General Purpose Supercomputing on SuperMUC

Dieter Kranzlmüller

Munich Network Management Team
Ludwig-Maximilians-Universität München (LMU) &
Leibniz Supercomputing Centre (LRZ)
of the Bavarian Academy of Sciences and Humanities

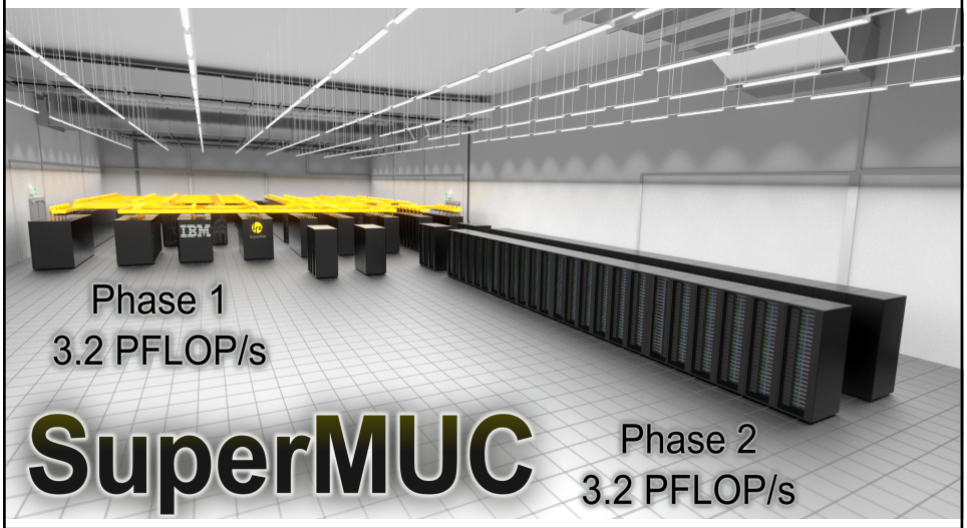
Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities

With approx. 250 employees
for more than 100.000 students and
for more than 30.000 employees
including 8.500 scientists

- European Supercomputing Centre
- National Supercomputing Centre ➔ GCS
- Regional Computer Centre for all Bavarian Universities
- Computer Centre for all Munich Universities

Photo: Ernst Graf

SuperMUC Phase 1 + 2



LRZ Application Mix

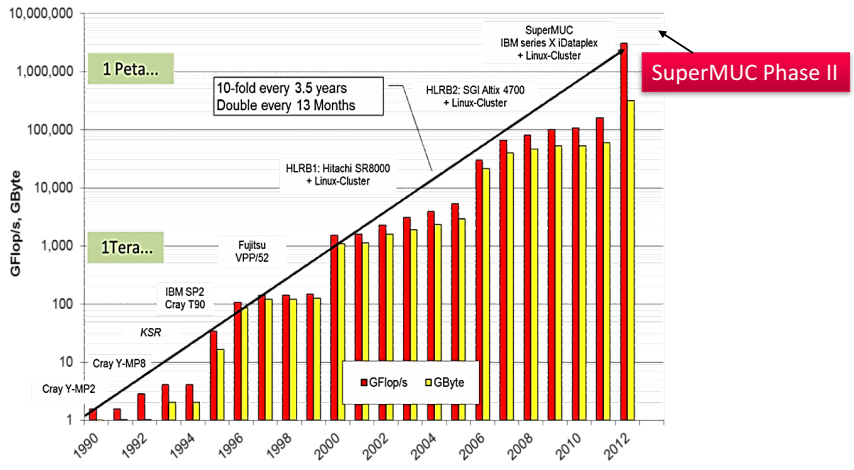
- **Computational Fluid Dynamics**: Optimisation of turbines and wings, noise reduction, air conditioning in trains
- **Fusion**: Plasma in a future fusion reactor (ITER)
- **Astrophysics**: Origin and evolution of stars and galaxies
- **Solid State Physics**: Superconductivity, surface properties
- **Geophysics**: Earth quake scenarios
- **Material Science**: Semiconductors
- **Chemistry**: Catalytic reactions
- **Medicine and Medical Engineering**: Blood flow, aneurysms, air conditioning of operating theatres
- **Biophysics**: Properties of viruses, genome analysis
- **Climate research**: Currents in oceans

Top 500 Supercomputer List (June 2012)

Rank	Site	Computer/Year Vendor	Cores	R_{max}	R_{peak}	Power
1	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom / 2011 IBM	1572864	16324.75	20132.66	7890.0
2	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect / 2011 Fujitsu	705024	10510.00	11280.38	12659.9
3	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	786432	8162.38	10066.33	3945.0
4	Leibniz Rechenzentrum Germany	**SuperMUC** - iDataPlex DX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR / 2012 IBM	147456	2897.00	3185.05	3422.7
5	National Supercomputing Center in Tianjin China	**Tianhe-1A** - NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050 / 2010 NUDT	186368	2566.00	4701.00	4040.0
6	DOE/SC/Oak Ridge National Laboratory United States	**Jaguar** - Cray XK6, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA 2090 / 2009 Cray Inc.	298592	1941.00	2627.61	5142.0
7	CINECA Italy	**Fermi** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	163840	1725.49	2097.15	821.9
8	Forschungszentrum Juelich (FZJ) Germany	**JuQUEEN** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom / 2012 IBM	131072	1380.39	1677.72	657.5
9	CEA/TGCC-GENCI France	**Curie thin nodes** - Bullx B510, Xeon E5-2680 8C 2.700GHz, Infiniband QDR / 2012 Bull	77184	1359.00	1667.17	2251.0
10	National Supercomputing Centre in Shenzhen (NSCS) China	**Nebulae** - Dawning TC3600 Blade System, Xeon X5650 6C 2.66GHz, Infiniband QDR, NVIDIA 2050 / 2010 Dawning	120640	1271.00	2984.30	2580.0

www.top500.org

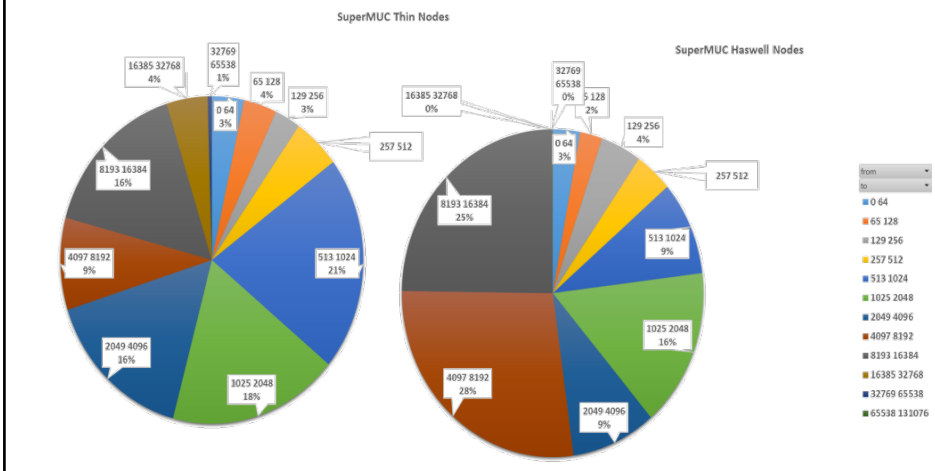
LRZ Supercomputers



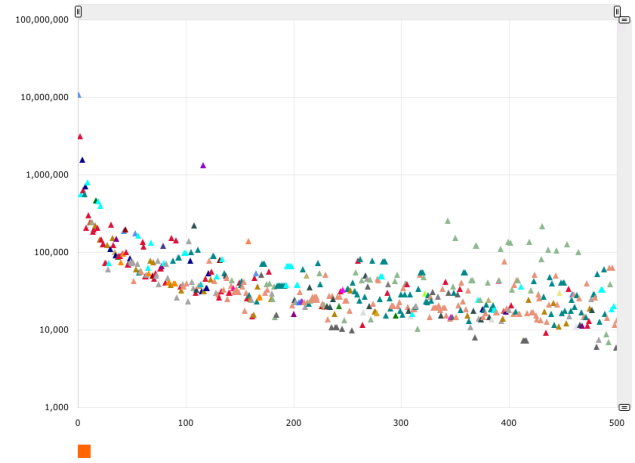
Increasing numbers

Date	System	Flop/s	Cores
2000	HLRB-I	2 Tflop/s	1512
2006	HLRB-II	62 Tflop/s	9728
2012	SuperMUC	3200 Tflop/s	155656
2015	SuperMUC Phase II	3.2 + 3.2 Pflop/s	229960

SuperMUC Jobsize 2015 (in Cores)



Statistics from www.top500.org



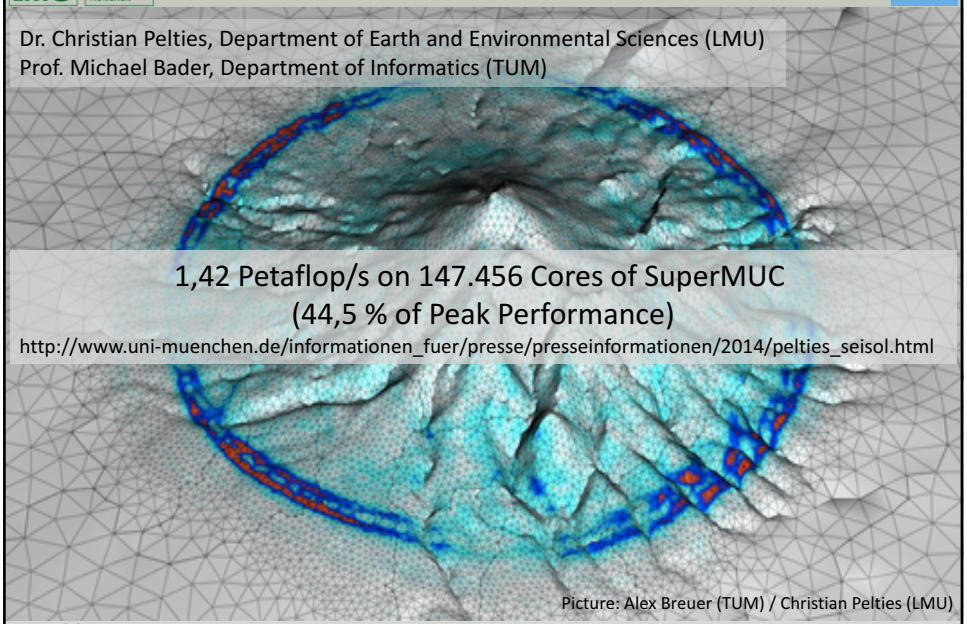
2016

Legend:

IBM, Hitachi/Fujitsu, HPE/SGI, Nvidia, NEC/HPE, Inspur, Huawei, T-Platforms, Sugon, ClusterVision, Koi Computers, Adtech, Dell / Intel, IPE, Nvidia, Tyan, Dell, SuperMicro/Mellanox, Cray Inc., Bull, ACTION, NUDT, Lenovo, NEC, Intel, DALCO, MEGWARE, Atipa, Self-made, IBM/Lenovo, Fujitsu, Penguin Computing, PEZY Computing / Exascaler Inc., , HPE, Megatel/Action, Dell/Sugon, EXXACT, Supermicro, Lenovo/IBM, Oracle, NRCPC, RSC Group,

SeisSol - Numerical Simulation of Seismic Wave Phenomena

Dr. Christian Pelties, Department of Earth and Environmental Sciences (LMU)
Prof. Michael Bader, Department of Informatics (TUM)

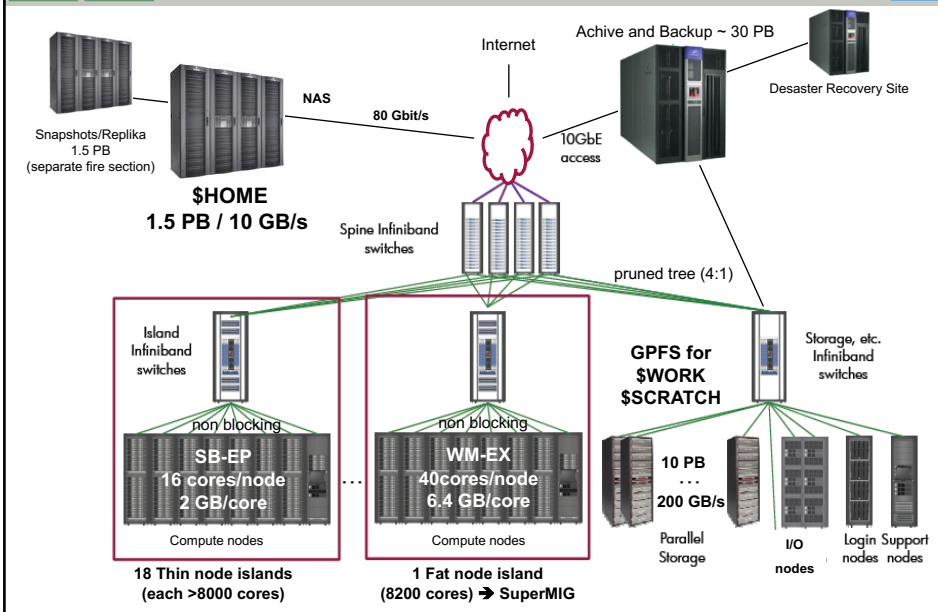


1,42 Petaflop/s on 147.456 Cores of SuperMUC
(44,5 % of Peak Performance)

http://www.uni-muenchen.de/informationen_fuer/presse/presseinformationen/2014/pelties_seisol.html

Picture: Alex Breuer (TUM) / Christian Pelties (LMU)

SuperMUC Architecture



SuperMUC System @ LRZ

Phase 1 (IBM System x iDataPlex):

- 3.2 PFlops peak performance
- 9216 IBM iDataPlex dx360M4 nodes in 18 compute node islands
- 2 Intel Xeon E5-2680 processors and 32 GB of memory per compute node
- 147,456 compute cores
- Network Infiniband FDR10 (fat tree)

Phase 2 (Lenovo NeXtScale WCT):

- 3.6 PFlops peak performance
- 3072 Lenovo NeXtScale nx360M5 WCT nodes in 6 compute node islands
- 2 Intel Xeon E5-2697v3 processors and 64 GB of memory per compute node
- 86,016 compute cores
- Network Infiniband FDR14 (fat tree)

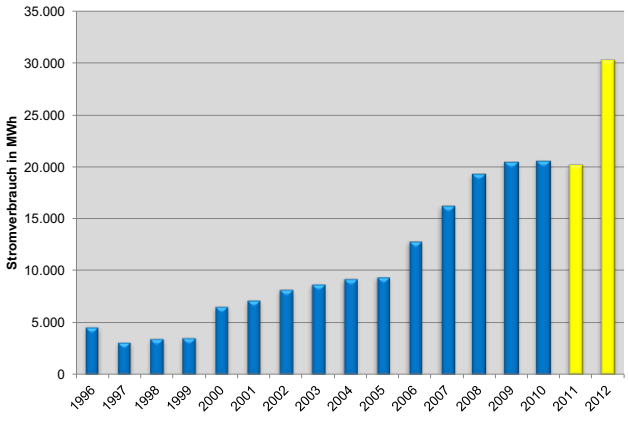
Common GPFS file systems with 10 PB and 5 PB usable storage size respectively
Common programming environment
Direct warm-water cooled system technology

Gauss Centre for Supercomputing (GCS)

High(est) Performance Computing in Germany

- Combination of the 3 German national supercomputing centers:
  - John von Neumann Institute for Computing (NIC), Jülich
  - High Performance Computing Center Stuttgart (HLRS)
  - Leibniz Supercomputing Centre (LRZ), Garching n. Munich
- Founded on 13. April 2007
- Hosting member of PRACE
  (Partnership for Advanced Computing in Europe)

Power Consumption at LRZ



Cooling SuperMUC

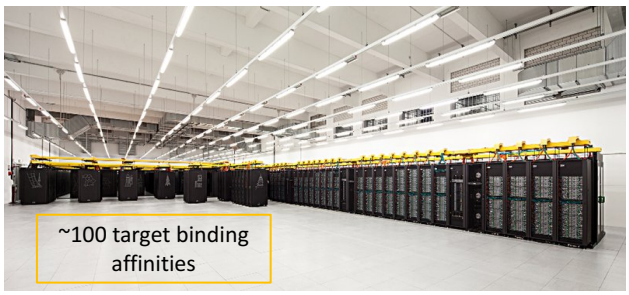
SuperMUC Phase 2 @ LRZ

- Intel Xeon E5 2697v3 CPU
- Direct liquid cooling
- Energy-aware scheduling
- Total annual savings: ~2 Mio. € for SuperMUC Phase 1 and 2

Photos: Torsten Bloth, Lenovo

Molecular Simulation for Personalized Medicinem (Prof. Peter Coveney, UCL)

- Running on all cores of SuperMUC Phase1+2



- Docking simulation of potentials drugs for breast cancer
- 37 hours total run time
- 241,672 cores
- 8.900.000 CPU hours
- 5 Terabytes of data produced

http://www.compat-project.eu

http://www.compbiomed.eu

What's next? A View into the Future

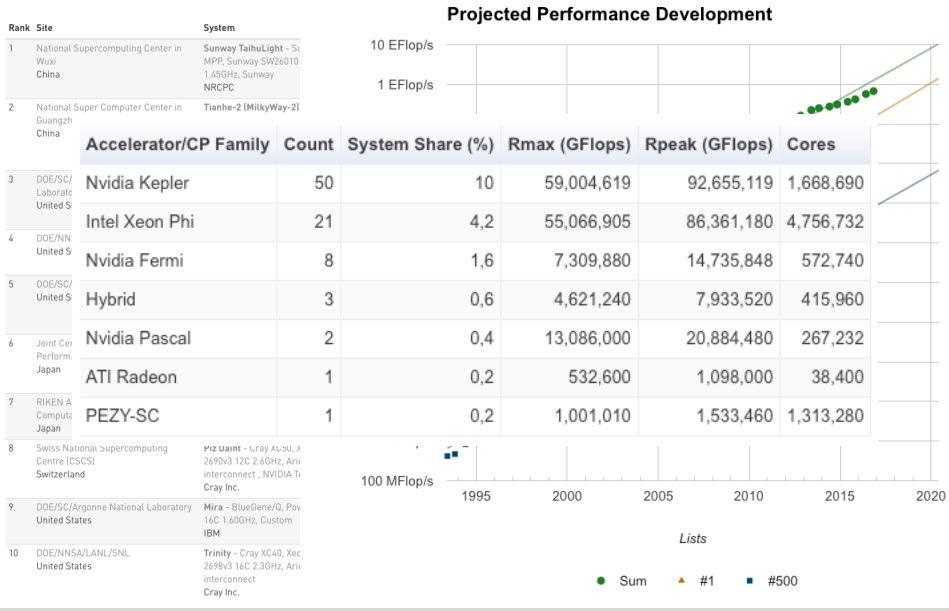
Until today:

- HLRB-II (pre-SuperMUC): Top 500 06/2007: 56,5 Tflop/s
- SuperMUC Phase 1: Top 500 06/2012: 2897 Tflop/s

Coming up:

- **SuperMUC NG** (Next Generation) – Procurement on-going

Consulting the Top 500 List - www.top500.org



Rank	Site	System
1	National Supercomputing Center in Wuxi China	Sunway TaihuLight - S MPP, Sunway SW26010 1.45GHz, Sunway NRCPC
2	National Super Computer Center in Guangzh China	Tianhe-2 (MilkyWay-2)
3	DOE/SC/ Laborat United S	
4	DOE/NN United S	
5	DOE/SC/ United S	
6	Joint Ce Perform Japan	
7	RIKEN A Comput Japan	
8	Swiss National Supercomputing Centre (CSCS) Switzerland	Piz Daint - Cray XC30, 2690v3 12C 2.6GHz, Ari interconnect , NVIDIA T Cray Inc.
9	DOE/SC/Argonne National Laboratory United States	Mira - BlueGene/Q, Pov 16C 1.60GHz, Custom IBM
10	DOE/NNSA/LANL/SNL United States	Trinity - Cray XC40, Xe 2698v3 16C 2.3GHz, Ari interconnect Cray Inc.

Accelerator/CP Family	Count	System Share (%)	Rmax (GFlops)	Rpeak (GFlops)	Cores
Nvidia Kepler	50	10	59,004,619	92,655,119	1,668,690
Intel Xeon Phi	21	4,2	55,066,905	86,361,180	4,756,732
Nvidia Fermi	8	1,6	7,309,880	14,735,848	572,740
Hybrid	3	0,6	4,621,240	7,933,520	415,960
Nvidia Pascal	2	0,4	13,086,000	20,884,480	267,232
ATI Radeon	1	0,2	532,600	1,098,000	38,400
PEZY-SC	1	0,2	1,001,010	1,533,460	1,313,280

What's next? A View into the Future

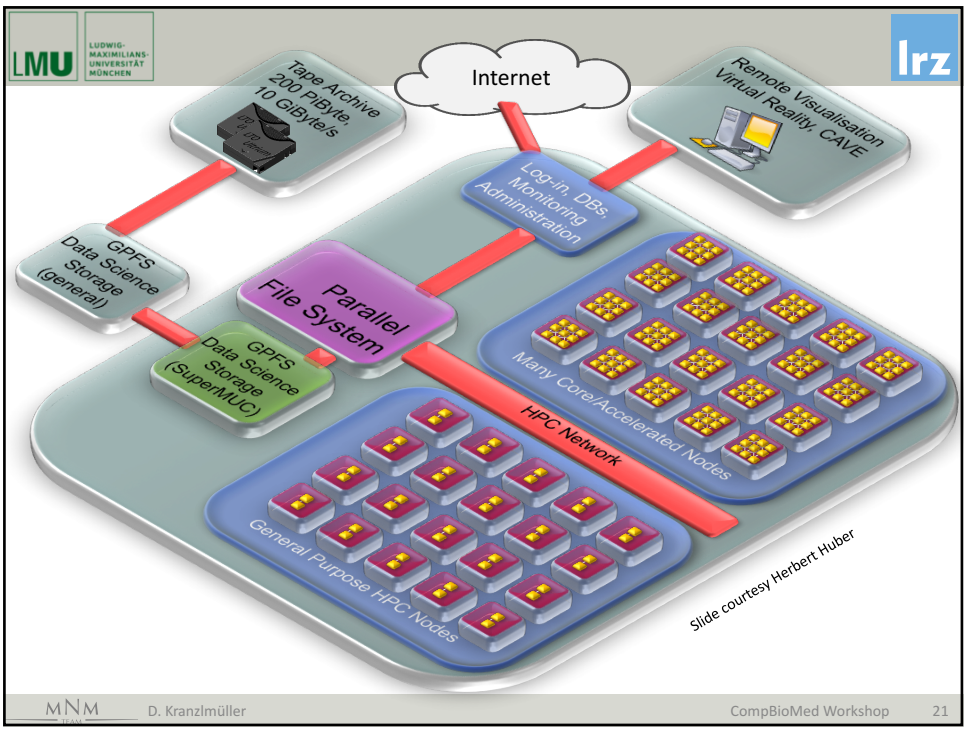
Until today:

- HLRB-II (pre-SuperMUC): Top 500 06/2007: 56,5 Tflop/s
- SuperMUC Phase 1: Top 500 06/2012: 2897 Tflop/s

Coming up:

- **SuperMUC NG** (Next Generation) – Procurement on-going

- **Selection criteria:**
  - LRZ application mix (compute,memory,bandwidth characteristics)
    - Number of cores
    - Memory per core
    - Interconnect
  - Accelerators (Manycore, GPGPU, ...)
  - Virtualization (Docker, Cloud, ...)
  - Workflow engines, HTC applications, ...
  - Power consumption (in total, over time, ...)



Partnership Initiative Computational Sciences πCS

- **Individualized services** for selected scientific groups – **flagship role**
  - Dedicated point-of-contact
  - Individual support and guidance and targeted training &
  - Planning dependability for use case specific optimi uctures
  - Early access to latest IT infrastructure (hard- a developments and specification of future requirements
  - Access to IT competence network a at CS and Math departments
- **Partner contribution**
  - Embedding IT expe oups
  - Joint research luding funding)
  - Scientifi – equal footing – joint publications
- **LRZ h**
  - U ding the (current and future) needs and requirements of the respective scientific domain
  - Developing future services for all user groups
  - Thematic focusing: **Environmental Computing**

The LRZ Partnership Initiative Computational Science (piCS)
http://www.sciencedirect.com/science/article/pii/S1877050914003433

piCS Cookbook

1. Choose focus topics to serve as lighthouse
   - National agreement within GCS: LRZ focuses on Environment (& Energy)
2. Choose user communities
   - Already active at LRZ?
   - Not active at LRZ?
3. Invite them for introductory piCS Workshops
   - Show faces & tour
   - Discussion on joint topics, requirements, interests, ...
4. Establish links between communities and specific points-of-contact
   - Whom to talk to, if there are questions?
   - When to talk to them? In general, as early as possible
   - Maybe, place people into the research groups (weekly, for a certain period)
5. Run joint lectures (e.g. hydrometeorology and computer science)
6. Apply for joint projects
7. Use **HPC Machines** efficiently to do science