brain), **biobanks** (BBMRI) and related **data mining** and many more.

CERIT-SC is an official partner of the **Czech national Node of the ELIXIR infrastructure**. CERIT-SC together with CESNET are responsible for developing, building, and operating the computational and storage infrastructure. As technology oriented partners, we provide expertise in the infrastructure areas. User authentication and authorization are probably the most visible outcomes of this cooperation.



Energy: We proposed and developed a **modular platform processing detailed weather forecast** information to predict production of photovoltaic and wind power plants. This information is used to control the power accordingly, minimizing negative impact of weather changes.

Large scale simulations for power grids: Wide introduction of smartmeters and accompanying technology requires large simulations to evaluate its impact on the CEZ power grid infrastructure. We used our testbed to run a simulation of 3.5 million smartmeters, evaluating impact of various behaviour patterns of non-reliable data transfers.

Cyber Security: Facing the threats posed by contemporary attackers requires to develop **techniques and methods for detection and prevention of attacks** continually. The CERIT-SC team collaborates on the development of a testbed which provides a generic way to simulate and study a wide range of cyber attacks.

Transport: We solved the problem of evaluating the **vulnerability of road network** by simulating impacts of simultaneous closure of multiple roads/links.

Earth observation sciences involve large-scale numeric **simulations of climate** (understanding the processes, important for geographical distribution of persistent organic pollutants in an environment, and time trends of their concentration), biomass growth, CO_2 cycle in environment, 3D three reconstruction from LiDAR scans, etc.

We developed **tools for transferring climate data between global models** that have been developed in different countries (including the **Czech Aladin model**).

BigData: Extensive attention is paid to extremely large datasets, which often require very simplistic techniques leveraging between analysis feasibility (e.g. response time, amount of used resources) and achievable information value. The example domains include **large networks of interconnected sensors** (present for instance within the concept of Internet of Things), cybersecurity assurance and **(cyber)crime detection** (dealing with vast amounts of heterogeneous data), as well as various **bioinformatics data portals and analyses** (e.g. genome DNA/RNA sequencing and analyses).

Computer Science e-Infrastructure Research

e-infrastructure research and development is primarily driven by the requirements of the e-infrastructure evolution and its flexible advanced use:

- **Algorithm development and optimization**, with specific emphasis on parallel and distributed computing environment and the use of accelerators (GP-GPU, Xeon Phi).
- **e-Infrastructure configuration and optimization**, including complex virtual (cloud) environments for critical information infrastructures, such as existing Cybernetic Proving Ground (research, development, and training in cybersecurity), or virtualized testbeds to support simulation of large systems (e.g., smart cities, energy distribution network etc.).
- **Work with very large datasets**—this is a recurring scheme in many RI's collaborations, ranging from security, through smart energy to life sciences)—and recently introduction of generic "big data" analysis techniques as a new research focus.
- **Job and task scheduling** (in collaboration with CESNET)
- **Security of large distributed e-infrastructures** (in collaboration with CESNET), in particular authentication and authorization in distributed environment, general cybersecurity.

The ICT technology is ubiquitous in research nowadays, virtually no research activities could be competitive without extensive use of ICT. Simultaneously, the technology grows more and more complex, becoming difficult to be used extensively by users who are experts in their scientific areas, and not computing and e-infrastructures themselves.

Research projects participations (selection)**:**

- 2012-2017 EU FP7: THALAMOSS 306201
- 2014-2017 EU FP7: SDI4Apps 621129
- 2014-2016 TACR: Advanced meteorological forecasts platform in Energy sector TA04020645
- 2015-2018 EU H2020 RI: West-Life 675858
- 2015-2016 GACR: Enhanced Sampling Simulations for Complex Systems 15-17269S
- 2017-2020 MV: Complex Analysis and Visualization of Large-scale Heterogeneous Data VI20172020096
- 2014-2016 GAMU: Advanced hybrid methods for study of transport processes in proteins and their use in the bio catalysers' design MUNI/M/1038/2013
- 2016-2017 GAMU: EDIRex (EU H2020 RI preparation phase)
- 2017–2022 BRIDGE (OP RDE) submitted
- 2018–2020 EDIREX (EU H2020) submitted
- 2017–2021 CETOCOEN (OP RDE) submitted

cerit scientific cloud

Contact

CERIT Scientific Cloud
Institute of Computer Science
Masaryk University
Botanická 68a
602 00 Brno
Czech Republic

www.cerit-sc.cz
support@cerit-sc.cz

Mission

Research Infrastructure CERIT-SC focuses on the **highly flexible computing and storage infrastructure provisioning** and use and on the related Computer Science research and development. Virtualized computing and storage resources, available through novel interfaces combining grid and cloud, represent an unique installation.

Research programme lie in **processing of large amounts of data** and **highly demanding computations**, with aim to increase the efficient usage of the CERIT-SC infrastructure and to advance the state of research in both areas.

The CERIT-SC role lies in a close long term **collaboration with end users and research teams from many research areas** (life sciences, material science, energy, transport, earth observation, cyber security, etc.) on implementation and improvement of new methods and tools.

CERIT-SC is an official partner of the Czech national Node of the ELIXIR – an infrastructure uniting Europe's leading life science organisations in managing, processing, and storing data generated in public-funded research. It also explicitly collaborates with infrastructures like BBMRI CZ or CzeCOS.

Funding

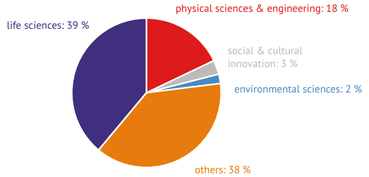
Infrastructure funding projects
- 2011–2014 CERIT Scientific Cloud (OP RDI) CZ.1.05/3.2.00/08.0144
- 2016–2019 CERIT Scientific Cloud LM2015085
- 2016–2019 ELIXIR CZ Czech National Infrastructure for Biological Data LM2015047
- 2017–2021 CERIT Scientific Cloud (OP RDE) – submitted
- 2017–2021 ELIXIR-CZ Capacity Building (OP RDE) – submitted

CERIT-SC in Numbers

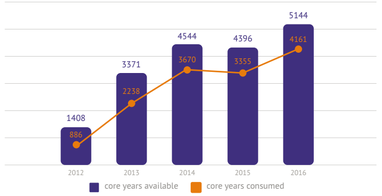
- **5,512 CPU cores** available
- **5,340 TB** for data storage
- **1,611 users** from Czech academic and research institutions and abroads
- usage **3 million CPU hours** per month

(State of January 2017)

Utilization by the discipline and real CPU time used



Number of CPU core years available and consumed



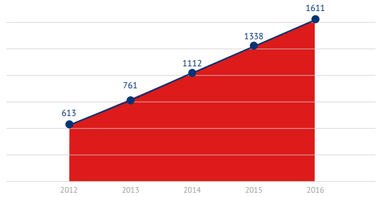
Equipment

Computing

Most of the resources are virtualized, supporting cloud interfaces (using primarily OpenNebula with OpenStack coming, but being open to experimental setup of other cloud management systems) and highly flexible access methods. The hardware is composed from computation nodes of four basic types:

- HD (High Density)—usually with two CPUs (i.e., 8-20 cores) in shared memory, offering maximal computation power/price ratio. They cover the needs of applications with limited or no internal parallelism that can make use of running many simultaneous instances (high-throughput computing).
- SMP (Symmetric MultiProcessing)—clusters with more CPUs (80 cores) in shared memory (up to 1 TB per node), oriented towards memory-demanding applications and applications requiring larger numbers of CPUs communicating with shared memory (suitable for finer parallelisation).

Number of users



- NUMA (Non-Uniform Memory Architecture) server (SGI UV2)—with extremely large shared memory (6TB) and reasonably high number of CPU cores (288). This machine is available for extremely demanding applications like climate modeling.
- Cluster with 6 newest Intel Xeon Phi 7210 many core processors with massively-parallel architecture consisting of high number of x86 cores for experiments with up-to-date manycore systems.

Storage

Storage facilities of two types are available. Standard disk arrays to keep data used for computations on the clusters (1 PB), and HSM (hierarchical storage management) system (4 PB) built on MAID (massive array of idle disks), intended for large scale and archival data.

Software

CERIT-SC provides access to a complex software system based on Debian Linux but covering other Linux flavours and also MS Windows, wide range of compilers and developer tools, and application software (both freely available and commercial). Nowadays we have hundreds of applications from tens research areas available, through the whole National Grid Infrastructure.

Among the most important application packages are Gaussian, CRC Biobench, Geneious, Turbomole, Gromacs, Amber, Matlab, Maple, Mathematica, Ansys Fluent, wide range of compilers and program development environments, ...

RI CERIT-SC

The Research Infrastructure CERIT-SC provides flexible storage and computing resources, related services and extensive expertise for both experimental and production use. Provision of these resources is complemented with extensive research activities, focusing both on the e-infrastructure itself and its own research challenges and on its flexible advanced use carried in cooperation with other scientific communities.