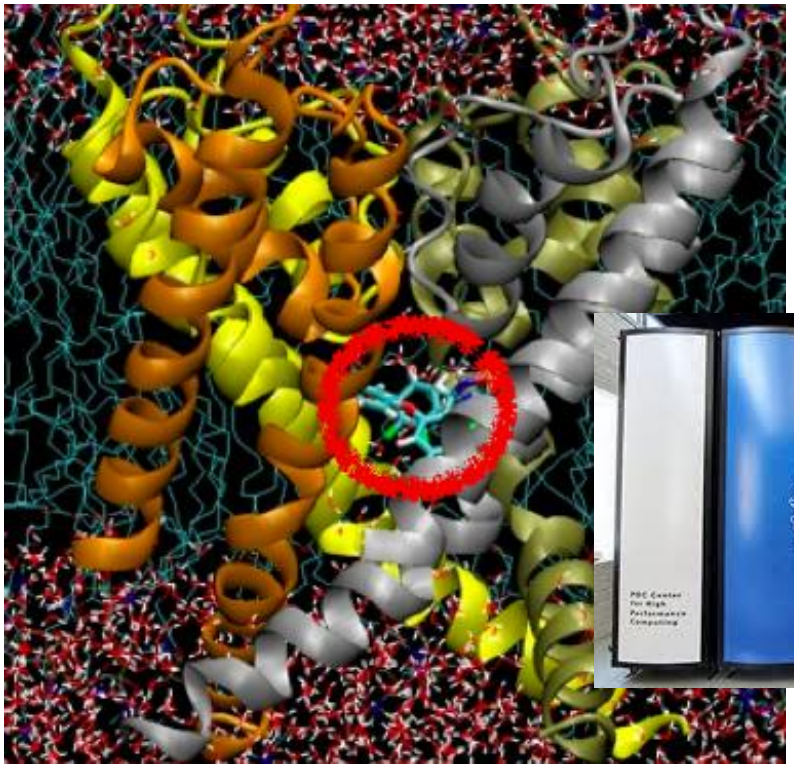
Scalable Software Services for Life Science

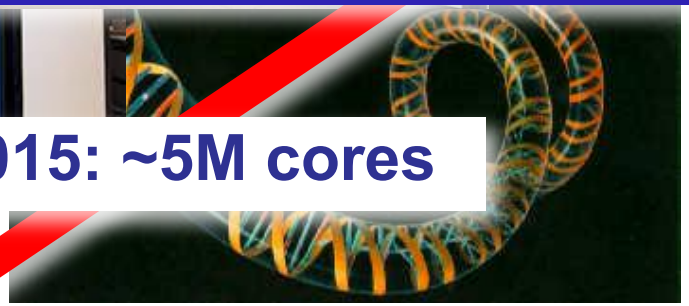
MAPPER School, 1-2 Februrary 2012

Mihai Duta, Oxford e-Research Centre

University of Oxford

The Multicore Challenge





2030: ~1B cores

2015: ~5M cores

2010: ~300,000 cores

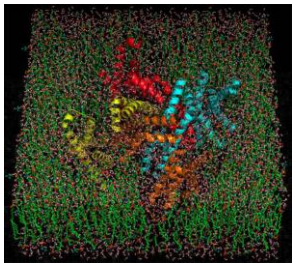
A “few” MD Packages

AMBER
GROMACS
HOOMD
ProtoMol
Tinker
LAMMPS
NAMD
ESPResSo
COSMOS
Presto
CHARMM
MacroModel
MDynaMix
ACEMD
Kalypso
ADUN
Desmond
Materials Studio
Ascalaph
RedMD
GULP
Hippo
LPMD
Culgi
MOLDY
MOSCITO
YASARA
GROMOS
XMD
ORAC
Abalone

Motivation

Life Science (LS) are amongst the largest European e-Infrastructure users, with a quickly growing demand on simulation software and support services. A gap between:

- e-Infrastructure projects (PRACE/DEISA) largely focused on (theoretical) scalability at the large-system end of the spectrum
- typical LS applications concern small-to-medium size molecules and run on small-to-medium systems
  - sequence bioinformatics
  - in silico virtual screening
  - free energy calculations (e.g. of drug binding)



At the same time, although there are efforts to integrate HPC practice and results into LS, they tend to be fragmented.

The EU **ScalaLife** project is a coordinated initiative to bring HPC to LS with a focus on the efficient use of LS applications on small-to-medium systems.

Partners + Collaborations

Leibniz-Rechenzentrum, Germany

Oxford e-Research Center, United Kingdom

Synective Labs, Sweden

Royal Institute of Technology, Sweden

PDC Center for High Performance Computing at KTH, Sweden

Institute for Research in BioMedicine, Spain

Barcelona Supercomputing Center, Spain

ScalaLife has important involvement from:

- academic and industrial partners
- major European Supercomputing centres
- application developers (of the pilot software)

ScalaLife collaborations:

- **ERINA+, INCF, EGI, Mapper, HealthGrid, PRACE**

Objectives

ScalaLife started in September 2010, with the following aims:

- to develop new hierarchical **high-throughput solutions**
  - parallel techniques for multi-core and streaming/GPGPU cluster architectures
  - open software standards for data storage/exchange
- to implement, document and maintain these solutions in "pilot" European **open-source codes**:
  - **Gromacs** (MD application)
  - **Dalton** (QM application)
  - **Discrete** (new ensemble simulation application)
  - newcomers: **Ergo** (QM) and **Music** (neural network simulations)
- to capture this knowledge/experience by establishing a **Competence Centre** for scalable LS software to
  - strengthen Europe as a major LS software provider
  - enable European LS communities to fully exploit e-Infrastructures

Project Outcomes

In the long run, the project results to disseminate will be:

- competitive **speed and scalability** demonstrated by the project pilot codebases (Gromacs and Dalton in particular) on small-to-medium molecules (and accelerated on GPGPU);
- **libraries and APIs** to facilitate the easy exploitation of hardware (*e.g.* acceleration on GPGPUs, hardware locality and process/thread placement) or of low-level message-passing communication to reduce the MPI latency;
- new **standards** for LS data storage and exchange;
- the creation of the **Competence Center**, which will provide an expert service for the LS community and be the main vehicle for knowledge transfer.

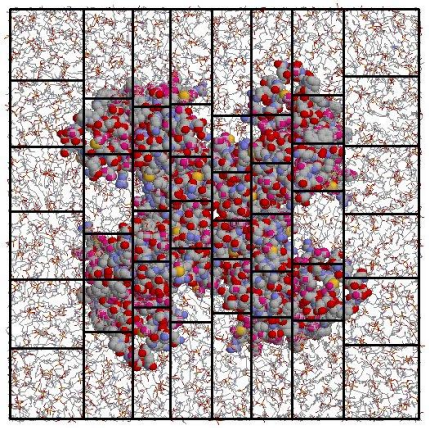
Example: Gromacs

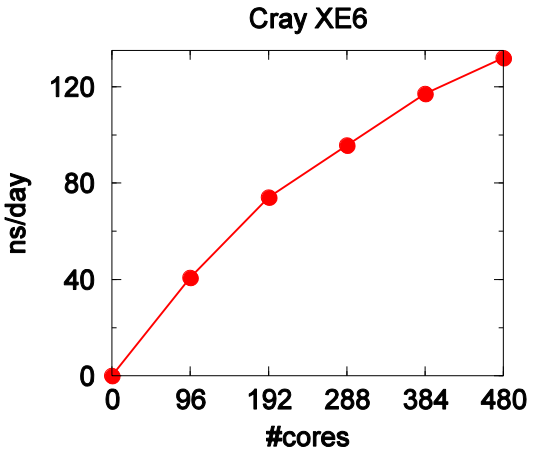
Gromacs is a European molecular dynamics package, supports all algorithms expected from a modern MD code but is competitive due to a number of features:

- **free** software (GNU GPL)
- **fast** on low-end desktops (hand-optimised code) and fast on high-end parallel computers using standard MPI communication

Benchmark: Kv1.2 ion channel, protein with 120k+ atoms:

- 2D PME particle-particle interaction mesh decomposition
- decomposition is dynamically load-balanced





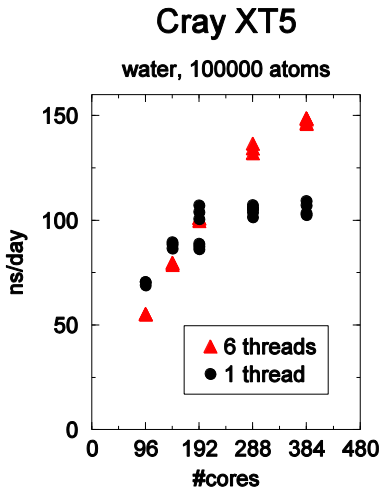
Example: Gromacs

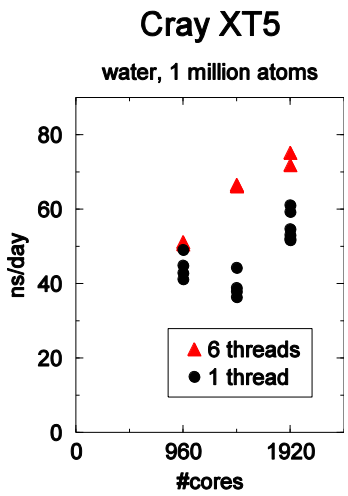
Hybrid programming for multi-core "fat" SMP compute node clusters:

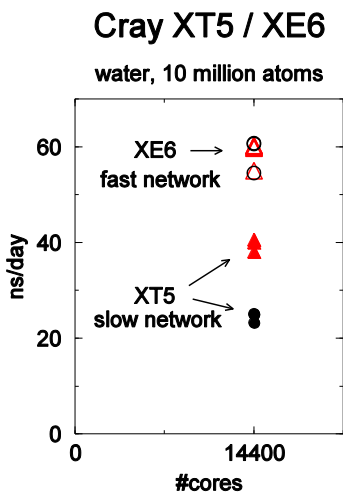
- use MPI for inter-node communication and OpenMP for in-node threading
- can have better scalability than pure MPI
  - avoids MPI communication overhead within node
  - allows large MPI message sizes, which can improve load balancing
  - increases the scope for overlapping computation and communication

Treads improve scalability considerably over a slow(er) network

- 60% improvement of threads on a Cray XT5 but none on XE6 (fast network)
- 10 million atoms, 14400 cores







Competence Centre

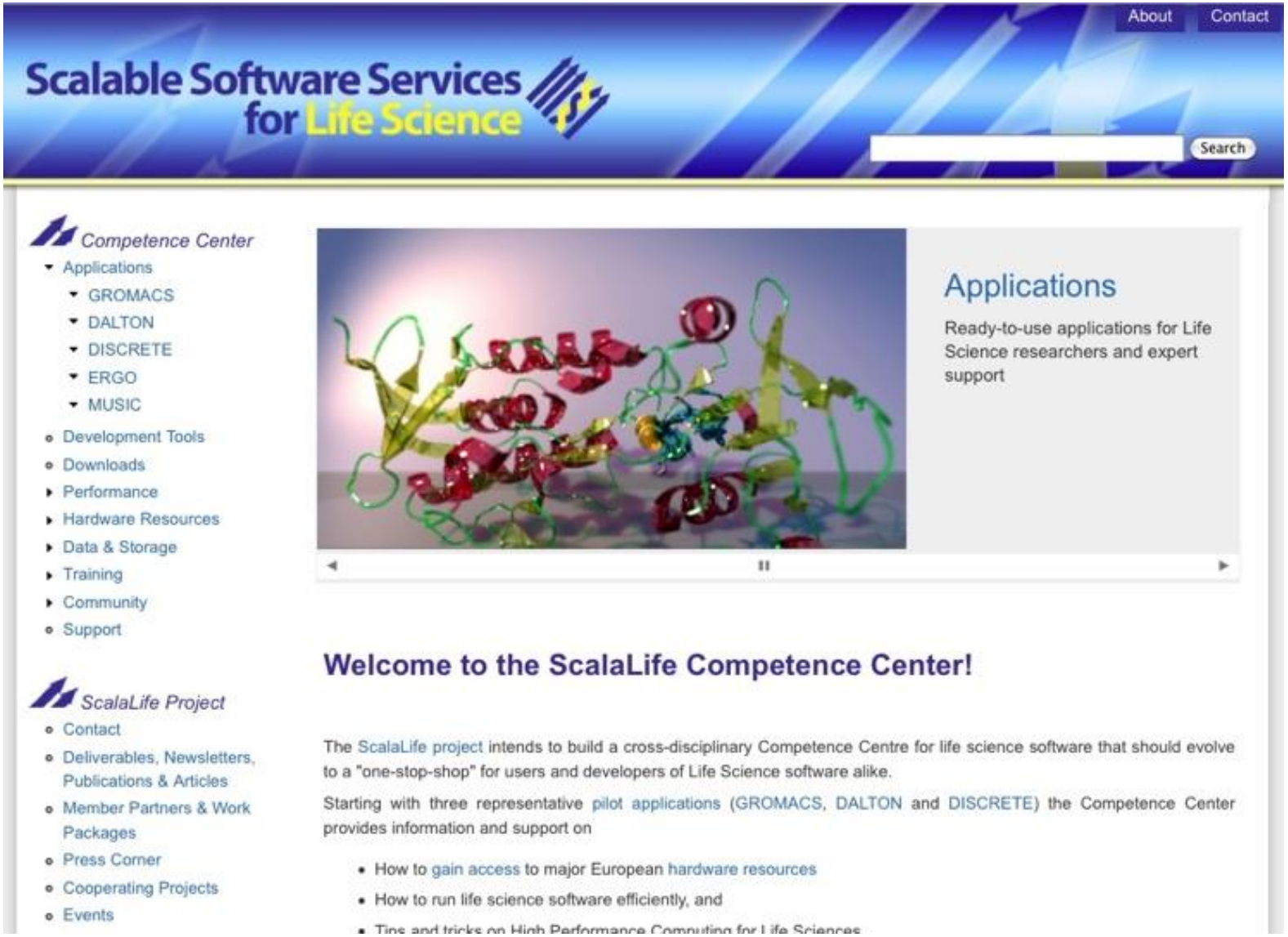
The long-term strategy of **ScalaLife** is the establishment of a **Competence Centre** to act as a "one-stop" shop for the European LS community seeking software guidance. Key steps:

- capturing of the experience gained with the pilot codes and make it available through a support and training network (for both software developers and users)
  - training (events and online material)
  - tutorials and best practice guides
- creating a (self-) sustainable "marketplace" for expertise and collaborative work
  - attract collaboration with other (similar) projects
  - provide expert advice and assistance (e.g. application profiling)
  - attract alpha users (offering privileged support, access to hardware, etc.)

The first version of the Centre went live before August 2011.

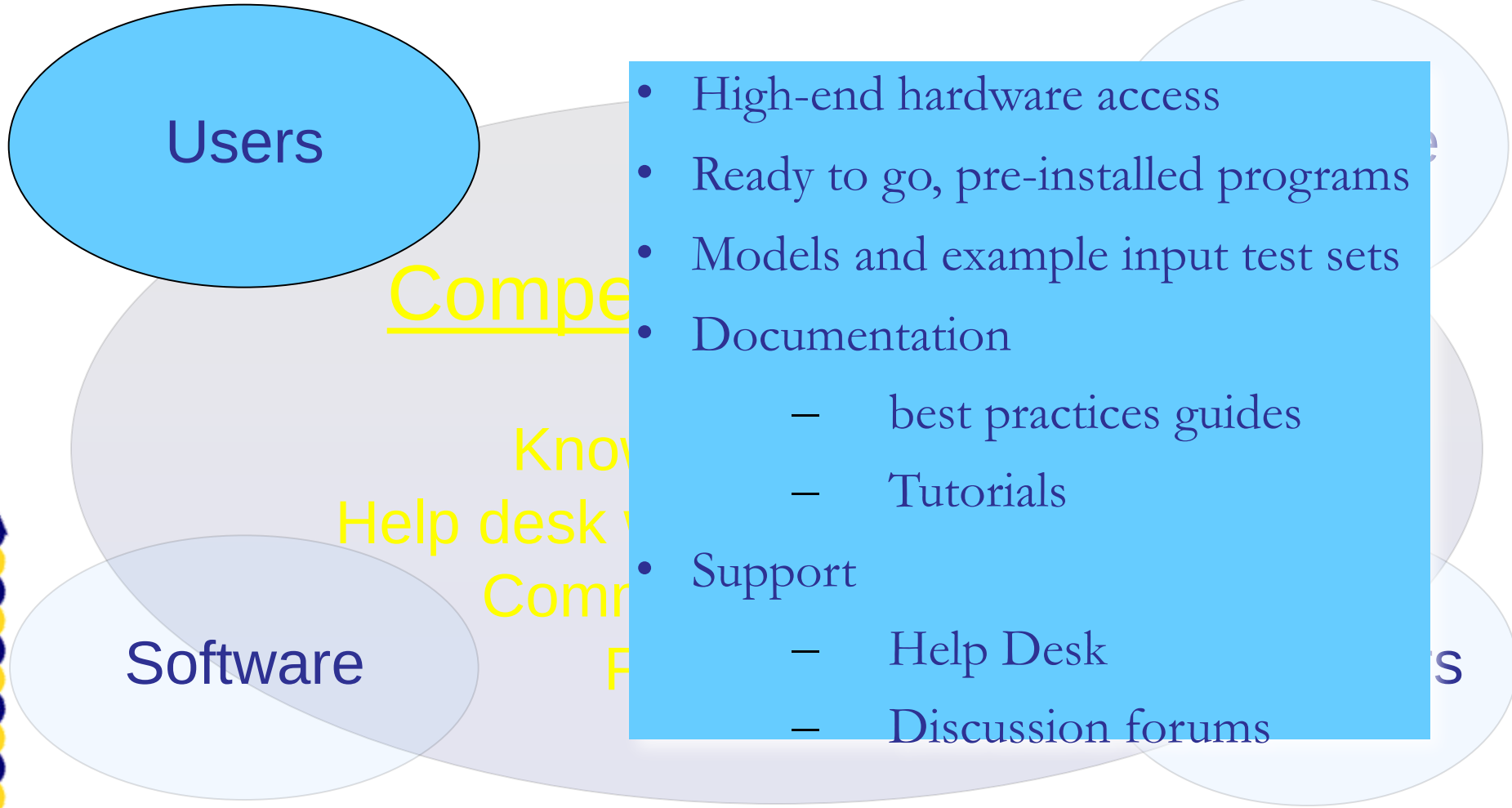
www.ScalaLife.eu

Scalable Software Services for Life Science



"One-stop-shop" for Life Science Software Communities

"One-stop-shop" for Life Science Software Communities



- High-end hardware access
- Ready to go, pre-installed programs
- Models and example input test sets
- Documentation
  - best practices guides
  - Tutorials
- Support
  - Help Desk
  - Discussion forums

"One-stop-shop" for Life Science Software Communities

Hardware

- Applications:
  - access the source code
  - test sets; best programming practices guides; performance reports on different architectures
- Development tools: Debuggers; Performance Tools etc.
  - Documentation; Best practices guides
- Algorithms:
  - Descriptions, implementations
  - Analysis of need for optimization

Developers

"O

Life Scien

Users

Com

Help de

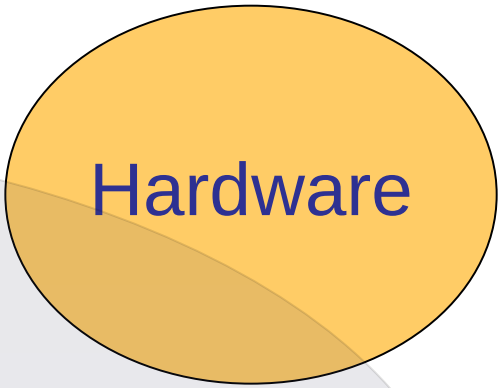
C

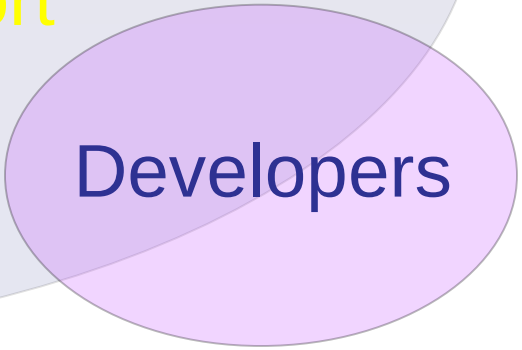
Software

- **Pilot applications**:
  - GROMACS, DALTON and DISCRETE
- **Development tools**
  - **Version control:** git, svn
  - **Build environments:** cmake, autotools
  - **Compilers:** gcc, icc, xlc
  - **Debuggers:** totalview, valgrind
  - **Performance tools:** valgrind, gprof, vampir, paraver, tau
  - **Programming models and libraries:** CUDA/OpenCL, MPI, COMPSs, FFTW
  - **Analysis tools**: Gromacs suite, R
  - **Visualization tools**: VMD, Molden, PyMol, Maestro

"One-stop-shop" for Life Science Software Communities

- Start with available hardware in the participating institutes
- Extend with external resources, e.g. from other EU projects such as EGI, DEISA and PRACE
- Include description of the usage of the systems; best practices
- Reports on resource usage; hardware monitoring
- Feedback and recommendations for upgrades and investment needs





Competence Centre

Services provided to LS software user communities:

- provision of software and support tools
  - pilot code binaries optimised for the Centres hardware (and similar)
  - parallel debugging tools, parallel performance tools
  - libraries and APIs generated by project
- provision of access to cutting-edge hardware
  - alpha users
  - new LS members and projects
- expert support
  - tutorials and best practice guides
  - assistance with new application profiling
  - assistance with using tools developed by ScalaLife
- training
  - events that are ScalaLife specific (e.g. workshops)
  - related (e.g. organised as part of existing events)

Competence Centre

Support structures:

- general project web portal
- knowledge base
  - the main source of information
  - provides training material, tutorials, best-practice guides, etc.
  - captures the experience gained with the pilot codes and alpha users
- discussion forums and mailing lists (general users)
  - foster interaction between communities
  - expand cooperation and collective knowledge
- help-desk (select users, communities and projects)
  - source code analysis and optimisation (new software applications)
  - custom software installation on specific hardware
  - application specific run-time issues
  - analysis of simulation results

Conclusions

ScalaLife project:

- integrates European efforts towards scalable LS software (both software exploitation and development)
- oversees collaboration between academia, industry, developers
- initiated the “one-stop shop” Competence Centre for the LS

Outlook:

- development of the Competence Centre to the point at which it is self-sustainable
  - collaboration with other LS software projects
  - integration of communities (e.g. Dalton users)
- delivery of software results from the project and their uptake by the LS community

Thank you!