E-Infrastructures
H2020- INFRAEDI-2018-2020

INFRAEDI-01-2018: Pan-European High Performance Computing infrastructure and services (PRACE)

PRACE-6IP

PRACE Sixth Implementation Phase Project

Grant Agreement Number: INFRAEDI-823767

D2.3

PRACE within the evolving European HPC ecosystem

Final

Version: 1.0
Author(s): Philippe Segers, GENCI, Stelios Erotokritou, CaSToRC
Date: 14/01/2022

Project and Deliverable Information Sheet

PRACE Project	**Project Ref. №: INFRAEDI-823767**	
	Project Title: PRACE Sixth Implementation Phase Project	
	Project Web Site: https://www.prace-ri.eu/about/ip-projects/	
	Deliverable ID: D2.3	
	Deliverable Nature: Report	
	Dissemination Level: PU*	**Contractual Date of Delivery:** 31 / 12 / 2021
		Actual Date of Delivery: 14 / 01 / 2022
	EC Project Officer: Leonardo Flores Añover	

* - The dissemination level are indicated as follows: **PU** – Public, **CO** – Confidential, only for members of the consortium (including the Commission Services) **CL** – Classified, as referred to in Commission Decision 2005/444/EC.

Document Control Sheet

Document	**Title: PRACE within the evolving European HPC ecosystem**	
	ID: D2.3	
	Version: 1.0	**Status:** ***Final***
	Available at: https://www.prace-ri.eu/about/ip-projects/	
	Software Tool: Microsoft Word 2016	
	File(s): D2.3_v1.0.docx	
Authorship	**Written by:**	Philippe Segers, GENCI, Stelios Erotokritou, CaSToRC
	Contributors:	Vania Markova, NCSA Oriol Pineda, BSC Veronica Teodor, JUELICH Jean-Philippe Nominé, CEA Debora Testi, CINECA Belgin Övet, UHEM Wahid Rofagha, PRACE aisbl Pedro Alberto, UC-LCA
	Reviewed by:	Thomas Eickermann, JUELICH Venkatesh Kannan, ICHEC
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.1	27/10/2020	1st Draft	1st compiled version
0.2	01/11/2021	2nd Draft	Updated draft
0.3	11/11/2021	3rd Draft	Updated draft
0.4	21/11/2021	4th Draft	1st full draft
0.5	25/11/2021	5th Draft	2nd full draft

Version	Date	Status	Comments
0.6	26/11/2021	6th Draft	1st almost finished
0.7	06/12/2021	7th Draft	Reviewed version
0.8	08/12/2021	8th Draft	Fully reviewed version
0.9	12/12/2021	Pre-Final	Version submitted to 2nd review
0.92	21/12/2021	Pre-final	Response to 2nd Review
0.93	22/12/2021	Release Candidate 1	Response to 3rd Review
0.94	23/12/2021	Release Candidate 2	Updated PRACE 3 Programme Mission and objectives
0.95	02/01/2022	Release Candidate 3	Final cleaning
0.99	05/01/2022	Release Candidate 4	References fixed
1.0	14/01/2022	Final	Final

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, Stakeholder management, EuroHPC, Peer Review, COVID-19, EOSC, Quantum Computing, AI

Table of Contents

List of Figures

List of Tables

References and Applicable Documents

[1] https://roadmap2021.esfri.eu/
[2] https://ec.europa.eu/digital-single-market/
[3] Impact Assessment accompanying the document "Proposal for a Council Regulation on establishing the EuroHPC Joint Undertaking", https://ec.europa.eu/digital-single-market/en/news/proposal-council-regulation-establishing-eurohpc-joint-undertaking-impact-assessment
[4] Top 500 The list, https://www.top500.org/
[5] Commission Staff Working Document Impact Assessment Accompanying the document Proposal for a Council Regulation on establishing the European High Performance Computing Joint Undertaking {COM(2018) 8 final} - {SWD(2018) 5 final}, http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=49295
[6] Council Regulation (EU) 2018/1488 of 28 September 2018 establishing the European High Performance Computing Joint Undertaking ST/10594/2018/INIT, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2018.252.01.0001.01.ENG&toc=OJ:L:2018:252:FULL
[7] https://eurohpc-ju.europa.eu/
[8] https://www.hpcqs.eu/
[9] PRACE-6IP, D2.4 “Report on PRACE strategic and operational developments towards EuroHPC-JU”
[10] http://prace-userforum.eu/
[11] https://prace-ri.eu/about/organisation/
[12] https://hpc-portal.eu
[13] https://www.european-processor-initiative.eu/
[14] https://eurohpc-ju.europa.eu/research-and-innovation-advisory-group-riag-0

[15] https://www.ppi4hpc.eu/
[16] https://www.hpccoe.eu/
[17] https://www.hpccoe.eu/hpc-coe-council/
[18] https://eosc.eu/
[19] https://eoscfuture.eu/
[20] https://qt.eu/app/uploads/2020/04/Strategic_Research-_Agenda_d_FINAL.pdf
[21] http://aqmo.irisa.fr/fr/accueil/
[22] https://digital-strategy.ec.europa.eu/en/policies/destination-earth
[23] www.prace-ri.eu/hpc-access/shape-programme
[24] http://www.prace-ri.eu/white-papers/
[25] https://lexis-project.eu/web/
[26] https://summerofhpc.prace-ri.eu/
[27] https://prace-ri.eu/about/diversity-inclusivity/
[28] https://prace-ri.eu/about/ip-projects/public-deliverables/ Repository for all above-mentioned PRACE Deliverables
[29] PRACE-PP, D2.5.1 "Analysis of the HPC Ecosystem"
[30] PRACE-PP, D2.5.2 "Report on links with HPC Ecosystem"
[31] PRACE-1IP, D2.3.2 "Funding scenarios for the PRACE infrastructure"
[32] PRACE-3IP, D2.2 "New PRACE Relationships with other Entities"
[33] PRACE-3IP, D5.3.3 "Report on the SHAPE implementation"
[34] PRACE-4IP, D2.1 "First Report on PRACE 2.0 Development"
[35] PRACE-6IP, D5.2 "Market and Technology Watch Report Year 2" (to be updated)
[36] https://ec.europa.eu/programmes/horizon2020/
[37] https://www.horizon-eu.eu/
[38] ETP4HPC www.etp4hpc.eu
[39] https://www.etp4hpc.eu/european-hpc-handbook.html
[40] https://www.etp4hpc.eu/sra.html
[41] http://www.etp4hpc.eu/en/euexascale.html
[42] http://www.eesi-project.eu/
[43] https://exdci.eu/
[44] https://exdci.eu/resources/prace-scientific-case
[45] https://exdci.eu/resources/sra
[46] https://www.humanbrainproject.eu/en/
[47] https://ebrains.eu/
[48] https://digital-strategy.ec.europa.eu/en/policies/european-quantum-communication-infrastructure-euroqci
[49] http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
[50] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0045:FIN:EN:PDF
[51] https://ec.europa.eu/digital-single-market/en/high-performance-computing-contractual-public-private-partnership-hpc-cppp
[52] http://ec.europa.eu/programmes/horizon2020/en/news/overview-eu-funded-centres-excellence-computing-applications
[53] http://www.etp4hpc.eu/cppp-monitoring.html
[54] http://www.etp4hpc.eu/news/182-etp4hpc-annual-report017-now-available-on-lin.html
[55] https://ec.europa.eu/digital-single-market/en/eurohpc-joint-undertaking
[56] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2018:252:FULL&from=EN
[57] http://montblanc-project.eu/
[58] EUDAT CDI https://www.eudat.eu/eudat-collaborative-data-infrastructure-cdi
[59] https://www.ff4eurohpc.eu/
[60] https://fenix-ri.eu/
[61] https://www.gaia-x.eu/

[62] https://en.wikipedia.org/wiki/Big_Tech#GAFAM_or_FAAMG
[63] https://en.wikipedia.org/wiki/BATX
[64] https://www.ovhtelecom.fr/
[65] http://82.116.198.186:8080/search-training-events
[66] https://prace-ri.eu/hpc-access/deci-access/
[67] https://github.com/
[68] https://repository.prace-ri.eu/git/CodeVault
[69] https://repository.prace-ri.eu/git/explore

List of Acronyms and Abbreviations

AAI	Authentication Authorisation Infrastructure
AI	Artificial Intelligence
aisbl	Association International Sans But Lucratif (legal form of the PRACE-RI)
BATX	Baidu, Alibaba, Tencent, and Xiaomi
CERN	Conseil Européen pour la Recherche Nucléaire (French: European Organisation for Nuclear Research)
CoE	Centre of Excellence
CPU	Central Processing Unit
DECI	Distributed European Computing Initiative
DL	Deep Learning
DSM	Digital Single Market
EC	European Commission
EDI	European Data Infrastructure
EOSC	European Open Science Cloud
EPI	European Processor Initiative
ETP4HPC	European Technology Platform for High Performance Computing
EUDAT	European Data Infrastructure
EXDCI	European eXtreme Data and Computing Initiative
FAIR	Findable, Accessible, Interoperable and Reusable
FETHPC	Future and Emerging Technologies HPC
GAMAM	Google, Amazon, Meta, Apple, Microsoft
GÉANT	Collaboration between National Research and Education Networks to build a multi-gigabit pan-European network.
GP	PRACE General Partners
GPU	Graphic Processing Unit
H2020	Horizon 2020
HLST	High Level Support Team
HPC	High Performance Computing; Computing at a high performance level at any given time; often used synonym with Supercomputing
IAC	PRACE Industrial Advisory Committee
ICEI	Interactive Computing E-Infrastructure for the Human Brain Project
JU	Joint Undertaking
ML	Machine Learning
ML	Maximum Likelyhood
NCP	National Contact Point in Horizon 2020
PA	Preparatory Access (to PRACE resources)
PTC	PRACE Training Centres
PMO	Project Management Office
PRACE	Partnership for Advanced Computing in Europe; Project Acronym
PRACE 1	Initial five years period of the PRACE Research Infrastructure
PRACE 2	Second period of the PRACE Research Infrastructure following the initial period
PRACE 3	The period of PRACE Research Infrastructure in the EuroHPC era
RI	Research Infrastructure
RIAG	Research and Innovation Advisory Group
SHAPE	PRACE SME HPC Adoption Programme in Europe
SKA	Square Kilometre Array
SME	Small and medium-sized enterprises

SSC	PRACE Scientific Steering Committee
SWG	PRACE Strategy Working Group
TB	Technical Board (group of Work Package leaders)
Tier-0	Denotes the apex of a conceptual pyramid of HPC systems. In this context the Supercomputing Research Infrastructure would host the Tier-0 systems; national or topical HPC centres would constitute Tier-1
Tier-1	National or topical HPC centres

List of Project Partner Acronyms

BADW-LRZ	Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, Germany (3rd Party to GCS)
BILKENT	Bilkent University, Turkey (3rd Party to UHEM)
BSC	Barcelona Supercomputing Center - Centro Nacional de Supercomputacion, Spain
CaSToRC	The Computation-based Science and Technology Research Center (CaSToRC), The Cyprus Institute, Cyprus
CCSAS	Computing Centre of the Slovak Academy of Sciences, Slovakia
CEA	Commissariat à l'Energie Atomique et aux Energies Alternatives, France (3rd Party to GENCI)
CENAERO	Centre de Recherche en Aéronautique ASBL, Belgium (3rd Party to UANTWERPEN)
CESGA	Fundacion Publica Gallega Centro Tecnológico de Supercomputación de Galicia, Spain, (3rd Party to BSC)
CINECA	CINECA Consorzio Interuniversitario, Italy
CINES	Centre Informatique National de l'Enseignement Supérieur, France (3 rd Party to GENCI)
CNRS	Centre National de la Recherche Scientifique, France (3 rd Party to GENCI)
CSC	CSC Scientific Computing Ltd., Finland
CSIC	Spanish Council for Scientific Research (3rd Party to BSC)
CYFRONET	Academic Computing Centre CYFRONET AGH, Poland (3rd Party to PNSC)
DTU	Technical University of Denmark (3rd Party of UCPH)
EPCC	EPCC at The University of Edinburgh, UK
EUDAT	EUDAT OY
ETH Zurich (CSCS)	Eidgenössische Technische Hochschule Zürich – CSCS, Switzerland
GCS	Gauss Centre for Supercomputing e.V., Germany
GÉANT	GÉANT Vereniging
GENCI	Grand Equipement National de Calcul Intensiv, France
GRNET	National Infrastructures for Research and Technology, Greece
ICREA	Catalan Institution for Research and Advanced Studies (3rd Party to BSC)
INRIA	Institut National de Recherche en Informatique et Automatique, France (3rd Party to GENCI)
IST-ID	Instituto Superior Técnico for Research and Development, Portugal (3rd Party to UC-LCA)
IT4I	Vysoka Skola Banska - Technicka Univerzita Ostrava, Czech Republic
IUCC	Machba - Inter University Computation Centre, Israel
JUELICH	Forschungszentrum Juelich GmbH, Germany

KIFÜ (NIIFI)	Governmental Information Technology Development Agency, Hungary
KTH	Royal Institute of Technology, Sweden (3rd Party to SNIC-UU)
KULEUVEN	Katholieke Universiteit Leuven, Belgium (3rd Party to UANTWERPEN)
LiU	Linkoping University, Sweden (3rd Party to SNIC-UU)
MPCDF	Max Planck Gesellschaft zur Förderung der Wissenschaften e.V., Germany (3rd Party to GCS)
NCSA	National Centre for Supercomputing Applications, Bulgaria
NTNU	The Norwegian University of Science and Technology, Norway (3rd Party to SIGMA2)
NUI-Galway	National University of Ireland Galway, Ireland
PRACE	Partnership for Advanced Computing in Europe aisbl, Belgium
PSNC	Poznan Supercomputing and Networking Center, Poland
SDU	University of Southern Denmark (3rd Party to UCPH)
SIGMA2	UNINETT Sigma2 AS, Norway
SNIC-UU	Uppsala Universitet, Sweden
STFC	Science and Technology Facilities Council, UK (3rd Party to UEDIN)
SURF	SURF is the collaborative organisation for ICT in Dutch education and research
TASK	Politechnika Gdańska (3rd Party to PNSC)
TU Wien	Technische Universität Wien, Austria
UANTWERPEN	Universiteit Antwerpen, Belgium
UC-LCA	Universidade de Coimbra, Labotatório de Computação Avançada, Portugal
UCPH	Københavns Universitet, Denmark
UEDIN	The University of Edinburgh
UHEM	Istanbul Technical University, Ayazaga Campus, Turkey
UIBK	Universität Innsbruck, Austria (3rd Party to TU Wien)
UiO	University of Oslo, Norway (3rd Party to SIGMA2)
UL	UNIVERZA V LJUBLJANI, Slovenia
ULIEGE	Université de Liège; Belgium (3rd Party to UANTWERPEN)
U Luxembourg	University of Luxembourg
UM	Universidade do Minho, Portugal, (3rd Party to UC-LCA)
UmU	Umea University, Sweden (3rd Party to SNIC-UU)
UnivEvora	Universidade de Évora, Portugal (3rd Party to UC-LCA)
UnivPorto	Universidade do Porto, Portugal (3rd Party to UC-LCA)
UPC	Universitat Politècnica de Catalunya, Spain (3rd Party to BSC)
USTUTT-HLRS	Universitaet Stuttgart – HLRS, Germany (3rd Party to GCS)
WCSS	Politechnika Wroclawska, Poland (3rd Party to PNSC)

Executive Summary

This deliverable presents the position of the PRACE ESFRI landmark [1] Research Infrastructure within the evolving European HPC ecosystem in the context of the Digital Single Market (DSM [2]).

The European HPC ecosystem changed substantially in the last few years. New stakeholders joined the landscape, with the establishment of the EuroHPC Joint Undertaking (JU) [3][5][6][7], where the European Commission and Member States invest considerably in developing a world class HPC and quantum computing ecosystem in Europe [8]. These systems will provide valuable compute cycles to the European academic and industrial research communities and to public administrations. PRACE RI has established, over the past decade, a strong relationship with European academic and industrial users and HPC centres in a bottom-up approach and developed a sound, recognised and science-driven peer review process for the access to world-class computing and data management resources in Europe, with many interactions to our stakeholder's ecosystem.

After a short Introduction in Chapter 1, an update on PRACE stakeholder relationship within European HPC ecosystem is performed in Chapter 2 with regard to the three pillars of European HPC strategy, namely the Infrastructure, the Application and the Technology, providing a global overview of the European HPC ecosystem, the place of PRACE within it and the specific role of coordination activities that are performed by some specific projects. Chapter 3 presents new services for industry that could be offered, derived from a survey performed by PRACE, and presenting a catalogue of software dedicated for industry on Chapter 4. A more detailed presentation of PRACE updated Business Model, PRACE-EuroHPC relation and PRACE specific Services in the EuroHPC era are provided in D2.4, a complementary Deliverable [9].

These many interactions with PRACE stakeholders took place in the context of a global competition, on which some players already reach Exascale [4], but what will matter the most in the long run is the efficiency of a whole ecosystem to harvest the benefit of these tools for society, who can rely on a persistent Research Infrastructure, and where services are offered on an operational basis, on a FAIR (Findable, Accessible, Interoperable and Reusable) and transparent way, based on international best practice of access to world class RIs.

At this stage, it remains a huge challenge to understand this European HPC ecosystem, which is more fragmented than ever through many projects. For instance, feedback from PRACE main stakeholders, our users, through PRACE User Forum [10], PRACE Scientific Steering Committee and Industrial Advisory Committee [11] is the lack of clarity of this European HPC ecosystem. PRACE aim, within this Deliverable, is to contribute to provide a humble tool to understand who are the main actors of this ecosystem, what are their relations (if any), and what level of coordination is put in place to ensure some coherency between them.

PRACE, with a clear proposal of services and activities, integrates the active involvement of user communities to support their scientific discoveries and research developments. Through an open and transparent governance, PRACE provides to these communities (represented by the Scientific Steering Committee and the Industrial Advisory Committee), jointly with HPC agencies and centres, the power to drive the PRACE activities. This also guarantees that principles underpinning scientific excellence are applied to the peer review processes distributing resources to the users.

In this context, the mission of PRACE is to enable high-impact scientific discovery and research and development across all disciplines to enhance European competitiveness for the benefit of society.

1 Introduction

PRACE-6IP is the 6th Implementation Phase of PRACE project, continuing a series of successful projects that started 13 years ago (see [28] to [35]) and now involving researchers collaborating from 49 organisations in 26 countries, with many interactions with other organisations and projects. A complex and rich ecosystem of internal stakeholders, managed by a PMO, using a sound Governance scheme, that links PRACE-6IP project to the PRACE aisbl Research Infrastructure and its bodies. That management is taking place within a fast-moving ecosystem, in the constant flow of innovative technology that is characteristic of High-Performance Computing, but also with the huge impact of many projects taking place with funding from the European Commission, and especially with the action of the EuroHPC Joint Undertaking [6], at the center of European strategy.

"Europe is facing challenges due to environmental and societal change. In order to sustain and develop European welfare, new approaches are necessary in which the research community uses scientific computing and reliable simulation models to make trustworthy predictions. This is only possible through wide ranging collaboration, crossing science disciplines and country borders." [29] These were the first words of the first Analysis of the HPC Ecosystem performed in 2008 by the PRACE-PP project. This is truer than ever 13th years after, with PRACE Research Infrastructure building-up these collaborations for more than 10 years, helping European HPC users to access world class resources (see below for European resources share within the global HPC ecosystem).

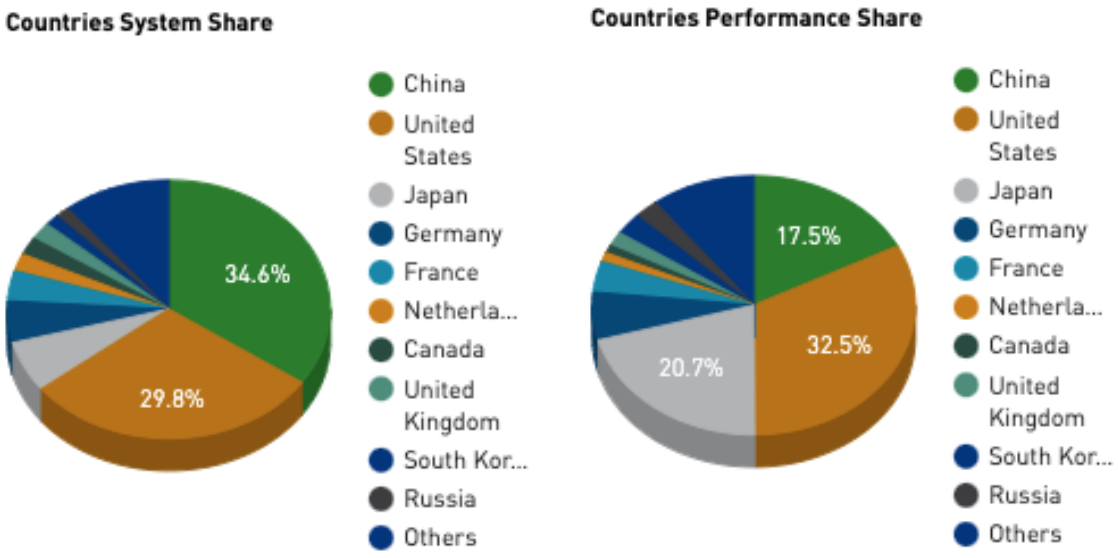

Figure 1: Countries Systems Share and Performance Share, from Nov. 2021 Top500

To provide a comprehensive view of the position of PRACE within the evolving European HPC ecosystem, we will provide in Chapter 2 an update on PRACE stakeholder relationship within the European HPC ecosystem with regard to the three pillars of European HPC strategy, namely the Infrastructure, the Application and the Technology, along with the coordination activities aiming at coordinating together all these projects related to these pillars, also providing a global overview of the European HPC ecosystem and all their major HPC projects. In Chapter 3 we will present some new services for industry that could be offered, derived from a survey performed by PRACE, and we will also present on Chapter 4 a catalogue of major HPC application dedicated for industry.

2 Update on the European HPC Ecosystem

The European HPC Ecosystem can be described through the view of three "virtual strategic pillars", Infrastructure, Technologies and Applications. Each pillar encompasses different entities or initiatives, together with many related European technical or coordination projects – activities which are closely tied to the EuroHPC JU or other European funding programmes and agencies (Digital Europe, Horizon Europe, INEA/CINEA, etc.). A comprehensive description of EU HPC related projects is provided by the ETP4HPC Handbook [39], and Table 2 in annex. This very complex ecosystem can be represented by the following figure:

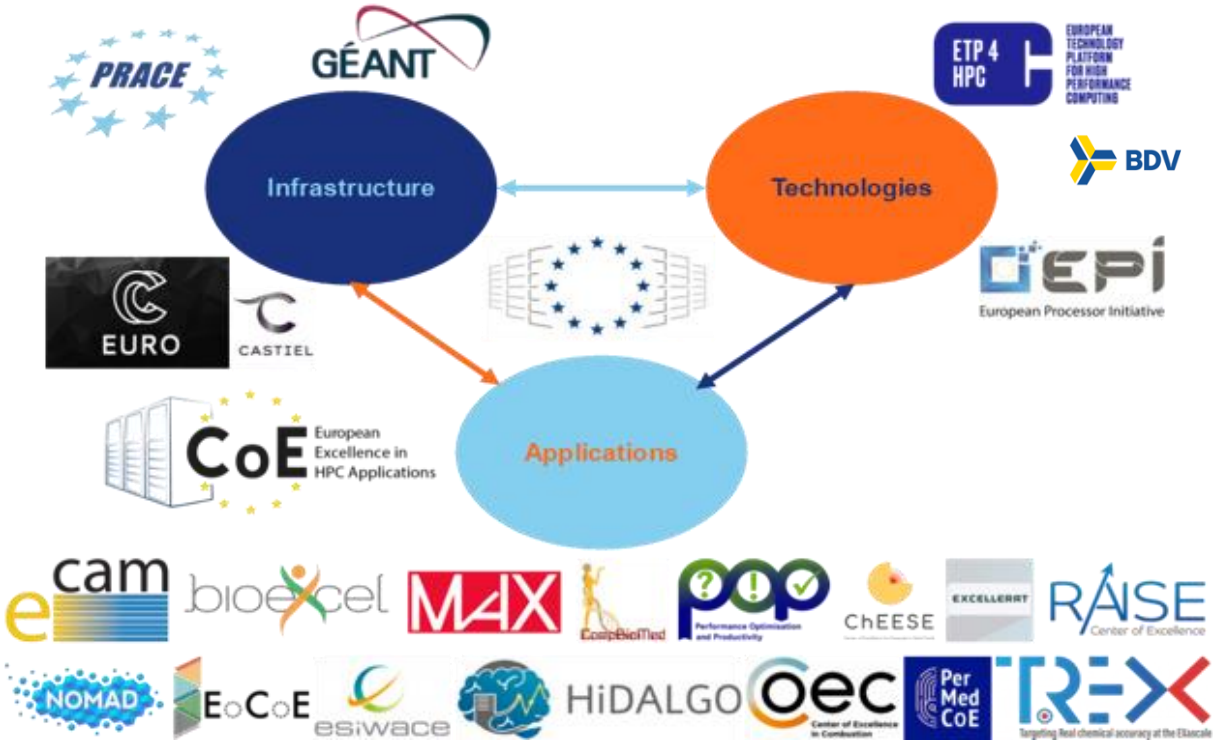

Figure 2: European HPC Ecosystem – some entities and projects

We will try to describe in the following sub chapters the main components of these three pillars, along with the projects in charge of trying to coordinating them and also providing a full list of main European HPC projects.

2.1 Infrastructure Pillar

The **Infrastructure Pillar** is built from the HPC Infrastructure – the bulk of which was provided by PRACE RI and its partners during the PRACE 1 and PRACE 2 Programme, as provider of the access to national computational resources and as operator of the Peer Review allocation process. The underlying network is provided by GEANT, which connects the PRACE HPC systems. We are actually in a transition period, where the main computational resources will be provided by HPC centres with systems owned by EuroHPC or co-owned by EuroHPC and Member states, and where the Peer Review allocation process will still be operated by PRACE, for at least a contractual time of two years. This new organisation will be framed in a so-called PRACE 3 Programme that is currently under preparation, organising the contractual relationship between PRACE and EuroHPC JU. This will provide a federated European HPC infrastructure for European HPC users and allows them to seamlessly use more than one system within their allocation period.

The PRACE 3 Programme is planned to handle the following updated missions and objectives:

- To provide European researchers in science and industry with access to a world-class Research Infrastructure, offering HPC and data management, Training, Support and Knowledge services;
- To support and organise the European HPC User Forum for academic and industrial users as well as public administrations;
- To foster international collaborations on the forefront of high-end computing in simulation and data science and to bring competence to the PRACE member states;
- To help develop Europe by promoting the European idea of bringing stability and peace through open scientific discourse between all members;
- To enhance the scientific output of the supercomputing systems through international Tier-0 allocations coupled with the assistance of local High-Level Support Teams;
- To foster healthy research competition through the unique and purely scientific review-based process that promotes scientific excellence;
- To stimulate the deployment of HPC in the knowledge economy in Europe and help European industry to become more competitive.

Other new players are also offering HPC resources at European level, such as FF4HPC [59], FENIX [60], GAIA-X [61] for publicly funded projects, or as new cloud offers from so-called "hyperscale Data Centre" (GAMAM (Google, Amazon, Meta, Apple, Microsoft [62]) and BATX (Baidu, Alibaba, Tencent, and Xiaomi [63])) or other private cloud providers such as the OVH Cloud, CloudSigma or 1&1 Ionos [64], some European companies. European Researchers can also access other digital resources through EOSC projects, whereas in the future EOSC could also offer a direct link to the HPC resources offered by/through PRACE, keep using the conventional Peer Review process for large allocations or other options for smaller allocations (such as the foreseen fast-track). This Infrastructure is and will continue to be a heterogeneous portfolio of systems, there is no such thing as one size fits all resources in HPC, with a constant evolution of concurring technologies and diverse usage models. The value of such infrastructure lies not only in its computational performance in floating-point operations per second, but also in data handling [58] and processing capacities, and in the surrounding knowledge services delivered to users, from Peer-Review allocation to HPC expertise, High-level support, code enabling and optimisation for applications, in a structured, persistent and efficient liaison with its stakeholders. In other words, services that can only be provided efficiently and on the long run by a Research Infrastructure.

2.2 Application Pillar

The **Application Pillar** encompasses 15 European Centres of Excellence (CoEs) for High-Performance Computing (HPC) applications (see Figure 2). They are independent projects which were selected under Horizon2020 [36] and HorizonEurope [37] calls for a period of up to three years, in three successive rounds between 2015 and 2020. They promote the use of upcoming exascale and extreme performance computing capabilities, and support the scaling up of existing parallel codes towards exascale performance for a selected set of organised communities, or expanding parametric or ensemble throughput HPC use, or else providing some generic cross-cutting services to all applications. Of course, other communities and teams are also developing applications for HPC, some of them with high scalability and targeting extreme scale. These applications are developed by academic teams or communities, in industry or in partnership between industry (sometimes SMEs) and academia.

This application pillar is a European asset, a sector where Europe has a very strong position due to long term investment in a skilled workforce that developed world-class codes, the largest of them representing the equivalent of hundreds of years of work (big teams working for decades). The drawback of such a position is the huge work for porting legacy codes on new architectures, with sometimes also some algorithm architecture hampering efficient porting. Subsequent validation/certification processes also require significant efforts for industrial grade applications. The role of the different CoEs is very important to keep that competitive advantage, which requires also a continuous effort of teaching and training to help these communities to acquire and keep the necessary computational skills. HPC centres bring their share of support, from helpdesk to application deployment, to more significant port codes porting. High-Level Support Team (HLST) like PRACE put in place can offer deeper and longer-term code adaptation. PRACE IP projects and their Work Package 4 supported PRACE Training Centres (PTCs), and more recently extra dedicated support started being provided by the network of National Centres of Competences (NCC) projects funded by EuroHPC. It also involved Work Package 7 on Code Enabling and Benchmark activity, along with the specific scope of the Work Package 8 on code re-engineering, offered extra application enabling resources.

The operational aspect of this Infrastructure is of course mostly relying on operational activity of the HPC Centres, with the help of Work Package 6 devoted to operation, from the management of the PRACE catalogue of services, that define a common set of services (also assessing and testing new innovative services), to operational activity on security, networking, data management or AAI (Authentication and Authorization Infrastructure), that are of huge importance for the Centres and for the users who are developing applications that rely on these operational services and their implementation. Strong links between PRACE and application communities are maintained through different bodies, such as the Scientific Steering Committee (SSC), the Industrial Advisory Committee (IAC) and User Forum.

Last but not least, in this very rapidly moving area, it is also very important to maintain technology awareness, and contact with Technology Pillar actors. In particular early access to prototypes and new equipment, both hardware and software, can benefit to the efficiency of applications, which requires a good understanding of the underlying technologies and of their foreseeable evolutions, which must be foreseen far ahead of the day when these new technologies are deployed.

2.3 Technology Pillar

The **Technology Pillar**, for the part that is funded by European agencies, encompasses many projects – in particular from 2014 onwards with Horizon 2020 HPC programme under the umbrella of HPC contractual Public Private Partnership, and now with EuroHPC.

The ETP4HPC [38] independent association develops roadmaps for the orientation of the EU HPC R&D programmes - in the area of HPC technology and usage, by issuing a Strategic Research Agenda (SRA) [40]. ETP4HPC is a private member of EuroHPC and is represented in the EuroHPC Research and Innovation Advisory Group (RIAG) [14].

The European Processor Initiative (EPI) [13] is a large H2020 and now EuroHPC project, aiming to design and build a new family of European low-power processors for HPC applications and emerging applications such as machine learning. EPI has Specific Grant Agreements under a Framework Partnership Agreement (FPA).

Figure 3: EPI roadmap [13]

A comprehensive list of such EC-funded Technological projects can be found in ETP4HPC Handbook on HPC project [39], and is summarised in Annex 1 Table 2.

The private sector (EU technology suppliers) also invests in R&D on their own, but even combined with EU public funding, Europe lags behind other large HPC players like the USA or China. This is reflected in Top 500 ranking and the shares of European vendors for instance. However, there is momentum since a few years, the EU providers share growing at a regular pace, and EuroHPC should further stimulate this with ambitious equipment plans.

Figure 4 and Figure 5 give an idea of vendors' presence, resp. directly via their Top500 shares, or by countries. The acquisition strategies differ according to the regions, Europe being the most open market.

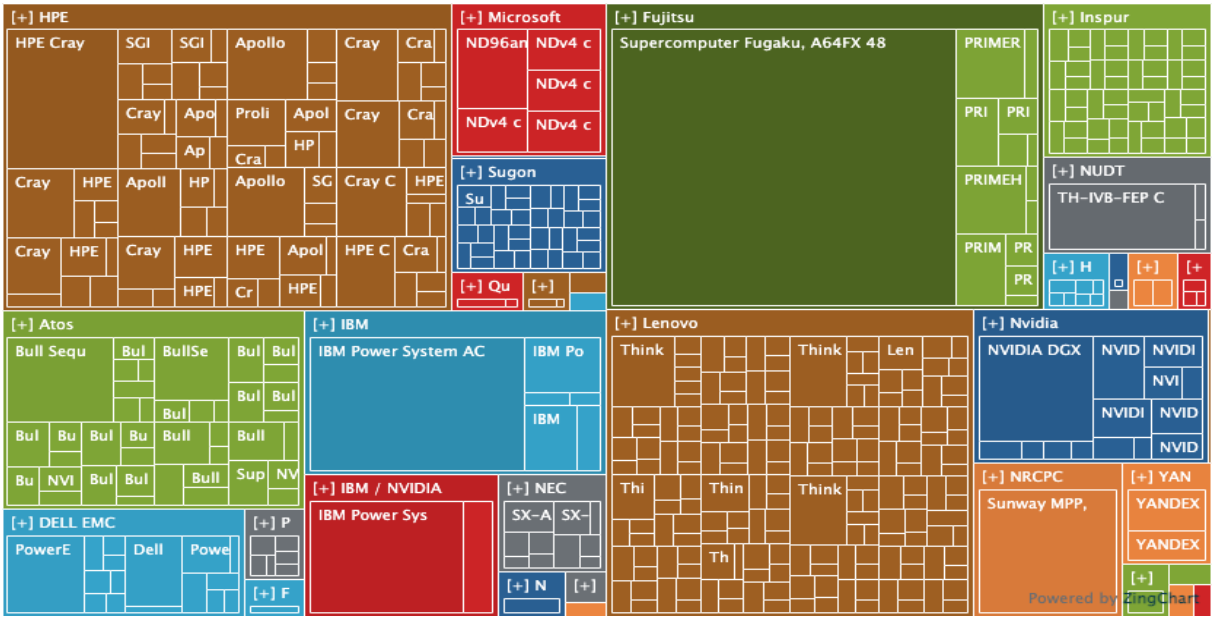

Figure 4: Share of HPC Technology from Nov. 2021 Top500

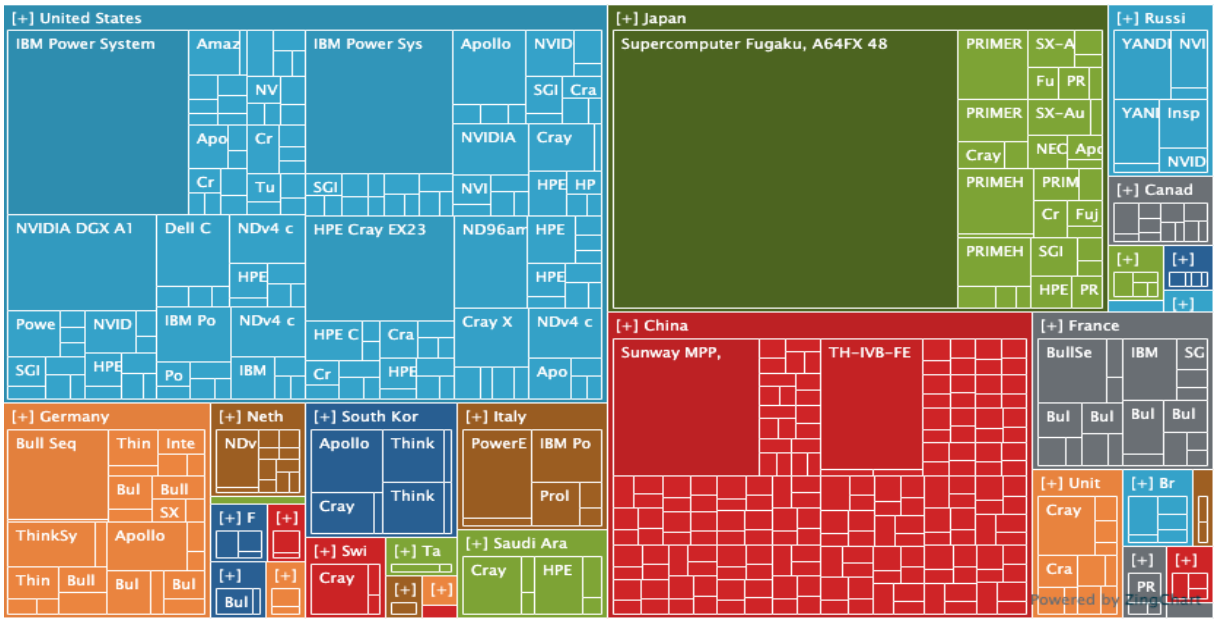

Figure 5: Share of performance per Country and per Technology provider, from Nov 2021 Top500

PRACE projects are not directly involved in the Technology pillar, but PRACE-3IP has led a successful Pre-Commercial Procurement (PCP), that has translated into products for the 3 selected technology providers, with pilot systems used for benchmarking activities in PRACE-4IP and -5IP. Some PRACE partners have also been involved in a Procurement for Innovative Solutions (PPI4HPC, with GENCI as lead procuring entity and BSC, CEA, CINECA and JSC as partners [15]), and within the FENIX projects (see below).

Another link to technology is also the Work Package 5 on Technology watch, which provides an independent analysis on emerging technologies, such as AI, Quantum Computing, or important technological issues such as energy efficiency (see below and [8] [35]).

2.4 Coordinating Project Activities (Ecosystem Development)

All the above projects and entities are representing a very rich, dense and complex HPC ecosystem. Without entering into details, projects in Europe are grouped by Work Programmes and related calls, with specific timelines and durations. From forming a consortium and submitting a project to kicking off a project, delays can be quite significant, and it is not easy to foresee future interaction with other projects not existing at the time of the preparation of one project, and also difficult to adapt during the project lifespan, as each project is liable to its own grant agreement.

The diversity of actors and activities adds to this complexity, and makes coordination a difficult task, since there is no overarching management. A level at which coordination and flexibility could better take place is inside a project, within its different Work Packages, which is one of the roles of a Technical Board and of a Management Board. Coordination between projects is far more difficult, to take an analogy with High-Performance Computing this is characteristic of two different situations, shared memory (inside a project where Work Packages can share information inside the Technical Board) and distributed memory (between projects, which needs to put in place ways to exchanges information to coordinate themselves, cross-activity or workshop for instance).

This is a reason why PRACE aisbl and PRACE IP projects have put in place some Memorandums of Understanding with other stakeholders, in addition to projects with formal coordination roles (CSA). This goes along with some more informal cross-cutting activities, such as workshops, participation to conferences or hackathons. Some projects have been put in place to coordinate activities inside a pillar, others for coordinating activities across pillars.

For instance, between the application and infrastructure pillars lies the EuroCC project – which is supported by CASTIEL. The EuroCC project and the 33 HPC National Competence Centres it embeds aim to increase the HPC competences of national computational communities - so they could better use the European HPC infrastructure and applications available to them.

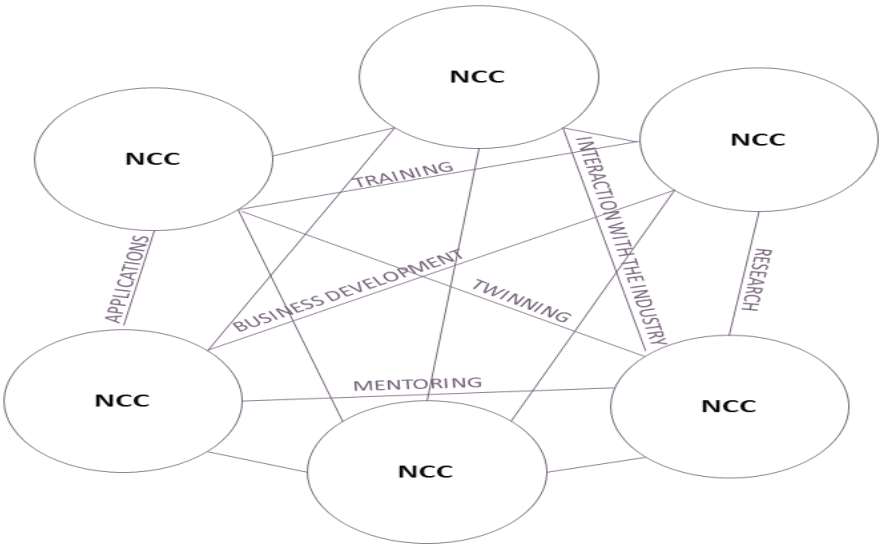

Figure 6: CASTIEL and EuroCC project goal to federate National Competence Centers across Europe

The EuroCC project brings together the necessary expertise to set up a network of National Competence Centres (NCCs) in HPC across participating, member and associated states. The goal is to provide a broad service portfolio tailored to the respective national needs of industry, academia and public administrations. All of this is implemented to support and strongly increase the national strengths of HPC competences as well as High Performance Data Analytics (HPDA) and Artificial Intelligence (AI) capabilities. This will close existing gaps and increase

usability of these technologies in the different states which will provide a European excellence baseline. The mandate of most of these NCCs is to support industry and SMEs in the uptake of HPC and related technologies in their work processes. CASTIEL – which is the Coordination and Support Action associated with EuroCC, aims to assist NCCs by providing industry related mentoring activities by organising respective workshops. By following these workshops, inexperienced NCCs can gain an understanding of what is required in order to stimulate the adoption of HPC, HPDA and AI technologies by industry – and especially by SMEs. To this end, CASTIEL has already organised two events in which speakers from NCCs or service companies report on their activities with regard to engaging with industry. They share their experiences with regards to what has worked and has not worked in their interactions with companies, including any issues with training, software and usage of HPC facilities whenever applicable.

FocusCoE [16] aims at coordinating the 15 CoEs inside the application Pillar. For that purpose, FocusCoE has initiated the foundation of the HPC CoE Council (HPC3) [17], a body comprising all active HPC CoEs and aiming at strengthening the role of applications in the European HPC ecosystem. FocusCoE is continually supporting the operations of HPC3, and is active in establishing connections to a number of other players in the HPC eco-system, such as CASTIEL, EuroCC or the petascale and pre-exascale consortia.

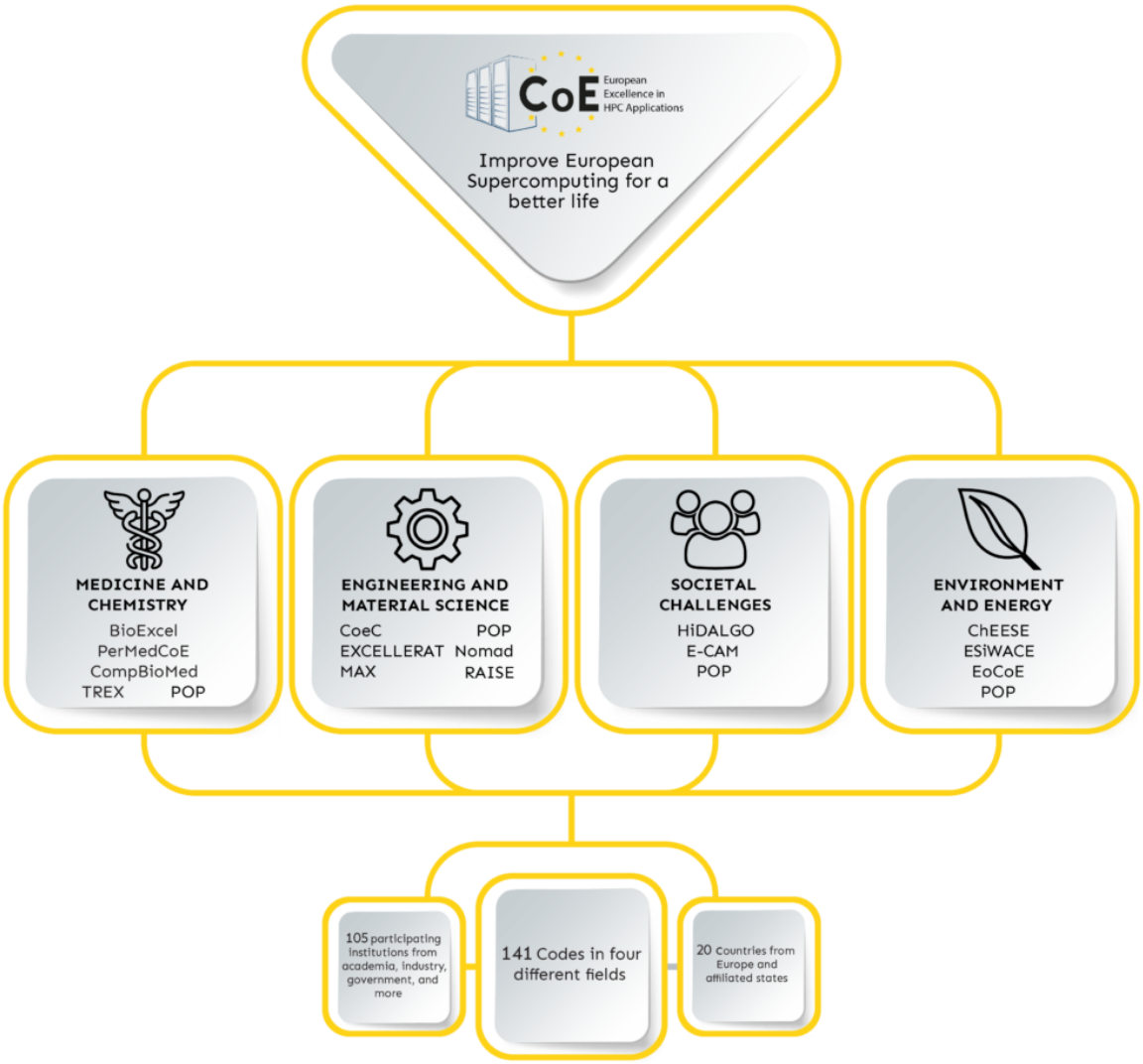

Figure 7: Focus CoE sketch-up of relations between all CoEs [16]

EXDCI-2 (European eXtreme Data and Computing Initiative [43]) played also an important role in the past (from 2018 to 2020) to coordinate HPC activity across Europe, with PRACE aisbl as PI and ETP4HPC as partners. The strategic objectives of EXDCI were the development and advocacy of a competitive European HPC Exascale Strategy and the coordination of the stakeholder community for European HPC at the Exascale. EXDCI has improved the HPC awareness by developing international event such as the EuroHPC Summit Week and by targeting specific audience through dedicated media by disseminating the achievements of the European HPC ecosystem. It also coordinates the edition of the "Scientific Case for Computing in Europe 2018 – 2026" [44] and of the 4th "Strategic Research Agenda" [45]. There is no other project continuing such coordination task at European level, but PRACE is still taking some coordinative role, including for instance maintaining the HPC in Europe Portal [12].

2.5 Other Pieces of the EU HPC Ecosystem

Further to the European HPC infrastructure depicted in Figure 2 other projects and initiatives are related and these can be considered as part of the extended European HPC Ecosystem. We describe the most significant ones of these below:

- EOSC
  - The ambition of the European Open Science Cloud (EOSC) is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation and educational purposes, aiming at a seamless access, FAIR (Findability, Accessibility, Interoperability and Reusability) management and reliable reuse of research data and all other digital objects produced along the research life cycle. EOSC ultimately aims to develop a Web of FAIR Data and services for science in Europe upon which a wide range of value-added services can be built. These range from visualisation and analytics to long-term information preservation or the monitoring of the uptake of open science practices. HPC resources and services can in the future become part of the offers to researchers and scientific communities.
    Since July 2020, EOSC is managed but the EOSC Association [18] counting 161 members (including PRACE aisbl) and over 200 observers. The implementation of EOSC in particular related to the EOSC Portal Marketplace and interoperability and composability of services is part of the EC-funded project EOSC Future [19].
  - The Data Infrastructure Capacities for EOSC (DICE) consortium brings together a network of computing and data centres, and research infrastructures, for the purpose of enabling a European storage and data management infrastructure for EOSC, providing generic services and building blocks to store, find, and access data in a consistent and persistent way. Specifically, DICE partners offer 14 state-of-the-art data management services together with more than 50 PB of storage capacity. The service and resource provisioning is accompanied by enhancing the current service offering in order to fill the gaps still present to the support of the entire research data lifecycle; solutions will be provided for increasing the quality of data and their re-usability, supporting long term preservation, managing sensitive data, and bridging the gap between data and computing resources. This last-mentioned activity has the objective to make it easier for users, who are making use of HPC resources, to manage their data according to the FAIR principle, with specific care to the interoperability of the different Authentication and Authorization mechanisms.

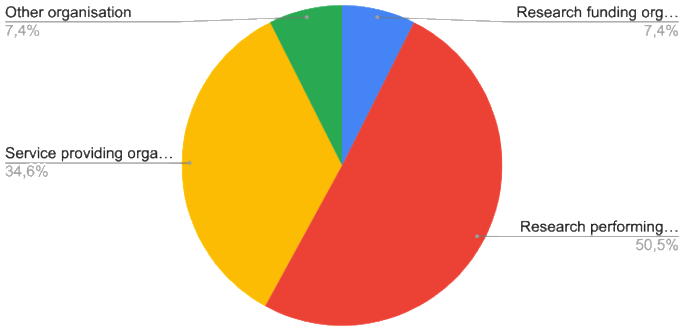

Figure 8: EOSC aisbl Members & Observers by Organisation type (Research funding org., Research performing org., Service providing org. and other kind of organisations)

- FENIX, HBP, EBRAINS
  - **FENIX**
    The Fenix Infrastructure [60] allows for access to data repositories and scalable supercomputing systems both of which are in close proximity and well-integrated. An initial version of this infrastructure is currently being realised through the ICEI project (Interactive Computing E-Infrastructure), which is part of the European

Human Brain Project (HBP). Fenix infrastructure and related resources are available to neuroscientists and European researchers at large. Collaboration has been established between PRACE and ICEI to support the requests of access to Fenix resources. The PRACE-ICEI Calls for Proposals have been offered since 2020 and combine the well-established PRACE peer review process with Fenix services and allows for proposals from European researchers in need of scalable computing resources, interactive computing services, Virtual Machine services and data storage for their projects.

- **HBP**
  The Human Brain Project (HBP) [46] is one of the largest research projects in the world with more than 500 scientists and engineers from over 140 universities, teaching hospitals, and research centres across Europe, all coming together to address one of the most challenging research targets – the human brain. To tame brain complexity, the project is building a research infrastructure to help advance neuroscience, medicine, computing and brain-inspired technologies - EBRAINS which is described below. The HBP is developing EBRAINS to create lasting research platforms that benefit the wider community. The HBP provides a framework where teams of researchers and technologists work together to scale up ambitious ideas from the lab, explore the different aspects of brain organisation, and understand the mechanisms behind cognition, learning, or plasticity. Scientists in the HBP conduct targeted experimental studies and develop theories and models to shed light on the human connectome, addressing mechanisms that underlie information processing, from the molecule to cellular signaling and large-scale networks. The project teams transfer the acquired knowledge to make an impact in health and innovation: Insights from basic research are translated into medical applications, to prepare the ground for new diagnoses and therapies.
- **EBRAINS**
  EBRAINS [47] is a new digital research infrastructure (RI), created by the EU-funded Human Brain Project, that gathers an extensive range of data and tools for brain related research. It draws on cutting-edge neuroscience, big data, computing, robotics and related technologies to help translate the latest scientific discoveries into innovation in medicine and industry, for the benefit of patients and society.
  EBRAINS' ambition is to provide the scientific community at large with an open state-of-the-art capability that fosters collaborative brain science, opens the way to ground-breaking discovery and aims to secure Europe's leading position in the dynamically growing field of multidisciplinary brain research and its exploitation.

- FF4EuroHPC
  - FF4EuroHPC [59] is a European initiative that facilitates access to all high-performance computing-related technologies for SMEs and thus increases the innovation potential of European industry. Whether it is running high-resolution simulations, doing large-scale data analyses, or incorporating AI applications into SMEs´ workflows, FF4EuroHPC connects business with cutting-edge technologies. Within the FF4EuroHPC project, European SMEs can develop unique products, innovative business opportunities and become more competitive by using European high-end HPC services though the project open calls.
- GAIA-X
  - **GAIA-X** [60] is a European project where representatives from business, politics, and science from Europe and around the globe are working together, to create a

federated and secure data infrastructure. Companies and citizens will collate and share data – in such a way that they keep control over them. They should decide what happens to their data, where it is stored, and always retain data sovereignty. The architecture of GAIA-X is based on the principle of decentralisation. GAIA-X is the result of a multitude of individual platforms that all follow a common standard – the GAIA-X standard. The developed data infrastructure is based on the values of openness, transparency, and trust. So, what emerges is not a cloud, but a networked system that links many cloud service providers together.

- Quantum Computing and Simulation
  - Since the theoretical presentation of the possibility that a quantum computer can be built in 1982, research towards constructing such a machine along with its associated required compilers has taken several successful steps. Still, the workforce needed for quantum system hardware and especially for quantum compilers is behind the desired level. In that respect, the U.S. National Science and Technology Council issued a warning about the lack of talented workforce in quantum studies. The EC has planned to support activities toward quantum computing and related skills under the EuroHPC JU (whereas core quantum technologies, incl. for computing but not only, are supported by the EU Quantum Flagship) [20].
  - As part of the EuroHPC JU, the Commission is planning to support the deployment of state-of-the-art pilot quantum computers in an HPC environment by 2023. These computers would act as accelerators interconnected with the Joint Undertaking's supercomputers, forming 'hybrid' machines that blend the best of quantum and classical computing technologies.
    End of 2021, work began on the Joint Undertaking's first quantum simulators [8], which would also be interconnected with PRACE Tier-0 supercomputers. This will be the first element of a European quantum simulation infrastructure available via the cloud on a non-commercial basis to public and private European users. The infrastructure will be used to address complex simulation and optimisation problems, especially in materials development, drug discovery, transportation and other real-world problems of high importance to industry.
  - **EuroQCI** [48] - Since June 2019, all 27 EU Member States have signed the European Quantum Communication Infrastructure (EuroQCI) Declaration, agreeing to work together, with the Commission and with the support of the European Space Agency, towards the development of a quantum communication infrastructure covering the whole of the EU.
- Digital twins
  - Urban digital twins are a virtual representation of a city's physical assets. They are connected to data related to those assets (real-time and historic), so that data analytics and machine learning can be used to help a city operate more efficiently. This is achieved by creating simulations, models or carrying out real-time monitoring. As cities have different needs and challenges, these digital twins may focus on extreme weather events, urban planning or crisis management (see for instance AQMO – Air Quality and Mobility CEF project [21])
  - **Destination Earth project** [22] aims at being a unique digital modelling of the Earth designed to uplift Europe's ability to simulate environmental changes, predict extreme events and to anticipate and adapt EU actions and policies to climate-related challenges. Building on the EU investment in high performance computing, on the massive geo-spatial and socio-economic data sources at our

disposal and on the European excellence in data and AI technologies, Destination Earth will develop a very accurate replica of the Earth.

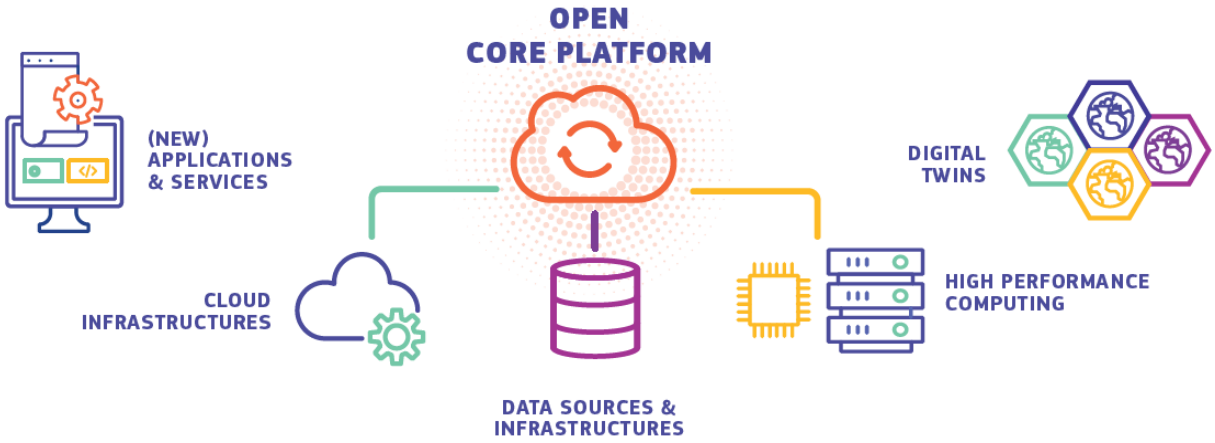

Figure 9: Destination Earth Digital © European Union, 2021

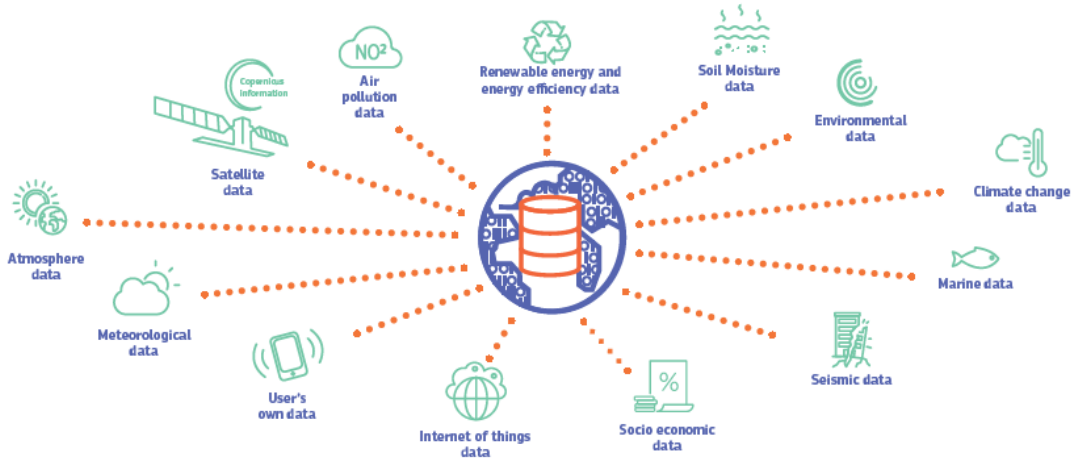

Figure 10: Destination Earth project datalake © European Union, 2021

These so-called "Digital Twins" will offer new ways to engage users with results of HPC simulation, with more interactivity but also with specific challenges (security, workflow, in situ visualization, etc.).

- Links with other infrastructures
  - The HPC Ecosystem is closely tied to other infrastructures in Europe and society. We describe this link through the help of the following diagram:

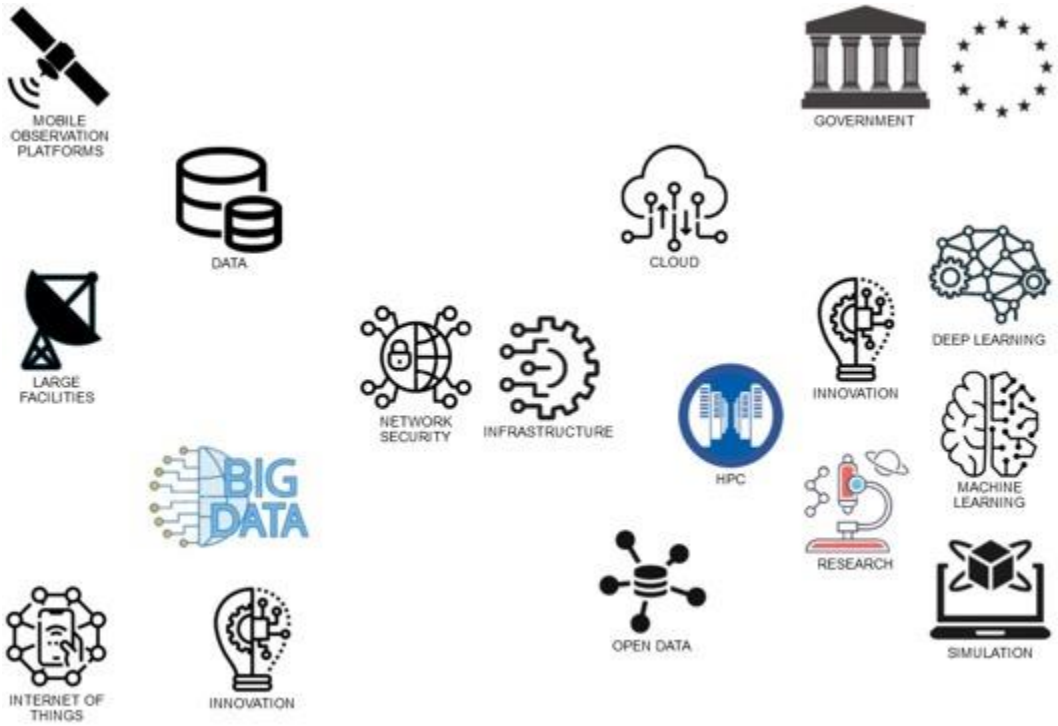

Figure 11: European infrastructure ecosystem

The HPC infrastructure is used for research and innovation purposes with applications in simulation, machine learning and deep learning. The European HPC infrastructure – which is provided by PRACE, is just one part of the **European integrated infrastructure** which encompasses a cloud infrastructure, a data and open data infrastructure. All of these are connected together over a network which is network secured against cyber threats. This network connects the infrastructure with edge devices and facilities which create a huge amount of data upon which research and innovation can take place. These edge devices include mobile devices and other smart devices (such as cell phones and sensors) owned by European citizens/industry/states – all of which fall under the categorisation of "Internet of Things". Other data creating facilities include large facilities – such as CERN and SKAO, and mobile observational platforms – such as observation satellites. All these fall under the authority and leadership of the European Union and Member States (governments) which support industry and academia to best serve the population and interests of society in Europe.

3 New Services for Industry

3.1 HPC services for industry

The HPC-landscape in Europe has been undergoing major changes over recent years. In the past, access to HPC went basically through a few regional and national centres which were to a large extent financed by governmental bodies, and in return were mainly used by publicly-funded organisations such as universities or research centres. Industry and indeed SMEs were almost absent in this user base, with companies for which HPC was part of their core competencies using almost only in-house HPC facilities (e.g. Oil and Gas, Air and Space, Automotive). Many of the national centres, if not all, were united under the umbrella of PRACE, with an architecture of Tier-0 (world class systems, resources offered through PRACE), Tier-1 (national systems, with resources exchanged among peers through the PRACE DECI access service [66]), and - not involved in PRACE - regional Tier-2 systems providing direct access

to academic users, and serving as a first step to enter the HPC ecosystem pyramid). The offer of access was free of charge at the point of use, but with the complexity needed to operate an access based only on the merit of excellence open science, through a world class level Peer-Review evaluation of proposals. This process is the best practice used to access large scale science facilities (e.g. CERN, astrophysics instruments, etc.). The arrival of cloud computing not only brought new players from the corporate sector into the arena but also enlarged the scope of technology with Big Data and Cloud. Furthermore, it fundamentally transformed the entire ecosystem by establishing vendors, users and strong policy-making bodies as fixed landmarks. As a consequence, new access points emerged. Among them, some were able to also offer High-Performance (cloud) Computing services, and not only 'regular' (cloud) Computing services. Where HPC could be defined as the highest capacity of its time, it also got some other characteristics that distinguish HPC from HTC, high performance from high capacity. Among them the capacity to perform (very) large parallel computing, which implies specific architecture of an HPC system such as high-speed interconnect, that are not used nor need inside a HTC centre. Most of the operational computing needs from the industry are HTC, with many computation taking place independently from each other (as for instance many end-users accessing the data base of a web marketplace, whereas to perform a numerical weather forecast the computation at each point of the grid representing the atmosphere is linked to all its neighbours).

It is arguably the entrance of multi-national heavyweights with their extensive and elaborated service portfolios of cloud-based offerings which had the biggest impact in this new marketplace. The main players in regards to revenue and market share are:

- Amazon (leveraging their own Data centre used for their core activity to offer cloud services);
- Microsoft;
- Alibaba;
- Google;
- IBM.

Their enormous growth rates are not only due to aggressive marketing strategies but also because these companies managed to re-define the service-function in software/application (the naming 'software' being more used in the market sector, the naming 'application' being more used in the academic sector, both referring to numerical computation). Software-as-a-Service (SaaS) has become the standard technological environment where complex analytics and big data-workloads are managed, in a standardised way that allow to decouple the use of the software service from the specific implementation of the platform on which it is performed. It has led to the creation of new service types, often used by the clients as single building blocks but also in a compatible and interconnected mix to maximise ROI:

- Cloud Business Process Services (BPaaS);
- Cloud Application Infrastructure Services (PaaS);
- Cloud Application Services (SaaS);
- Cloud Management and Security Services (CMSS);
- Cloud System Infrastructure Services (IaaS);
- New Cloud HPC Services (HPCaaS).

For the user, this diversity of services has an undeniable added value because it can experiment with different application designs in the most diverse business scenarios. In order to make this working approach even more attractive for their clients, the above-mentioned service providers eliminated the biggest burden which is 'investments' and replaced it with a user-friendly 'pay as you go-model' in numerous and often tailor-made variations. This outsourcing includes not only the capital expenditure on a Data Centre, but also the outsourcing of relatively rare high-tech talent, for operating and maintaining the Data Centre, but also for application knowledge

which are not seen as core business of many companies. For HPC these new services are offering the same attractiveness as for other cloud services users, along with the same limitations, but with some specificity inherent to the HPC, where skills are more rare than on mainstream IT, and more difficult to acquire as it requires access to (rare) HPC facilities to learn very specific skills related to the scalability of applications. Furthermore, it often requires at least some knowledge of the scientific field of these applications, to be able to provide efficient support. Many HPC specialists have a dual background, in sciences and in computer science, dual Master or dual PhD. Attracting creating and keeping such talents for providing consultancy and support is a particular strength of PRACE, while Cloud service providers do not offer any such services. There are many factors entering into account for the attractivity (or lack of attractivity) of career in STEM, IT, HPC, and their discussion goes beyond the scope of this deliverable, but we could shortly mention action toward the promotion of diversity and inclusivity and to bridge gender gap [27], and the fact that for such careers non-monetary parameters play an important role, such as the interest of participating to a scientific endeavor (fighting Covid19 and other diseases, modeling climate change, understanding dark matter or searching for clean energies of tomorrows), or the personal gratification that one could acquire from participating in the building of a European Research Infrastructures. Such non-monetary personal gratification could compensate other rewards, up to a certain point, and are also intricated with other factors, such as job security, which could be higher on private or in public sector, depending on the situations. To get back to services for industry, being able to rely on a persistent (at the scale of the project or product life-cycle) high-level team is an important factor.

An important drawback of a cloud HPC offer could be the availability of high level very specific services. In HPC, the performance came with a cost, which is not only related to the performance of the HPC centre capacity (measured in the potential number of floating-point operations per second, that result from the number of interconnected cores available within the HPC centre). Performing many parallel computations at a time requires an overhead of information exchange between them. The scalability of an application is never perfect, and the ways to minimise this burden is highly technical, and require a deep knowledge of the application algorithm structure AND of the HPC system on which it runs. This is the complete opposite of the standard business model of Software as a Service, where standard non-optimised software runs on a standard operating system on standard hardware architecture. Depending on the lifecycle of a digital application, the return on investment in optimising the code to obtain faster or more efficient applications, allowing some saving on the infrastructure bill could be low with regard to the addition cost of development. Depending on the sector, time to market could be more important than perfection of the product. Moreover, “standard” SaaS have a short lifespan, and “low” complexity, in comparison with “standard” HPC applications (for academic or for industrial use). Their lifecycle and sizes fit well with the lifecycle of the hardware on which they operate, new generations of software service being designed to work on their up-to-date new hardware, without much stickiness from their previous generations. The recent tendency to produce “low code” applications, by assembling black boxes in a very short cycle of try and errors and direct test to users, allowing very fast development results in a very different corporate culture than the traditional V cycle of design, development and test that is more common to big industry or academic applications. Code Saturne for instance, the free open-source code developed and released by EDF to solve computational fluid dynamics (CFD) applications (see Annexe 2 for a comprehensive description), has been developed in early 2000, based upon Code Mercure that have been designed by previous generation of EDF engineers and physicist in 90s. Its millions of lines of code and thousands of subroutines have little in common to a standard web-based service, and requires complex non-regression and validation tests at each update, in a very different manner from try and errors of “low code”. Some projects are aiming at bridging the gap between cloud and HPC, such as Lexis [25], which is building

an advanced engineering platform, leveraging large-scale geographically-distributed resources from the existing HPC infrastructure, employs Big Data analytics solutions and augments them with Cloud services.

The second limitation of a cloud-based HPC solution is inherent to the construction of its cost business model, the so-called 'pay as you go-model'. This model offers an attractive solution in comparison to in-house data-centre total cost of ownership, especially as it does not require to build up a data-centre to cope with peak need but not fully used most of the time. But for a more continuous high work-load, in-house solutions become financially more attractive given the current cost of hours offered by the main private operators.

A strong aspect of private offers of Cloud-HPC is that they are a one stop shop, offering a portfolio of services, allowing computing at (almost) any scale, providing support with service level agreements, and the basic fact that they are liable and persistent, which is not the case of EC funded projects of 2 to 3 years. This is very important when it came to investment decisions that will have impact on decades, this is an obvious reason why a Research Infrastructure is needed, with a strong brand, and not just a collection of projects put together with little coherence.

In this competitive environment of demand and best price offer, the number of available cores and computational power is a second layer which influences the decision of HPC-access. On a global level, Asia and the US largely outnumbered Europe. This shortcoming led the EC to introduce 'EuroHPC Joint Undertaking' along with the installation of several pre-exascale and petascale systems such as f.eg. 'LUMI' in Finland, 'Mare Nostrum 5' in Spain and 'Leonardo' in Italy.

There are six modes to access the machines:

- Extreme Scale Access;
- Regular Access;
- Benchmark Access;
- Development Access;
- Fast Track Access for Academia;
- Fast Track Access for Industry Access.

The table below describe in detail these six access mode.

Access Mode	**Extreme Scale**	**Regular**	**Benchmark**	**Development**	**Academic Fast Track**	**Industry Fast Track**
Duration	1 y renewable	1 y renewable	2 to 3 months	1 y renewable	< 6 months	1 y renewable
Periodicity	Continuous calls, bi-yearly cut-offs	Continuous calls, quarterly year cut-offs	Continuous calls, monthly cut-offs	Continuous calls, monthly cut-offs	Continuous calls, cut-offs ev. 2w/1m	Continuous calls, cut-offs ev. 2w/1m
Share of resources	70% Mostly pre-exascale	20 to 30% Mostly multi-exascale	Few % All systems	Few % All systems	5 % All systems	5 % All systems
Data storage needs	Large storage for medium to long term	Large storage for medium to long term	Limited	Data Processing environment and platform		
Accessible to industry	Yes - Open R&D with specific evaluation criteria	Yes - Open R&D with specific track	Yes - Open R&D	Yes - Open R&D	No - use industry Fast Track instead	Exclusively Open R&D

Access Mode	Extreme Scale	Regular	Benchmark	Development	Academic Fast Track	Industry Fast Track
External Sc. Peer-Review	Yes	Yes	No	No	No / Pre-identified	No / Pre-identified
Tech. assessment	Yes	Yes	Yes	Yes	Yes	Yes
Data Management Plan required	Yes	Yes	No	No	Yes	Yes
Application type	Full application	Full application	Technical application	Technical application	Light request + support documents	Full application
Prerequisite	Benchmark	Benchmark	None	None	Previous allocation on Benchmark	Benchmark
Submission period	> 2 months	> 2 months	N/A	N/A	N/A	N/A
Duration of evaluation process	3 months	2 months	≥ 1 week < 2 weeks	≥ 1 week < 2 weeks	≥ 2 weeks < 1 month	≥ 2 weeks < 1 month

Table 1: Access mode offered by PRACE to EuroHPC systems

As to the commercial access, 20 % of the computing resources are reserved for commercial purposes and offered on a pay-per-use basis. It is stipulated in the access regulations that pricing should be comparable to market prices and not disrupt the market.

3.2 Survey to Industry

The PRACE-6IP project performed a "Survey to Industry" in the frame of the task 2.1 new services for industry.

The survey ran for two months from the 14 September until 15 November 2021. In order to take advantage of the wide coverage of the CASTIEL/EuroCC projects and National Competence Centre (NCC) network across Europe, that share similar objectives, the survey has been sent to CASTIEL to be sent to NCC, to be distributed to their local contacts. The survey has also been distributed to the list of industrial and SMEs contact of PRACE, from the SHAPE initiative [23], and through the PRACE Industrial Advisory Committee [11]. We received 52 responses, from 18 countries (including Turkey (2), UK (4) and USA (3) outside European Union), Austria (3), Belgium (2), Bulgaria, Croatia (2), Denmark, Finland, France (8), Germany (6), Hungary, Italy (4), Montenegro, Netherlands (2), Portugal, Spain (4) and Sweden, but only 30 responses were complete and exploitable. 32 responses were from SMEs, 12 from large enterprises (over 250 employees) and 8 from associations. The format of the survey was a questionnaire that contains 20 short questions (the last 4 being optional) and takes on average about 20 min to complete. The rate of return was quite low, but one must take into account that it is difficult to reach Industries and SMEs on such matter, that many of them consider as strategic and sensitive.

The objective of this Survey was to gain further understanding of Industry's needs in the area of HPC, HPDA and AI as well as to gain insights into the trends and applications of innovative technologies relating to best practices in the areas of HPC, HPDA and AI.

The survey aimed at assessing in more detail the specific difficulties and issues faced by SMEs when attempting to apply High Performance Computing, Big Data Analysis and AI to their work and when navigating the maze of software covering virtually all industry areas and research related to them.

The outcome of this survey will help prepare a more detailed survey regarding industry's needs which will be undertaken by the HPC-GIG project.

HPC/HPDA/AI are modern paradigms that refer to computing activities and resources, far beyond the capabilities, capacity and productivity of mass computer technology that most companies, organisations and individuals work with. A brief introduction at the beginning of the Survey defines these paradigms, so that our audiences are fully aligned and access to wider, less experienced audiences, is facilitated too.

In order to shed some light on the content of the Survey and the basis for the 20 questions eventually selected, we elaborate further details below.

As a rule, SMEs, private laboratories and start-ups are poorly acquainted with the capabilities of supercomputers and cannot fully grasp the benefits of using them.

The Survey helps to guide them in which areas it is beneficial to use supercomputers in order to improve product quality, reduce time to delivery, or create innovative new services.

On the other hand, there are industrial areas in which the use of supercomputers is not necessary or beneficial. For example, food production, the garment industry, architecture and housing, household design and many more. The tasks required by those can be solved by personal computers or on several multiprocessor servers (in case the calculations are more complex). A convenient platform for them to use would be cloud computing.

These are some of the main areas identified to have a proven benefit of using supercomputers:

- Genetic engineering and agriculture selections;
- HPC modelling and simulations in biochemistry and biophysics, molecular biology, micro biology;
- Virtual drug discovery;
- Clinical trials and joint pharmaceutical and Artificial Intelligence predictions of drug side effects;
- Personal medicine, immunotherapy, gene analysis and genome malformation identifications;
- Computer aided synthesis and study of nanomaterials for hydrogen production, car catalyst, for water and air purifications, chemistry synthesis;
- High Performance Simulation of Turbulence and Large Eddy Flow;
- Application of computational fluid dynamics in drone and small aircraft design;
- High performance simulations of turbulent combustion in new generation engines;
- Real time planning and reliability of the electrical power generation and distribution systems;
- Simulation of 3D seismic wave propagation in any region of the Earth;
- FinTech and financial services;
- Smart cities projects: digital solutions.

These certainly do not exhaust the areas of effective application of HPC.

With these in mind we also wanted to make sure that the Survey informs businesses, private laboratories and institutes, whose business interests are in these areas, that if they use HPC they will improve their financial results and increase their competitiveness.

Of course, SMEs businesses generally do not have software engineers as staff, they would need some external professional assistance to help them select the appropriate software. There are such specialists at PRACE and EuroHPC Supercomputing Centres, Centers of Excellence or other institutes of the PRACE member states.

We would think that PRACE SHAPE should take on the role of SME assistant to put a group of support members together.

Lastly, please also refer to the "Applied Software Suite" for our attempt to facilitate access to available HPC Software. The most used HPC software repository is GitHub [67]. According to various publications, it stores between 300 and 320 high performance computing codes, created in Japan, China, India, Russia, the European Union, the United States and other countries. Each program has a sufficiently detailed description. From these descriptions you can relatively easily find out which software you should use to solve your specific task. PRACE have also developed a dedicated CodeVault repository of programs developed by organisations of PRACE member countries [68][69]. This guide to European software requires some constant work to be kept up to date, but all the programs in it are widely used today and through it you can find their latest modifications - i.e. latest software release. These have been a helpful resource to our work.

3.3 Applied Software Catalogue for Industry

The second pillar of PRACE services comprises the coordination activities undertaken, both in operations and technology watch, software development activities, knowledge compilation and their effective communication and dissemination. Among these services, PRACE has put in place a catalog to facilitate access to applied software available on PRACE HPC systems, open to all audiences, but having SMEs and Industry as the principal target. It is important to ensure that SMEs are both aware of the resources available to them as well as facilitating access to them. We have therefore put together an "HPC Applied Software Library", which is a comprehensive list of the ready-to-use software available on our HPC systems with a concise summary overview description added for each software. The library is classified based on Industry sectors and scientific disciplines in order to facilitate finding the appropriate software solution. This structured approach will allow new users who want to start using HPC, and notably Industry, a clear overview of the initial possibilities and how these could be applied to their specific challenges. Relevant training will be provided where possible or cross-referenced with existing events. The Library will be uploaded onto the HPC in Europe portal, further complementing the services already available there, a copy of its current version is provided in Annex 2 - HPC Software Suites widely used for supercomputer simulations and virtual experiments.

4 Conclusions

PRACE, the ESFRI landmark Research Infrastructure offers many links to the European ecosystem of its stakeholders, including Industry, building upon its many competences.

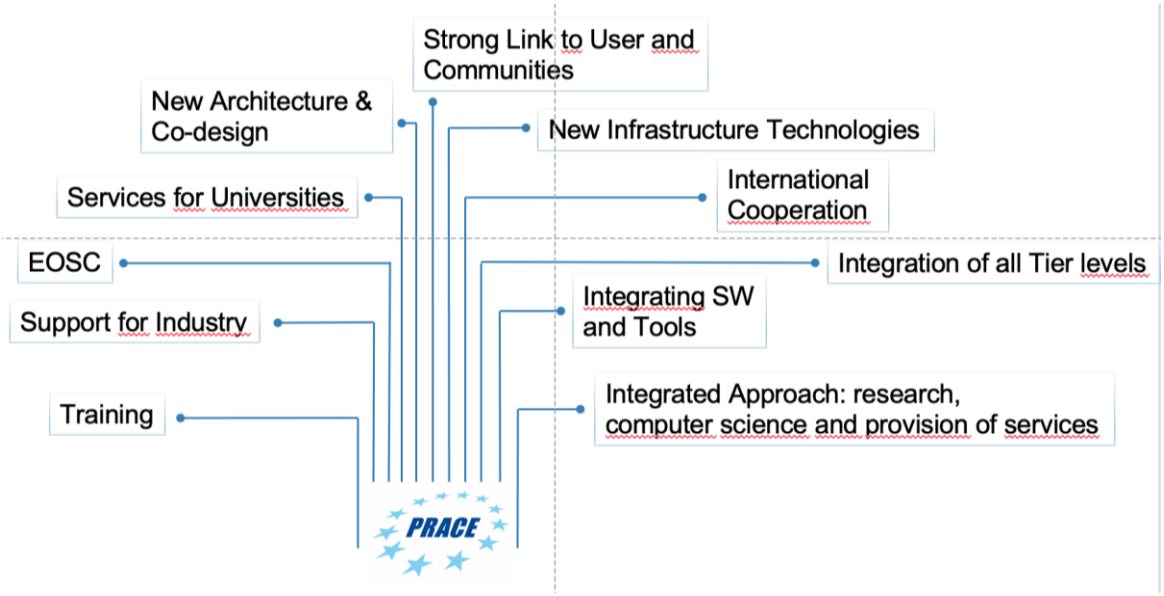

Figure 12: PRACE tree of competences, possible links to our Stakeholders

PRACE integrates the active involvement of User Communities to support their scientific discoveries and research developments. Through an open and transparent Governance, PRACE provides to these communities, jointly with HPC agencies and centres, the power to drive the PRACE activities. This also guarantees that principles underpinning scientific excellence are applied to the peer review processes distributing resources to the users.

PRACE elevates the competence of PRACE partners by coordinating and aligning the support for users of the European High-Performance Computing and Data Infrastructure. PRACE enables a continuous, aggregated collection of feedback and scientific roadmaps from users, and performs European-wide independent technology watch, and identifies new services to anticipate the needs of potential new users. The organisation provides a platform for national computing centres and HPC stakeholders to exchange practices and to discuss ways to improve the HPC services within a converged and evolving HPC, AI, and soon, quantum computing framework.

In this context, the mission of PRACE is to enable high-impact scientific discovery and research and development across all disciplines to enhance European competitiveness for the benefit of society.

PRACE seeks to realise its mission by pooling high-level services to foster Scientific Collaboration and Innovation, leveraging the power of the available European Computing and Data Infrastructure.

The organisation also maintains, and further develops international collaboration beyond Europe whenever it strengthens the development of the HPC services in Europe and benefits Scientific Collaboration.

5 Annex 1 - HPC Ecosystem References

5.1 List of active EC funded HPC projects (from [39])

ACROSS, www.acrossproject.eu
ADMIRE, www.admire-eurohpc.eu
ASPIDE, www.aspide-project.eu
BioExcel-2, bioexcel.eu/
CASTIEL, eurocc-access.eu
ChEESE, www.cheese-coe.eu
CoEC, www.coec-project.eu
CompBioMed2, www.compbiomed.eu
CYBELE, www.cybele-project.eu
DComEX, mgroup.ntua.gr/dcomex
DEEP-EST, www.deep-projects.eu
DeepHealth, deephealth-project.eu
DEEP-SEA, www.deep-projects.eu
E-cam, www.e-cam2020.eu
eFlows4HPC, eflows4hpc.eu
ENERXICO, www.enerxico-project.eu
EoCoE-II, www.eocoe.eu
EPEEC, www.epeec-project.eu
EPI, eprocessor.eu
EPiGRAM-HS, epigram-hs.eu
eProcessor, eprocessor.eu
ESCAPE-2, www.esiwace.eu
ESiWACE2, www.esiwace.eu
EuroCC, eurocc-access.eu
EuroExa, www.euroexa.eu
EVOLVE, www.evolve-h2020.eu
EXA MODE, www.examode.eu
EXA2PRO, exa2pro.eu/
exaFOAM, exafoam.eu
ExaQUte, www.exaqute.eu
EXCELLERAT, www.excellerat.eu
EXSCALATE4COV, exscalate4cov.eu
FF4EUROHPC, ff4eurohpc.eu
FocusCoE, www.hpccoe.eu
HEROES,
HiDALGO,
HiPEAC,
HPC-EUROPA3,
HPCWE,
IO-SEA,
LEXIS, https://lexis-project.eu/web/
LIGATE,
MAELSTROM,
MAESTRO,
MaX,
MEEP,
MICROCARD,
Mont-Blanc 2020,
NEASQC,
NextSim,
NOMAD,
OPTIMA,
PerMedCoE,
Phidias,
POP2,
RAISE,
RECIPE,
RED-SEA,
REGALE,
Sage2,
SCALABLE,
SODALITE,
SPARCITY,
TEXTAROSSA,
TIME-X,
TRex,
VECMA,
VESTEC,

5.2 List and link of the 15 HPC Centres of Excellence

BioExcel : https://www.hpccoe.eu/index.php/bioexcel/
ChEESE : https://www.hpccoe.eu/index.php/cheese/
CoEC : https://www.hpccoe.eu/index.php/coec/
CompBioMed : https://www.hpccoe.eu/index.php/compbiomed-2/
E-CAM : https://www.hpccoe.eu/index.php/e-cam-2/
EoCoE: https://www.hpccoe.eu/index.php/eocoe-2/
ESiWACE: https://www.hpccoe.eu/index.php/esiwace/
EXCELLERAT: https://www.hpccoe.eu/index.php/excellerat-2/
HiDALGO: https://www.hpccoe.eu/index.php/hidalgo-2/
MaX: https://www.hpccoe.eu/index.php/max2/
NOMAD: https://www.hpccoe.eu/index.php/nomad/
PerMedCoE: https://www.hpccoe.eu/index.php/permedcoe/
POP: https://www.hpccoe.eu/index.php/pop-2-2/
RAISE: https://www.hpccoe.eu/index.php/raise/
TREX: https://www.hpccoe.eu/index.php/trex-coe/

5.3 EC funded Exascale projects

Pillars	Corresponding calls	Comments
Technology: HW and SW building blocks for exascale	EuroHPC-01-2019 Extreme scale computing and data driven technologies	10 projects started in 2021, most of them with a duration of 3 years
	EuroHPC-06-2019 Advanced Experimental Platform To-wards Exascale	1 project started in 2020
	FETFLAG-05-2020 Complementary call on Quantum Computing	1 project started in 2020
	FETHPC-01-2016	2 projects started in 2017 and ending at the beginning of 2021
	FETHPC-02-2017 Transition to Exascale	11 projects started in 2018 with a duration of around 3 years
	ICT-05-2017 Customised and low energy computing (including Low power processor technologies)	Prequel to EPI
	ICT-16-2018 Software Technologies	1 project related to HPC
	ICT-42-2017 Framework Partnership Agreement in European low-power microprocessor technologies SGA-LPMT-01-2018 Specific Grant Agreement under the Framework Partnership Agreement "EPI FPA"	The European Processor Initiative (EPI)
Applications: Centres of Excellence in computing applications	E-INFRA-5-2015 Centres of Excellence for computing applications	1 remaining project ending at the beginning of 2021
	INFRAEDI-02-2018 Centres of Excellence for computing applications	9 projects and 1 Coordination and Support Action started in 2018
	INFRAEDI-05-2020 Centres of Excellence in exascale computing	5 projects started in 2020 and 2021 with a duration of around 3 years
Applications: in the HPC Continuum	CEF-TC-2018-5	1 project related to HPC
	EuroHPC-02-2019 HPC and data centric environments and application platforms	5 projects started in 2021
	EuroHPC-03-2019 Industrial software codes for extreme scale computing environments and applications	5 projects started in 2021
	FETHPC-01-2018 International Cooperation on HPC	2 projects started in June 2019
	ICT-11-2018-2019 Subtopic (a) HPC and Big Data enabled Large-scale Test-beds and Applications	4 Innovative Actions started in 2018
	ICT-12-2018-2020 Big Data technologies and extreme-scale analytics	1 project related to HPC
	SC1-PHE-CORONAVIRUS-2020 Advancing knowledge for the clinical and public health response to the 2019-nCoV epidemic	Special call
Ecosystem development	FETHPC-2-2014 HPC Ecosystem Development	2 Coordination and Support Actions
	FETHPC-03-2017 Exascale HPC ecosystem development	2 Coordination and Support Actions
	INFRAIA-01-2016-2017 Integrating Activities for Advanced Communities	HPC-EUROPA3
	EuroHPC-04-2019 Innovating and Widening the HPC use and skills base	2 projects starting in 2020
	EuroHPC-05-2019 Innovating and Widening the HPC use and skills base	1 project starting in 2020
	ICT-01-2019 Computing technologies and engineering methods for cyber-physical systems of systems	1 project related to HPC

Table 2: List of EC funded Exascale projects, by strategic pillars. From ETP4HPC HPC projects handbook.

6 Annex 2 - HPC Software Suites widely used for supercomputer simulations and virtual experiments

Bioinformatics and Genomics

1.ALLPATHS-LG
https://software.broadinstitute.org/allpaths-lg/blog/
National Human Genome Research Institute and National Institute of Allergy and Infectious Diseases
Massively parallel DNA sequencing technologies are revolutionizing genomics by making it possible to generate billions of relatively short (~100-base) sequence reads at very low cost. LLPATHS-LG implements the algorithm for genome assembly and its application to massively parallel DNA sequence data from the human and mouse genomes, generated on the Illumina platform. The resulting draft genome assemblies have good accuracy, short-range contiguity, long-range connectivity, and coverage of the genome. In particular, the base accuracy is high (≥99.95%) and the scaffold sizes (N50 size = 11.5 Mb for human and 7.2 Mb for mouse) approach those obtained with capillary-based sequencing.

2.BLAST
https://blast.ncbi.nlm.nih.gov/Blast.cgi
NCBI BLAST is Basic Local Alignment Search Tool is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold.

3.ClustalW2
https://www.ebi.ac.uk/Tools/msa/clustalw2/
EMBL-EBI
ClustalW2 is platform for multiple alignment of nucleic acid to a genomic DNA sequence and protein sequences, compares a sequence, allowing for introns and frame shifting errors. Parallel ClustalW2 software package to allow faster alignment of very large data sets and to increase alignment accuracy. This software calculates the best matches and aligns the sequences according to the identified similarities. The number of nucleotides are compared to the genomic DNA sequence with the number of nucleotides in 10^8.

4.BEAST2
https://beast.community/
http://www.beast2.org/
BEAST 2 is a cross-platform program for Bayesian phylogenetic analysis of molecular sequences. It estimates rooted, time-measured phylogenies using strict or relaxed molecular clock models. It can be used as a method of reconstructing phylogenies but is also a framework for testing evolutionary hypotheses without conditioning on a single tree topology. BEAST 2 uses Markov chain Monte Carlo (MCMC) to average over tree space, so that each tree is weighted proportional to its posterior probability. BEAST 2 includes a graphical user-interface for setting up standard analyses and a suit of programs for analysing the results.

5.BWA
http://bio-bwa.sourceforge.net/
BWA is a software package for mapping low-divergent sequences against a large reference genome, such as the human genome. It consists of three algorithms: BWA-backtrack, BWA-SW and BWA-MEM. The first algorithm is designed for Illumina sequence reads up to 100bp, while the rest two for longer sequences ranged from 70bp to 1Mbp. BWA-MEM and BWA-SW share similar features such as long-read support and split alignment, but BWA-MEM, which is the latest, is generally recommended for high-quality queries as it is faster and more accurate. BWA-MEM also has better performance than BWA-backtrack for 70-100bp Illumina reads.

6.FALCON
https://github.com/PacificBiosciences/FALCON
Pacific Biosciences

Falcon a set of tools for fast aligning long reads for consensus and assembly. The Falcon tool kit is a set of simple code collection which I use for studying efficient assembly algorithm for haploid and diploid genomes.

6. FreeBayes 1.3.1
https://mybiosoftware.com/freebayes-0-9-4-bayesian-genetic-variant-detector.html
FreeBayes is a Bayesian genetic variant detector designed to find small polymorphisms, specifically SNPs (single-nucleotide polymorphisms), indels (insertions and deletions), MNPs (multi-nucleotide polymorphisms), and complex events (composite insertion and substitution events) smaller than the length of a short-read sequencing alignment.

7.IQ-TREE
http://www.iqtree.org/
Large phylogenomics data sets require fast tree inference methods, especially for maximum-likelihood (ML) phylogenies. Fast programs exist, but due to inherent heuristics to find optimal trees, it is not clear whether the best tree is found. Thus, there is need for additional approaches that employ different search strategies to find ML trees and that are at the same time as fast as currently available ML programs. IQ-TREE is an efficient phylogenomic inference software designed for the reconstruction of trees by maximum likelihood and assessing branch supports. This fast and effective stochastic algorithm is widely used in molecular systematics and IQ-TREE has essentially greater results when compared to RAxML and PhyML in terms of likelihood. IQ-TREE found higher likelihoods between 62.2% and 87.1% of the studied alignments, thus efficiently exploring the tree-space. If we use the IQ-TREE stopping rule, RAxML and PhyML are faster in 75.7% and 47.1% of the DNA alignments and 42.2% and 100% of the protein alignments, respectively. However, the range of obtaining higher likelihoods with IQ-TREE improves to 73.3–97.1%. IQ-TREE is freely available at http://www.cibiv.at/software/iqtree.

8.MEMSAT-SVM
https://bio.tools/memsat-svm
http://bioinfadmin.cs.ucl.ac.uk/downloads/memsat-svm/
University College London
A method capable of automatically identifying pore-lining regions in transmembrane proteins from sequence information alone, which can then be used to determine the pore stoichiometry. MEMSAT-SVM provides a way to characterise pores in transmembrane proteins and may even provide a starting point for discovering novel routes of therapeutic intervention in a number of important diseases.

9.RAY
https://sourceforge.net/projects/denovoassembler/
Ray is a parallel de novo genome assemblies for parallel DNA sequencing. It uses the message passing interface (MPI) for passing messages, while assembling reads obtained with new sequencing technologies (Illumina, 454, SOLiD) for next-generation sequencing data.

10.Samtool
https://hpc.nih.gov/apps/samtools.html
The NIH HPC group
Samtools is a suite of programs for interacting and post-processing with high-throughput sequencing data and short DNA sequence read alignments in SAM format. The SAM generic format is used for storing large nucleotide sequence alignments. The tools provided support complex tasks, such as variant calling and alignment viewing as well as sorting, indexing, data extraction. Samtools is a suite of applications for processing high throughput sequencing data. Amtools is used for working with SAM, BAM, and CRAM files containing aligned sequences; Bcftools is used for working with BCF2, VCF, and gVCF files containing variant calls; htslib is a library for reading and writing these formats. Samtools and Bcftools are based on htslib and format conversion.

Computational Chemistry
11.NWChem
https://www.nwchem-sw.org/
Pacific Northwest National Laboratory, US Department of Energy

NWChem is the DOE flagship quantum chemistry - molecular mechanics code, which was designed from scratch to run on supercomputers. NWChem aims to provide its users with computational chemistry tools that are scalable both in their ability to treat large scientific computational chemistry problems efficiently, and in their use of available parallel computing resources from high-performance parallel supercomputers to conventional workstation clusters NWChem development strategy is focused on providing new and essential scientific capabilities to its users in the areas of kinetics and dynamics of chemical transformations, chemistry at interfaces and in the condensed phase, and enabling innovative and integrated research at EMSL. NWChem software can handle:

- Biomolecules, nanostructures, and solid-state;
- From quantum to classical, and all combinations;
- Ground and excited-states;
- Gaussian basis functions or plane-waves;
- Ab-initio molecular dynamics (Carr-Parinello);
- In general: single-point calculations, geometry optimizations, vibrational analysis;
- Extended (solid-state) systems DFT;
- Classical force-fields (Molecular Mechanics: AMBER, CHARMM, etc.).

Classical molecular dynamics capabilities provide for the simulation of macromolecules and solutions, including the computation of free energies using a variety of force fields. NWChem is scalable, both in its ability to treat large problems efficiently, and in its utilization of available parallel computing resources. NWChem has been optimized to perform calculations on large molecules using large parallel computers, and it is unique in this regard. This document is intended as an aid to chemists using the code for their own applications. Users are not expected to have a detailed understanding of the code internals, but some familiarity with the overall structure of the code.

12. Open Babel
http://openbabel.org/wiki/Main_Page
https://sourceforge.net/projects/openbabel/
Open Babel is a free, open-source version of the Babel chemistry file translation program. Open Babel is a project designed to pick up where Babel left off, as a cross-platform program and library designed to interconvert between many file formats used in molecular modeling, computational chemistry, and many related areas.

13.DIRAC
http://www.diracprogram.org/doku.php
DIRAC a relativistic ab initio electronic structure program for Atomic and Molecular Direct Iterative Relativistic All-electron Calculations. The DIRAC computes molecular properties using relativistic quantum chemical methods. It is named after P.A.M. Dirac, the father of relativistic electronic structure theory.

Molecular Dynamics, Molecular Mechanics and Molecular interactions
14.GROMACS
https://bioexcel.eu/
Centre of Excellence for Computational Biomolecular Research (BioExcel)
GROMACS is an engine to perform molecular dynamics simulations and energy minimization. These are two of the many techniques that belong to the realm of computational chemistry and molecular modeling. Molecular modeling indicates the general process of describing complex chemical systems in terms of a realistic atomic model, with the aim to understand and predict macroscopic properties based on detailed knowledge on an atomic scale. Often molecular modeling is used to design new materials, drugs, nano structure and atomic clusters for which the accurate prediction of physical properties of realistic systems is required. Macroscopic physical properties can be distinguished in (a) static equilibrium properties, such as the binding constant of an inhibitor to an enzyme, the average potential energy of a system, or the radial distribution function in a liquid, and (b) dynamic or non-equilibrium properties, such as the viscosity of a liquid, diffusion processes in membranes, the dynamics of phase changes, reaction kinetics, or the dynamics of defects in crystals. The choice of technique depends on the question asked and on the feasibility of the method to yield reliable results at the present state of the art.

15. NAMD

https://www.ks.uiuc.edu/Research/namd/
The Theoretical and Computational Biophysics Group at the University of Illinois
NAMD is a parallel, object-oriented molecular dynamics code designed for high-performance simulation of large biomolecular systems. Simulation preparation and analysis is integrated into the visualization package VMD. NAMD pioneered the use of hybrid spatial and force decomposition, a technique used by most scalable programs for biomolecular simulations, including Blue Matter. NAMD is developed using Charm++ and benefits from its adaptive communication-computation overlap and dynamic load balancing. We describe some recent optimizations including: pencil decomposition of the Particle Mesh Ewald method, reduction of memory footprint, and topology sensitive load balancing. Unlike most other MD programs, NAMD not only runs on a wide variety of platforms ranging from commodity clusters to supercomputers, but also scales to over one hundred thousand processors thousands. NAMD was tested 1.07 billion-atom complex benchmark up to 64,000 processors.

16. LAMMPS
https://www.lammps.org/
Sandia National Laboratories, Department of Energy,USA
LAMMPS is a classical molecular dynamics simulation code designed to run efficiently on parallel computers. It was developed at Sandia National Laboratories, a US Department of Energy facility, with funding from the DOE. In the most general sense, LAMMPS integrates Newton's equations of motion for collections of atoms, molecules, or macroscopic particles that interact via short - or long - range forces with a variety of initial and/or boundary conditions. On parallel machines, LAMMPS uses spatial decomposition techniques to partition the simulation domain into small 3d sub domains, one of which is assigned to each processor. LAMMPS is most efficient (in a parallel sense) for systems whose particles fill a 3d rectangular box with roughly uniform density. *Kinds of systems LAMMPS can simulate:*

- bead−spring polymers;
- united−atom polymers or organic molecules;
- all−atom polymers, organic molecules, proteins, DNA;
- metals;
- granular materials;
- coarse−grained mesoscale models;
- ellipsoidal particles;
- point dipolar particles;
- hybrid systems.

17. DL_POLY
https://www.scd.stfc.ac.uk/Pages/DL_POLY.aspx
DL POLY was developed at Daresbury Laboratory, the Science & Technology Facilities Council UK, with support from the Engineering and Physical Sciences Research Council the Natural Environment Research Council (NERC). DL_POLY is a general purpose classical molecular dynamics (MD) simulation software. The package is used to model the atomistic (coarse-grained or DPD) evolution of the full spectrum of models commonly employed in the materials sciences as solid-state chemistry, biological simulation and soft condensed-matter communities. Calculates molecular dynamics and solves problems of molecular mechanics for very large molecules containing up to 10^9 atoms. DL POLY is a fully data distributed code, employing methodologies such as spatial domain decomposition (DD), link-cells (LC) built Verlet neighbour. The particle density of the modelled systems close to uniform in space and time (ensuring load balancing).

18. GENESIS MD
https://www.r-ccs.riken.jp/labs/cbrt/
GENESIS (GENeralized-Ensemble SImulation System) mainly developed by the Sugita groups in RIKEN, Japan. (Computational Biophysics Research Team (R-CCS), Theoretical Molecular Science Laboratory (CPR), and Laboratory for biomolecular function simulation (BDR))
The GENESIS program package is composed of two MD programs (ATDYN, SPDYN) and trajectory analysis tools:

- CHARMM force field, AMBER force field, MARTINI model, and Go models;
- Energy minimization and molecular dynamics simulations;

- SHAKE/RATTLE, SETTLE, and LINCS algorithms for bond constraint Bussi;
- Langevin, and Berendsen thermostat/barostat;
- Replica-exchange molecular dynamics method (REMD) in temperature, pressure, and surface-tension space;
- Generalized replica-exchange with solute tempering (gREST) and replica-exchange umbrella sampling (REUS) with collective variables;
- Multi-dimensional REMD methods;
- Gaussian accelerated molecular dynamics method;
- String method for reaction pathway search;
- Hybrid quantum mechanics/molecular mechanics calculation (QM/MM);
- Implicit solvent model (Generalized Born/Solvent Accessible Surface Area model);
- Free-energy perturbation method (FEP);
- An harmonic vibrational analysis using SINDO;
- Steered MD and Targeted MD simulations;
- Restrained MD simulations (Distance, angle, dihedral angle, position, etc);
- Hybrid MPI+OpenMP, hybrid CPU+GPGPU, mixed double+single precision calculations;
- Scalable MD simulations for huge systems > 100,000,000 atoms;
- Spatial decomposition analysis (SPANA).

19. ls1 mardyn:
https://www.ls1-mardyn.de/home.html
The molecular dynamics (MD) simulation program ls1 mardyn mainly developed by High Performance Computing Center Stuttgart; Leibniz Supercomputing Centre, Scientific Computing in Computer Science; University of Kaiserslautern, Laboratory of Engineering Thermodynamics.
ls1 mardyn was optimized for massively parallel execution on supercomputing architectures. With an efficient MD simulation engine, explicit particle-based force-field models of the intermolecular interactions can be applied to length and time scales which were previously out of scope for molecular methods. Employing a dynamic load balancing scheme for an adaptable volume decomposition, *ls1 mardyn* delivers a high performance even for challenging heterogeneous configurations. The program is an interdisciplinary endeavor, whose contributors have backgrounds from engineering, computer science and physics, aiming at studying challenging scenarios with up to trillions of molecules. In the considered systems, the spatial distribution of the molecules may be heterogeneous and subject to rapid unpredictable change. This is reflected by the algorithms and data structures as well as a highly modular software engineering approach. It is more specialized than most of the molecular simulation programs mentioned above. In particular, it is restricted to rigid molecules, and only constant volume ensembles are supported, so that the pressure cannot be specified in advance. Electrostatic long-range interactions, beyond the cut-off radius, are considered by the reaction field method, which cannot be applied to systems containing ions. However, ls1 mardyn is highly performant and scalable. Holding the present world record in simulated system size, it is furthermore characterized by a modular structure, facilitating a high degree of flexibility within a single code base. Thus, ls1 mardyn is not only a simulation engine, but also a framework for developing and evaluating simulation algorithms, e.g. different thermostats or parallelization schemes.

Quantum Chemistry

20. CPMD (Car-Parrinello Molecular Dynamics).
https://bioexcel.eu/software/cpmd/
Physical Chemistry Institute of the Zurich University; Max Planck Institute, Stuttgart; IBM Research Laboratory, Zurich and Centre of Excellence for Computational Biomolecular Research (Bioexcel).
The CPMD code is a parallelized plane wave/pseudopotential implementation of Density Functional Theory, particularly designed for ab-initio molecular dynamics. CPMD is currently the most HPC efficient code that allows performing quantum molecular dynamics simulations by using the Car-Parrinello molecular dynamics scheme. CPMD simulations are usually restricted to systems of few hundred atoms.
The main characteristics of the CPMD code include:

- works with norm conserving or ultrasoft pseudopotentials;
- free energy density functional implementation;
- isolated systems and system with periodic boundary conditions; k-points;

- molecular and crystal symmetry;
- wavefunction optimization: direct minimization and diagonalization;
- geometry optimization: local optimization and simulated annealing;
- molecular dynamics: constant energy, constant temperature and constant pressure;
- - molecular dynamics: NVE, NVT, NPT ensembles;
- response functions and many electronic structure properties;
- time-dependent Density Functional Theory (excitations, molecular dynamics in excited states);
- Hybrid quantum mechanical / molecular mechanics calculations (QM/MM).

21. CP2K
https://www.cp2k.org/
The CP2K Foundation; University of Zurich, Computational Chemistry; Paul Scherrer Institute
CP2K is a quantum chemistry and solid-state physics software package that can perform atomistic simulations of solid state, liquid, molecular, periodic, material, crystal, and biological systems. It is especially aimed at massively parallel and linear-scaling electronic structure methods and state-of-the-art ab initio molecular dynamics simulations. Excellent performance for electronic structure calculations is achieved using novel algorithms implemented for modern high-performance computing systems. CP2K is a suite of modules, collecting a variety of molecular simulation methods at different levels of accuracy, from ab-initio density functional theory (DFT) to classical Hamiltonians, passing through semi-empirical neglect of diatomic differential overlap (NDDO) approximation It is used routinely for predicting energies, molecular structures, vibrational frequencies of molecular systems, reaction mechanisms, and ideally suited for performing molecular dynamics studies. CP2K can do simulations of molecular dynamics, metadynamics, Monte Carlo, Ehrenfest dynamics, vibrational analysis, core level spectroscopy, energy minimization, and transition state optimization using NEB or dimer method.

22.Quantum Espresso
https://www.quantum-espresso.org/
Quantum ESPRESSO Foundation, Centre of Excellence MaX (MAterials design at the eXascale)
Quantum Espresso is an integrated suite of computer codes for electronic-structure calculations and materials modeling at the nanoscale. It is based on density-functional theory, plane waves, and pseudo potentials. The core plan wave DFT functions of QE are provided by the Plane-Wave Self-Consistent Field (Wscf) component. Quantum Espresso can currently perform the following kinds of calculations:

- ground-state energy and one-electron (Kohn-Sham) orbitals;
- atomic forces, stresses, and structural optimization;
- molecular dynamics on the ground-state Born-Oppenheimer surface, also with variable cell;
- Nudged Elastic Band (NEB) and Fourier String Method Dynamics (SMD) for energy barriers and reaction paths;
- macroscopic polarization and finite electric fields via the modern theory;
- of polarization (Berry Phases).

23. General Atomic and Molecular Electronic Structure System (GAMES)
https://www.msg.chem.iastate.edu/gamess/
Mark Gordon's Quantum Theory Group, Ames Laboratory, Iowa State University
GAMESS is a program for ab-initio molecular quantum chemistry. Briefly, GAMESS can compute SCF wave functions ranging from RHF, ROHF, UHF, GVB, and MCSCF. Correlation corrections to these SCF wave functions include Configuration Interaction, second order perturbation Theory, and Coupled-Cluster approaches, as well as the Density Functional Theory approximation. Excited states can be computed by CI, EOM, or TD-DFT procedures.

Supercomputing atomistic simulation tools for understanding structure-property relation of nanomaterial

23. FLEUR
http://www.max-centre.eu/codes-max/fleur
FLEUR is mainly developed at the Forschungszentrum Jülich at the Institute of Advanced Simulation and the Peter Grünberg Institut. Full-potential Linearised augmented plane wave in EURope (FLEUR) is a code family for calculating groundstate as well as excited-state properties of solids within the context of density functional theory (DFT). The Fleur code implements the all-electron full-potential linearized augmented-plane-wave (FLAPW) approach to density functional theory (DFT). It allows the calculation

of properties obtainable by DFT for crystals and thin films composed of arbitrary chemical elements. For this it treats all electrons on the basis of DFT and does not rely on the pseudopotential approximation. There are also no shape approximations to the potential required. However, this comes at the cost of complex parametrizations of the calculations. The Fleur approach to this complex parametrization is the usage of an input generator that itself only requires basic structural input. Using this it generates a completely parametrized Fleur input file with material adapted default parameters.

24. BigDFT
http://www.max-centre.eu/codes-max/bigdft
European Centre of Excellence MAX
BigDFT is an electronic structure pseudopotential code that employs Daubechies wavelets as a computational basis, designed for usage on massively parallel architectures. It features high-precision cubic-scaling DFT functionalities enabling treatment of molecular, slab-like as well as extended systems, and efficiently supports hardware accelerators such as GPUs since 2009. Also, it features a linear-scaling algorithm that employs adaptive Support Functions (generalized Wannier orbitals) enabling the treatment of system of many thousand atoms. The code is developed and released as a software suite made of independent, interoperable components, some of which have already been linked and distributed in other DFT codes. BigDFT is fast, precise, and flexible code for ab-initio atomistic simulation.

Virtual Drug Discovery

Force field and Free Energy Cakculations

25. AMBER
https://ambermd.org/
https://ambermd.org/AmberModels.php
The term "Amber" refers to two things. First, it is a set of molecular mechanical force fields.
Amber is designed to work with several simple types of force fields, although it is most commonly used with parametrizations developed by Peter Kollman and his co-workers and “descendents”. The traditional parametrization uses fixed partial charges, centered on atoms. Less commonly used modifications add polarizable dipoles to atoms, so that the charge description depends upon the environment; such potentials are called “polarizable” or “non-additive”. An alternative is to use force fields originally developed for the CHARMM or Tinker (AMOEBA) codes. Since various choices make good sense, as of Amber 16 we have implemented a new scheme for users to specify the force fields they wish to use. Depending on what components are in your system, you may need to specify for the simulation of biomolecules (these force fields are in the public domain, and are used in a variety of simulation programs):

- a protein force field;
- a DNA force field;
- an RNA force field;
- a carbohydrate force field;
- a lipid force field;
- a water model with associated atomic ions (more variable, but the most common choice is still tip3p);
- a general force field, for organic molecules like ligands;
- other components (such as modified amino acids or nucleotides, other ions).

Second, AmberTool21 molecular simulation programs
https://ambermd.org/AmberTools.php
AmberTools 21consists of several independently developed packages that work well by themselves, and with Amber20 itself. The suite can also be used to carry out complete molecular dynamics simulations, with either explicit water or generalized Born solvent models. The AmberTools suite is free of charge, and its components are mostly released under the GNU General Public License (GPL). A few components are included that are in the public domain or which have other, open-source, licenses. The sander program has the LGPL license.

26. CHARMM
https://www.charmm.org/
Chemistry at HARvard Macromolecular Mechanics (CHARMM)

CHARMM is actively maintained by a large group of developers led by Martin Karplus. A molecular simulation program with broad application to many-particle systems with a comprehensive set of energy functions, a variety of enhanced sampling methods, and support for multi-scale techniques including Quantum Mechanics/Molecular Mechanics Quantum (QM/MM), hybrid molecular mechanics/coarse-grained (MM/CG) simulation, and a range of implicit solvent models. CHARMM primarily targets biological systems including peptides, proteins, prosthetic groups, small molecule ligands, nucleic acids, lipids, and carbohydrates, as they occur in solution, crystals, and membrane environments. CHARMM also finds broad applications for inorganic materials with applications in materials design. CHARMM contains a comprehensive set of analysis and model building tools. CHARMM achieves high performance on a variety of platforms including parallel clusters and GPUs.
Modules:

- Coordinate manipulation and analysis | corman;
- Energy commands | energy;
- Non-bonded options | nbonds;
- Minimization | minimiz;
- Molecular dynamics | dynamc;
- Constraints and restraints | cons;
- Time series and correlation functions | correl;
- Atom selections | select;

Modules in alphabetical order.
https://academiccharmm.org/documentation

Massively parallel ligand screening

27. VirtualFlow
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Harvard University, Boston,USA; Department of Physics, Faculty of Arts and Sciences, Harvard University, Boston; Department of Cancer Biology, Dana-Farber Cancer, Boston, USA; Institute; Department of Pharmacy, Pharmaceutical and Medicinal Chemistry, Saarland University, Saarbrücken, Germany; Enamine, National Taras Shevchenko University of Kyiv, Kyiv, Ukraine; Zuse Institute Berlin, Berlin, Germany; Institute of Mathematics, Technical University Berlin, Berlin, Germany; Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany;
https://github.com/VirtualFlow/VFVS
On average, an approved drug currently costs US$2–3 billion and takes more than
10 years to develop. In part, this is due to expensive and time-consuming Wet-laboratory experiments, poor initial hit compounds and the high attrition rates in the (pre-) clinical phases. Structure-based virtual screening has the potential to mitigate these problems. With structure-based virtual screening, the quality of the hits improves with the number of compounds screened. However, despite the fact that large databases of compounds exist, the ability to carry out large-scale structure-based virtual screening on supercomputers in an accessible, efficient and flexible manner has remained difficult. Virtual Flow (VF) is highly automated and versatile open-source drug discovery platform with perfect scaling behavior that is able to prepare and efficiently screen ultra-large libraries of compounds. VF is able to use a variety of the most powerful docking programs. Using VF, we prepared one of the largest and freely available ready-to-dock ligand libraries, with more than 1.4 billion commercially available molecules. Screening 1 billion compounds on a single processor core, with an average docking time of 15 s per ligand, would take approximately 475 years. Virtual Flow can dock 1 billion compounds in Approximately 2 weeks when leveraging 100,000 CPU cores simultaneously. VFLP prepares ligand databases by converting them from the SMILES format to any desired target format (for example, the PDBQT format, which is required by many of the AutoDock-based docking programs). VFLP uses the JChem package of ChemAxon as well as Open Babel to desalt ligands, neutralize them, generate (one or multiple) tautomeric states, compute protonation states at specific pH values, calculate three-dimensional coordinates and convert the molecules into desired target formats.
Preparation of the Enamine REAL library
One of the largest vendor libraries that is currently available is the REAL library of Enamine, which contains approximately 1.4 billion make-on-demand compounds.
https://enamine.net/
https://enamine.net/library-synthesis/real-compounds/real-database

The ZINC 15 database contained 1.46 billion compounds, but only provided 630 million molecules in a ready-to-dock format.
https://zinc15.docking.org/
The entire database has a six-dimensional lattice architecture, the general concept of which was modelled after the ZINC 15 database, in which each dimension corresponds to a physico-chemical property of the compounds (molecular mass, partition coefficient, number of hydrogen bond donors and acceptors, number of rotatable bonds and the topological polar surface area).

28. EXSCALATE
E4C is a public-private consortium supported by the European Commission's. The E4C consortium, coordinated by Dompé Farmaceutici , is composed by 18 institutions from seven European countries.
https://www.exscalate4cov.eu/
EXSCALATE a Ultra High Performance Virtual Screening Platform for computer aided drug design (CADD), based on LiGen, an exascale software able to screen billion of compounds in a very short time, and a library of trillion of compounds. Since 2010, Dompé SpA, has invested in a proprietary software for computer aided drug design (CADD), through its dedicated Drug Discovery Platform. The most relevant tool is the de novo structure-based virtual screening software LiGen(Ligand Generator), co-designed in collaboration with the Italian super computer center, CINECA. The distinguishing feature of LiGen is that it has been designed and developed to run on High-Performance Computing (HPC) architectures. To maintain the performance primate beyond 2020, LiGen is tools for molecular de novo design are actively sought incorporating sets of chemical rules for fast and efficient identification of structurally new chemotypes endowed with a desired set of biological properties. He is tools for molecular de novo design are actively sought incorporating sets of chemical rules for fast and efficient identification of structurally new chemotypes endowed with a desired set of biological properties. In its standard application, LiGen modules are used to define input constraints, either structure-based, through active site identification, or ligand-based, through pharmacophore definition, to docking and to de novo generation. Alternatively, individual modules can be combined in a user-defined manner to generate project-centric workflows. Specific features of LiGen are the use of a pharmacophore-based docking procedure which allows flexible docking without conformer enumeration and accurate and flexible reactant mapping coupled with reactant tagging through substructure searching.

Biological Target – Ligand Docking

29. High Ambiguity Driven protein-protein DOCKing (HADDOCK)
Centre of Excellence for Computational Biomolecular Research (Bioexcel).
https://bioexcel.eu/software/haddock/
HADDOCK is a versatile information-driven flexible docking approach for the modelling of biomolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods in the fact that it can integrate information derived from biochemical, biophysical or bioinformatics methods to enhance sampling, scoring, or both. The information that can be integrated is quite diverse: interface restraints from NMR or MS, mutagenesis experiments, or bioinformatics predictions; various orientational restraints from NMR and, recently, cryo-electron maps.

30.DOCK 6.9
Department of Pharmaceutical Chemistry, University of California, San Francisco .USA
http://dock.compbio.ucsf.edu/Online_Licensing/index.htm
DOCK 6.9 algorithm addressed rigid body docking using a geometric matching algorithm to superimpose the ligand onto a negative image of the binding pocket. Important features that improved the algorithm's ability to find the lowest-energy binding mode, including force-field based scoring, on-the-fly optimization, an improved matching algorithm for rigid body docking and an algorithm for flexible ligand docking. DOCK 6 include Delphi electrostatics, ligand conformational entropy corrections, ligand desolvation, receptor desolvation; Hawkins-Cramer-Truhlar GB/SA solvation scoring with optional salt screening; PB/SA solvation scoring; and AMBER scoring-including receptor flexibility, the full AMBER molecular mechanics scoring function with implicit solvent, conjugate gradient minimization, and molecular dynamics simulation capabilities.

31. VinaL.C.
Lawrence Livermore National Laboratory, Department of Energy, USA
https://github.com/XiaohuaZhangLLNL/VinaLC

A mixed parallel scheme that combines message passing interface (MPI) and multithreading was implemented in the AutoDock Vina molecular docking program. The resulting program, named VinaLC, was tested on the petascale high performance computing (HPC) machines at Lawrence Livermore National Laboratory. Parallel performance analysis of the VinaLC program shows that the code scales up to more than 15K CPUs with a very low overhead cost of 3.94%. One million flexible compound docking calculations took only 1.4 h to finish on about 15K CPUs. The docking accuracy of VinaLC has been validated against the DUD data set by the re-docking of X-ray ligands and an enrichment study, 64.4% of the top scoring poses have RMSD values under 2.0 Å. The program has been demonstrated to have good enrichment performance on 70% of the targets in the DUD data set. An analysis of the enrichment factors calculated at various percentages of the screening database indicates VinaLC has very good early recovery of actives.

32. SwissDOCK
Swiss Institute of Bioinformatics
http://www.swissdock.ch/
SwissDock is based on the docking software EADock DSS, whose algorithm consists of the following steps:

- many binding modes are generated either in a box (local docking) or in the vicinity of all target cavities (blind docking);
- simultaneously, their CHARMM energies are estimated on a grid;
- the binding modes with the most favorable energies are evaluated with FACTS, and clustered;
- the most favorable clusters can be visualized online and downloaded on your computer.

SwissDock, a web service to predict the molecular interactions that may occur between a target protein and a small molecule.
S3DB, a database of manually curated target and ligand structures, inspired by the Ligand-Protein Database.

3D Protein Structure Prediction
33. High - Resolution Protein Structure Prediction Codes (ROSETTA 3)
Hughes Institute, University of Washington.
https://www.rosettacommons.org/
ROSETTA is a library based object-oriented software suite which provides a robust system for predicting and designing protein structures, protein folding mechanisms, and protein-protein interactions. The Rosetta3 codes have been successful in the Critical Assessment of Techniques for Protein Structure Prediction (CASP7) competitions. The Roseta3 method uses a two-phase Monte Carlo algorithm to sample the extremely large space of possible structures in order to find the most favorable one. The first phase generates a low-resolution model of the protein backbone atoms while approximating the side chains with a single dummy atom. The high-resolution phase then uses a more realistic model of the full protein, along with the corresponding interactions, to find the best candidate for the native structure. The library contains the various tools, such as Atom, ResidueType, Residue, Conformation, Pose, ScoreFunction, ScoreType, and so forth. These components provide the data and services Rosetta uses to carry out its computations.
Rosetta Functionality Summary:

- Rosetta Abinitio: Performs de novo protein structure prediction;
- Rosetta Design: Low free energy sequences for target protein backbones;
- Rosetta Design pymol plugin: A user-friendly interface for submitting Protein Design simulations using Rosetta Design;
- Rosetta Dock: Predicts the structure of a protein-protein complex from the individual structures of the monomer components;
- Rosetta Antibody: Predicts antibody Fv region structures and performs antibody-antigen docking;
- Rosetta Fragments Generates: Fragment libraries for use by Rosetta ab initio in building protein structures;
- Rosetta NMR: Incorporates NMR data into the basic Rosetta protocol to accelerate the process of NMR structure prediction;
- Rosetta DNA: For the design of proteins that interact with specified DNA sequences;
- Rosetta RNA: Fragment assembly of RNA;

- Rosetta Ligand: For small molecule - protein docking.

Seismic Wave Impact Simulation

34. SPECFEM3D - seismic wave propagation
https://geodynamics.org/cig/software/specfem3d/
California Institute of Technology, USA; University of Pau, France.
SPECFEM3D is Computational Infrastructure for Geodynamics. Unstructured hexahedral mesh generation is a critical part of the modeling process in the Spectral-Element Method (SEM). We present some examples of seismic wave propagation in complex geological models, automatically meshed on a parallel machine based upon CUBIT (Sandia Laboratory), an advanced 3D unstructured hexahedral mesh generator that offers new opportunities for seismologist to design, assess, and improve the quality of a mesh in terms of both geometrical and numerical accuracy. The main goal is to provide useful tools for understanding seismic phenomena due to surface topography and subsurface structures such as low wave-speed sedimentary basins.

35. SEISSOL.
http://www.seissol.org/
Software
https://github.com/SeisSol/SeisSol
SeisSol is a software package for simulating wave propagation and dynamic rupture based on the arbitrary high-order accurate derivative discontinuous Galerkin method (ADER-DG). Computational earthquake dynamics is emerging as a key component in physics-based approaches to strong motion prediction for seismic hazard assessment and in physically constrained inversion approaches to earthquake source imaging from seismological and geodetic observations. Typical applications in both areas require the ability to deal with rupture surfaces of complicated, realistic geometries with high computational efficiency. In our implementation, tetrahedral elements are used which allows for a better fit of the geometrical constraints of the problem, i.e., the fault shape, and for an easy control of the variation of element sizes using smooth refining and coarsening strategies. Characteristics of the SeisSol simulation software are:

- use of tetrahedral meshes to approximate complex 3D model geometries and rapid model generation;
- use of elastic, viscoelastic and viscoplastic material to approximate realistic geological subsurface properties;
- use of arbitrarily high approximation order in time and space to produce reliable and sufficiently accurate synthetic seismograms or other seismological data sets.

Computational Fluid Dynamics

36. Code Saturne
Électricité de France
Code_Saturne is the free, open-source software developed and released by EDF to solve computational fluid dynamics (CFD) applications.
https://www.code-saturne.org/cms/
It solves the Navier-Stokes equations for 2D, 2D-axisymmetric and 3D flows, steady or unsteady, laminar or turbulent, incompressible or weakly dilatable, isothermal or not, with scalars transport if required. Several turbulence models are available, from Reynolds-Averaged models to Large-Eddy Simulation models.
Physical modelling:

- Laminar and turbulent flows;
- Compressible flow;
- Radiative heat transfer;
- Conjugate heat transfer;
- Combustion coal, fuel, gas;
- Electric arc and Joule effect;
- Lagrangian module for dispersed particle tracking;
- ALE method for deformable meshes;
- Specific engineering modules for nuclear waste surface storage and cooling towers;
- Derived version for atmospheric flows;

- Derived version for eulerian multiphase flows;
- Lagrangian method - stochastic modelling with 2-way coupling (momentum, heat, mass);
- Transport and deposit of droplets, ashes, coal, corrosion products, radioactive particles, chemical forces;
- Gas combustion;
- Coal combustion.

37. OpenFOAM
The OpenFOAM Foundation
https://www.openfoam.com/code
https://www.openfoam.com/download/
https://www.openfoam.com/documentation/tutorial-guide
OpenFOAM (Open source Field Operation And Manipulation) is a C++ toolbox for the development of customized numerical solvers, and pre-/post-processing utilities for the solution of continuum mechanics problems, including computational fluid dynamics (CFD). It has a large user base across most areas of engineering and science, from both commercial and academic organizations. OpenFOAM has an extensive range of features to solve anything from complex fluid flows involving chemical reactions, turbulence and heat transfer, to solid dynamics and electromagnetics.

38. Delft3D
Delft3D Open Source Community
https://oss.deltares.nl/web/delft3d/download
The Delft3D Flexible Mesh Suite (Delft3D FM) allows you to simulate the interaction of water, sediment, ecology, and water quality in time and space. The suite is mostly used for the modelling of natural environments like coastal, estuarine, lakes and river areas, but it is equally suitable for more artificial environments like harbours, locks, urban areas, etc. Delft3D FM consists of a number of well-tested and validated modules, which are linked to and integrated with each other. Delft3D is a integrated modelling suite, which simulates two-dimensional (in either the horizontal or a vertical plane) and three-dimensional flow, sediment transport and morphology, waves, water quality and ecology and is capable of handling the interactions between these processes. The suite is designed for use by domain experts and non-experts alike, which may range from consultants and engineers or contractors, to regulators and government officials, all of whom are active in one or more of the stages of the design, implementation and management cycle. As a second option we tried to use a communication-hiding conjugate gradient method, PETSc's linear solver KSPPIPECG, to solve the linear system arising from the spatial discretisation, but we were not able to get any performance. Currently the full source code is available of the Delft3D-FLOW (including morphology), Delft3D-WAVE, DELWAQ (D-Water Quality and D-Ecology) and PART (D-Particle Tracking) engines under GPLv3 conditions.

Massively parallel Large Eddy Simulations (LES) and Direct Numerical Simulation (DES) for the study of complex flows

https://coec-project.eu/references-codes/
39. CIAO
https://www.fz-juelich.de/ias/jsc/EN/Expertise/High-Q-Club/CIAO/_node.html
Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)
Compressible/Incompressible Advanced reactive turbulent simulations with Overset.
CIAO performs Direct Numerical Simulations (DNS) as well as Large-Eddy Simulations (LES) of the Navier-Stokes equations along with multiphysics effects (multiphase, combustion, soot, spark). It is a structured, finite difference code, which enables the coupling of multiple domains and their simultaneous computation. Moving meshes are supported and overset meshes can be used for local mesh refinement. A fully compressible as well as an incompressible/low-Mach solver are available within the code framework. Spatial and temporal staggering of flow variables are used in order to increase the accuracy of stencils. The sub-filter model for the momentum equations is an eddy viscosity concept in form of the dynamic Smagorinsky model with Lagrangian averaging along fluid particle trajectories. While the fully compressible solver uses equation of states or tabulated fluid properties, a transport equation for internal/total energy, and a low-storage five-stage, explicit Runge-Kutta method for time integration, the incompressible/low-Mach solver uses Crank-Nicolson time advancement and an

iterative predictor corrector scheme. The resulting Poisson equation for pressure is solved by HYPRE's multi-grid solver.

40. Alya
https://www.bsc.es/research-development/research-areas/engineering-simulations/alya-high-performance-computational
Barcelona Supercomputing Centre
Alya is a high-performance computational mechanics code to solve complex coupled multi-physics / multi-scale / multi-domain problems, which are mostly coming from the engineering realm. Among the different physics solved by Alya we can mention: incompressible/compressible flows, non-linear solid mechanics, chemistry, particle transport, heat transfer, turbulence modeling, electrical propagation, etc. From scratch, Alya was specially designed for massively parallel supercomputers, and the parallelization embraces four levels of the computer hierarchy:

- A substructuring technique with MPI as the message passing library is used for distributed memory supercomputers;
- At the node level, both loop and task parallelisms are considered using OpenMP as an alternative to MPI. Dynamic load balance techniques have been introduced as well to better exploit computational resources at the node level.
- At the CPU level, some kernels are also designed to enable vectorization.
- Finally, accelerators like GPU are also exploited through OpenACC pragmas or with CUDA to further enhance the performance of the code on heterogeneous computers.

Multiphysics coupling is achieved following a multi-code strategy, relating different instances of Alya. MPI is used to communicate between the different instances, where each instance solves a particular physics. This powerful technique enables asynchronous execution of the different physics.

41. AVBP
https://cerfacs.fr/en/computational-fluid-dynamics-softwares/
CERFACS (Centre de recherche fondamentale et appliquée spécialisé dans la modélisation et la simulation numériques, également centre de formation avancée).
AVBP is a LES (Large Eddy Simulation) code dedicated to unsteady compressible flows in complex geometries with combustion or without combustion. It is applied to combustion chambers, turbo machinery, safety analysis, optimization of combustors, pollutant formation (CO, NO, soot), UQ analysis. AVBP uses a high-order Taylor Galerkin scheme on hybrid meshes for multi species perfect of real gases. Its spatial accuracy on unstructured hybrid meshes is 3 (4 on regular meshes). The AVBP formulation is fully compressible and allows to investigate compressible combustion problems such as thermoacoustic instabilities (where acoustics are important) or detonation engines (where combustion and shock must be computed simultaneously). AVBP is a world standard for LES of combustion in engines and gas turbines, owned by CERFACS and IFP Energies Nouvelles. It is used by multiple laboratories (IMFT in Toulouse, EM2C in Centralesupelec, TU Munich, Von Karmann Institute, ETH Zurich, etc) and companies (SAFRAN AIRCRAFT ENGINES, SAFRAN HELICOPTER ENGINES, ARIANEGROUP, HERAKLES, etc). AVBP is also used today to compute turbomachinery (compressors and turbines) and to compute full engine configurations. Being able to compute simultaneously the compressor and the chamber of the chamber and the turbine or all three is now possible with AVBP. This is critical for multiple problems such as new propulsion concepts (such as Rotating Detonation Engines) or to study coupled phenomena such as the noise emitted from a gas turbine. AVBP has always been at the forefront of HPC research at CERFACS: its efficiency has been verified up to 250 000 cores with grids of 2 to 4 billion cells.

42. YALES2
https://www.coria-cfd.fr/index.php/YALES2#Solvers
CORIA lab: Joint lab from CNRS, INSA and University of Rouen
YALES2 aims at the solving of two-phase combustion from primary atomization to pollutant prediction on massive complex meshes. It is able to handle efficiently unstructured meshes with several billions of elements, thus enabling the Direct Numerical Simulation of laboratory and semi-industrial configurations. YALES2 is based on a large numerical library to handle partitioned meshes, various differential operators or linear solvers, and on a series of simple or more complex solvers:

- Scalar solver (SCS);

- Level set solver (LSS);
- Incompressible solver (ICS);
- Variable density solver (VDS);
- Spray solver (SPS = ICS + LSS + Ghost-Fluid Method);
- Lagrangian solver (LGS);
- Compressible solver (ECS);
- Magneto-hydrodynamic solver (MHD);
- Mesh movement solver (MMS);
- Radiative solver (RDS);
- Linear acoustics solver (ACS);
- Heat transfers solver (HTS);
- Immersed boundary solver (IBS);
- Granular Flow solver (GFS).

43. Nek5000
https://nek5000.mcs.anl.gov/
Argonne National Laboratory and Swiss Federal Institute of Technology, Zurich.
Simulation code Nek5000 sheds light on the turbulent flow fields of internal combustion engines, nuclear reactors, airplane wings, and more. The open-source software, which has evolved for over 30 years, features scalable algorithms that are fast and efficient on platforms ranging from laptops to the world's fastest computers. Highlights:

- Incompressible and low Mach-number Navier-Stokes;
- Spectral element disrectization;
- Runs on all POSIX compliant operating systems;
- Proven scalability to over a million ranks using pure MPI for parallelization;
- Easy-to-build with minimal dependencies;
- High-order conformal curved quadrilateral/hexahedral meshes;
- Semi-implicit 2nd/3rd order adaptive timestepping;
- Conjugate fluid-solid heat transfer;
- Efficient preconditioners;
- Parallel I/O;
- Lagrangian phase model;
- Moving and deforming meshes;
- Overlapping overset grids;
- Basic meshing tools including converters;
- LES and RANS(Reynolds-Averaged Navier-Stokes) turbulence models;
- VisIt and Paraview support for data analysis and visualization.

44. PRECISE_UNS
https://www.dolfyn.net/
Rolls-Royce and Institute Energy and Power Plant Technology (EKT) of Darmstadt University
Numerical Modeling Methods for Prediction of Ignition Processes in Aero-Engines. Code PRECISE-UNS (Predictive-System for Real Engine Combustors - Unstructured) is a finite volume based unstructured CFD solver for turbulent multi-phase and reacting flows. It is a pressure-based code, which uses the pressure correction scheme / PISO scheme to achieve pressure velocity coupling. It is applicable to both low-Mach number and fully compressible flows. Discretisation in time and space is up to second order. The linearized equations are solved using various well-known libraries such as PETSc, HYPRE and AGMG. Several turbulence models are available: k-epsilon, k-ω-SST, RSM, SAS, LES. Different combustion models are available, ranging from the classical conserved scalar (flamelet) models and global reaction mechanism, to FGM and detailed chemistry, PRECISE-UNS is built on Dolfyn, an open-source code written in Fortran. All investigations documented in this work have been performed using PRECISE-UNS.

Finite Element Computer Simulation
45.SALOME Library
Électricité de France ; Le Commissariat à l'énergie atomique et aux énergies alternatives (CEA)

https://www.edf.fr/groupe-edf/inventer-l-avenir-de-l-energie/r-d-un-savoir-faire-mondial/nos-offres/codes-de-calcul/salome
SALOME platform is an open software framework for integration of numerical solvers in various physical domains. The CEA and EDF use SALOME to realize a wide range of simulations, which typically concern industrial equipment in nuclear production plants. Among primary concerns are the design of new-generation reactor types, nuclear fuel management and transport, material ageing for equipment life-cycle management, and the reliability and safety of nuclear installations. To satisfy these challenges, SALOME integrates a CAD/CAE modeling tool, industrial meshing algorithms, and advanced 3D visualization functionalities. SALOME is a generic platform for numerical simulation with the following aims:

- Facilitate interoperation between CAD modelling and computing codes;
- Facilitate implementation of coupling between computing codes in a distributed environment;
- Provide a generic user interface;
- Pool production of developments (pre and post processors, calculation distribution and supervision) in the field of numerical simulation.

46. Weather Research and Forecasting Model (WRF)
https://www.mmm.ucar.edu/weather-research-and-forecasting-model
National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (represented by the National Centers for Environmental Prediction (NCEP) and the Earth System Research Laboratory), the U.S. Air Force, the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA).
The Weather Research and Forecasting (WRF) Model is a next-generation mesoscale numerical weather prediction system designed for both atmospheric research and operational forecasting applications. It features two dynamical cores, a data assimilation system, and a software architecture supporting parallel computation and system extensibility. The model serves a wide range of meteorological applications across scales from tens of meters to thousands of kilometers. For researchers, WRF can produce simulations based on actual atmospheric conditions (i.e., from observations and analyses) or idealized conditions. WRF offers operational forecasting a flexible and computationally-efficient platform, while reflecting recent advances in physics, numerics, and data assimilation contributed by developers from the expansive research community. The WRF system contains two dynamical solvers, referred to as the ARW (Advanced Research WRF) core and the NMM (Nonhydrostatic Mesoscale Model) core. The NMM core was developed by the National Centers for Environmental Prediction (NCEP), and is currently used in their HWRF (Hurricane WRF) system.

Libraries

47. PETCs -Portable, Extensible Toolkit for Scientific Computing
Mathematics and Computer Science Division, Argonne National Laboratory, USA.
https://www.mcs.anl.gov/petsc/
PETSc has been used for modeling in all of these areas: Acoustics, Aerodynamics, Air Pollution, Arterial Flow, Bone Fractures, Brain Surgery, Cancer Surgery, Cancer Treatment, Carbon Sequestration, Cardiology, Cells, CFD, Combustion, Concrete, Corrosion, Data Mining, Dentistry, EarthQuakes, Economics, Esophagus, Fission, Fusion, Glaciers, Ground Water Flow, Linguistics, Mantel Convection, Magnetic Films, Material Science, Medical Imaging, Ocean Dynamics, Oil Recover, Page Rank, Polymer Injection Molding, Polymeric Membranes, Quantum computing, Seismology, Semiconductors, Rockets, Relativity, Surface Water Flow.

48.ParMETIS - Parallel Graph Partitioning and Fill-reducing Matrix Ordering
Department of Computer Science and Engineering , University of Minnesota,
http://glaros.dtc.umn.edu/gkhome/metis/parmetis/overview
The algorithms implemented in METIS are based on the multilevel recursive-bisection, multilevel k-way, and multi-constraint partitioning schemes. The fill-reducing orderings produced by METIS are significantly better than those produced by other widely used algorithms including multiple minimum degree. For many classes of problems arising in scientific computations and linear programming, METIS is able to reduce the storage and computational requirements of sparse matrix factorization, by up to an order of magnitude. Moreover, unlike multiple minimum degree, the elimination trees produced by METIS are suitable for parallel direct factorization

49. Linear Algebra PACKage (LAPACK)
http://www.netlib.org/lapack/
LAPACK provides routines for solving systems of simultaneous linear equations, least-squares solutions of linear systems of equations, eigenvalue problems, and singular value problems. The associated matrix factorizations (LU, Cholesky, QR, SVD, Schur, generalized Schur) are also provided, as are related computations such as reordering of the Schur factorizations and estimating condition numbers. Dense and banded matrices are handled, but not general sparse matrices. In all areas, similar functionality is provided for real and complex matrices, in both single and double precision. The original goal of the LAPACK project was to make the widely used EISPACK and LINPACS libraries run efficiently on shared-memory vector and parallel processors. LAPACK requires that highly optimized block matrix operations be already implemented on each machine.

50.SuperLU
National Energy Research Scientific Computing Center (NERSC), Department of Energy, USA
https://portal.nersc.gov/project/sparse/superlu/
SuperLU is a general-purpose library for the direct solution of large, sparse, nonsymmetric systems of linear equations on high performance machines. The library is written in C and is callable from either C or Fortran. The library routines will perform an LU decomposition with partial pivoting and triangular system solves through forward and back substitution. The LU factorization routines can handle non-square matrices but the triangular solves are performed only for square matrices.

51. Multifrontal massively parallel sparse direct solver (MUMSPS)
CERFACS, CNRS, ENS Lyon, INP Toulouse, Inria, Mumps Technologies, University of Bordeaux.
http://mumps.enseeiht.fr/
Multifrontal massively parallel sparse direct solver (MUMSPS) for solution of large linear systems with symmetric positive definite matrices, general symmetric matrices and general unsymmetrical matrices. Several reorderings interfaced: AMD, QAMD, AMF, PORD, METIS, PARMETIS, SCOTCH, PT-SCOTCH.

52.Trilinos
Trilinos Community
https://trilinos.github.io/index.html
Trilinos is a collection of open-source software libraries, called packages, intended to be used as building blocks for the development of scientific applications. Trilinos facilitate the design, development, integration and ongoing support of mathematical software libraries within an object-oriented framework for the solution of large-scale, complex multi-physics engineering and scientific problems. Trilinos addresses two fundamental issues of developing software for these problems: Providing a streamlined process and set of tools for development of new algorithmic implementations; Promoting interoperability of independently developed software.

53. Scalable Linear Solvers and Multigrid Methods HYPRE
Lawrence Livermore National Laboratory, Department of Energy, USA
https://computing.llnl.gov/projects/hypre-scalable-linear-solvers-multigrid-methods
HYPRE library of linear solvers makes possible larger, more detailed simulations by solving problems faster than traditional methods at large scales. It offers a comprehensive suite of scalable solvers for large-scale scientific simulation, featuring parallel multigrid methods for both structured and unstructured grid problems. The HYPRE library is highly portable and supports a number of languages. The HYPRE team was one of the first to develop algebraic multigrid algorithms and software for extreme-scale parallel supercomputers.

Monte Carlo Simulations

54.Geant4
CERN
https://geant4.web.cern.ch/node/1
GEANT4 is a software toolkit for the simulation of the passage of particles through matter. It is used by a large number of experiments and projects in a variety of application domains, including high energy physics, astrophysics and space science, medical physics and radiation protection. As a Monte Carlo simulation toolkit, Geant4 profits from improved throughput via parallelism derived from the

independence of modeled events and their computation. Until Geant4 version 10.0, parallelization was obtained with a simple distribution of inputs: each computation unit (e.g. a core of a node in a cluster) ran a separate instance of Geant4 that was given a separate set of input events and associated random number seeds. Given a computer with k cores, the design goal of multithreaded Geant4 was to replace k independent instances of a Geant4 process with a single, equivalent process with k threads using the many-core machine in a memory-efficient, scalable manner. The corresponding methodology involved transforming the code for thread safety and memory footprint reduction.

Libraries for Artificial Intelligence and Data Analysis

55. The R Project for Statistical Computing
https://www.r-project.org/
R is a programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. The basic package (inlcuding the HPC package) of R is installed on the HPC. Users can install their own packages in their home directories. The packages available include: Chemometrics and Computational Physics, Clinical Trial Design, Monitoring, and Analysis, Econometrics, Analysis of Ecological and Environmental Data, Empirical Finance, Statistical Genetics, Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization, Hydrological Data and Modeling, Machine Learning & Statistical Learning, Medical Image Analysis, Multivariate Statistics, Natural Language Processing, Psychometric Models and Methods, Analysis of Spatial Data, among others.

56. Python
Modern, interpreted, object-oriented, full featured high-level programming language. Versions to include are 2.7.x and 3.x. Packages for Computational Science include: numpy and pandas for data operation and analysis, scipy for higher level computational routines, matplotlib for plotting. Additional packages can be installed in user's home directory.

57.Tensorflow
https://www.tensorflow.org/learn
TensorFlow is an open source software library for machine learning developed by Google. Its mission is to train and build neural networks. It can be used on CPU and GPU architectures. It is furthermore an open source software library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. The flexible architecture allows you to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API.

58. Machine Learning Libraries – Keras, Caffe, Pytorch, Theano.
A Python version of Torch, known as Pytorch, was open-sourced by Facebook in January 2017. PyTorch offers dynamic computation graphs, which let you process variable-length inputs and outputs. Torch is a computational framework with an API written in Lua that supports machine-learning algorithms. Some version of it is used by large tech companies such as Facebook and Twitter, which devote in-house teams to customizing their deep learning platforms. Lua is a multi-paradigm scripting language that was developed in Brazil in the early 1990s. Caffe is a well-known and widely used machine-vision library that ported Matlab's implementation of fast convolutional nets to C and C++ (see Steve Yegge's rant about porting C++ from chip to chip if you want to consider the tradeoffs between speed and this particular form of technical debt). Caffe is not intended for other deep-learning applications such as text, sound or time series data. Like other frameworks mentioned here, Caffe has chosen Python for its API. Caffe2 is the second deep-learning framework to be backed by Facebook after Torch/PyTorch. The main difference seems to be the claim that Caffe2 is more scalable and light-weight. It purports to be deep learning for production environments. Like Caffe and PyTorch, Caffe2 offers a Python API running on a C++ engine.

Libraries for Neuroscience

59. The Neural Simulation Tool NEST

Prof. Dr. Markus Diesmann Institute of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Jülich Research Center, Jülich, Germany, Prof. Dr. Marc-Oliver Gewaltig École Polytechnique Fédérale de Lausanne, Switzerland
https://www.nest-simulator.org/
NEST is a simulator for spiking neural network models from small-scale microcircuits to brain-scale networks of the order of 10^8 neurons and 10^12 synapses. Main features include: Integrate-and-fire neuron models with current- and conductance-based synapses, Adaptive threshold integrate-and-fire neuron models (AdEx, MAT2), Hodgkin-Huxley type neuron models with one compartment, Simple multi-compartmental neuron models, Static and plastic synapse models (STDP, short-term plasticity, neuromodulation), Grid based spike interaction and interaction in continuous time, Exact Integration for linear neuron models and appropriate solvers for others, Topology Module and support for CSA for creating complex networks.

60.NEURON - flexible and powerful simulator of neurons and networks
Yale University, USA
https://neuron.yale.edu/neuron/
It was primarily developed by Michael Hines, John W. Moore, and Ted Carnevale at Yale and Duke.
Simulation environment for modeling individual neurons and networks of neurons. It provides tools for conveniently building, managing, and using models in a way that is numerically sound and computationally efficient. It is particularly well-suited to problems that are closely linked to experimental data, especially those that involve cells with complex anatomical and biophysical properties. NEURON's computational engine employs special algorithms that achieve high efficiency by exploiting the structure of the equations that describe neuronal properties. It has functions that are tailored for conveniently controlling simulations, and presenting the results of real neurophysiological problems graphically in ways that are quickly and intuitively grasped.

61. Extreme Parallel Tools for Brain Neural Network Simulations
Prof. Stoyan Markov, Dr. Kristina Kapanova , Mag. , Jasmine Brune
New tool developed by NCSA, Bulgaria giving the capability to simulate Hodgkin-Huxley type neuron models, while working in a pipeline algorithmic manner. Different to other simulation tools, the simulation is cell driven, with a user specified number of cycles, which describe the simulation time of the system. Each cycle consists of a 300ms window, during which the HH membrane potential is computed and the reaction of the cell is recorded. The algorithm works on a densely interconnected and sparse outside connection model, thus significantly reducing the volume of communications. Correspondingly during the simulation of the neural network each neuron is considered as an independent entity by means of the cell data structure, which records all the required communication.

The Performance Optimisation and Productivity Centre of Excellence in Computing Applications.

62 .Extrae
Barcelona Supercomputing Centre, Barcelona, Spain
https://tools.bsc.es/
Extrae is the package devoted to generate Paraver trace-files for a post-mortem analysis. It is a tool that uses different interposition mechanisms to inject probes into the target application so as to gather information regarding the application performance.

63. Scalasca
Forschungszentrums Jülich, Jülich, Germany
https://www.scalasca.org/
Scalasca is a software tool that supports the performance optimization of parallel programs by measuring and analyzing their runtime behavior. The analysis identifies potential performance bottlenecks – in particular those concerning communication and synchronization – and offers guidance in exploring their causes. Scalasca targets mainly scientific and engineering applications based on the programming interfaces MPI and OpenMP, including hybrid applications based on a combination of the two. The tool has been specifically designed for use on large-scale systems.

64. Cube
https://www.scalasca.org/scalasca/software/cube-4.x/download.html

Cube, which is used as performance report explorer for Scalasca and Score-P, is a generic tool for displaying a multi-dimensional performance space consisting of the dimensions (i) performance metric, (ii) call path, and (iii) system resource. Each dimension can be represented as a tree, where non-leaf nodes of the tree can be collapsed or expanded to achieve the desired level of granularity. In addition, Cube can display multi-dimensional Cartesian process topologies.

65.Score-P
Forschungszentrums Jülich, Jülich, Germany
http://scorepci.pages.jsc.fz-juelich.de/scorep-pipelines/doc.r14401/quickstart.html
Scalable Performance Measurement Infrastructure for Parallel Codes (Score-P) measurement infrastructure is a highly scalable and easy-to-use tool suite for profiling, event tracing, and online analysis of HPC applications. Score-P offers the user a maximum of convenience by supporting a number of analysis tools. Currently, it works with Periscope, Scalasca, Vampir, and Tau and is open for other tools.

66.Vampir
Forschungszentrums Jülich, Jülich, Germany
Vampir provides an easy-to-use framework that enables developers to quickly display and analyze arbitrary program behavior at any level of detail. The tool suite implements optimized event analysis algorithms and customizable displays that enable fast and interactive rendering of very complex performance monitoring data. Vampir and Score-P provide a performance tool framework with special focus on highly-parallel applications. Performance data is collected from multi-process (MPI, SHMEM), thread-parallel (OpenMP, Pthreads), as well as accelerator-based paradigms (CUDA, OpenCL, OpenACC).

67. Extra-P
Forschungszentrums Jülich, Jülich, Germany https://www.scalasca.org/software/extra-p/
Extra-P is an automatic performance-modeling tool that supports the user in the identification of scalability bugs. A scalability bug is a part of the program whose scaling behavior is unintentionally poor, that is, much worse than expected. Extra-P uses measurements of various performance metrics at different processor configurations as input to represent the performance of code regions (including their calling context) as a function of the number of processes. All it takes to search for scalability issues even in full-blown codes is to run a manageable number of small-scale performance experiments, launch Extra-P, and compare the asymptotic or extrapolated performance of the worst instances to the expectations.

68. Paraver: a flexible performance analysis tool
Barcelona Supercomputing Centre, Barcelona, Spain. https://tools.bsc.es/paraver
Paraver was developed to respond to the need to have a qualitative global perception of the application behavior by visual inspection and then to be able to focus on the detailed quantitative analysis of the problems. Expressive power, flexibility and the capability of efficiently handling large traces are key features addressed in the design of Paraver. The clear and modular structure of software plays a significant role towards achieving these targets. Some features of its features include:

- Detailed quantitative analysis of program performance;
- Concurrent comparative analysis of several traces;
- Customizable semantics of the visualized information;
- Cooperative work, sharing views of the tracefile;
- Building of derived metrics;
- The following are major features of the Paraver philosophy and functionality.