



Scalable Software Services for Life Science

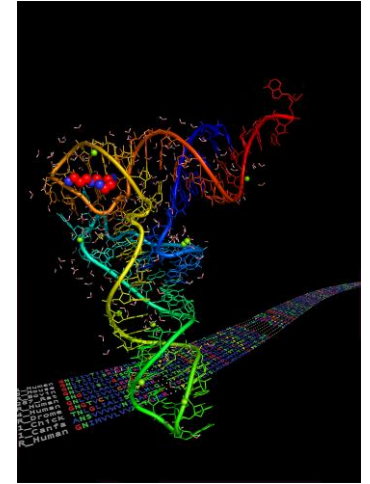
Erwin Laure

erwinl@pdc.kth.se

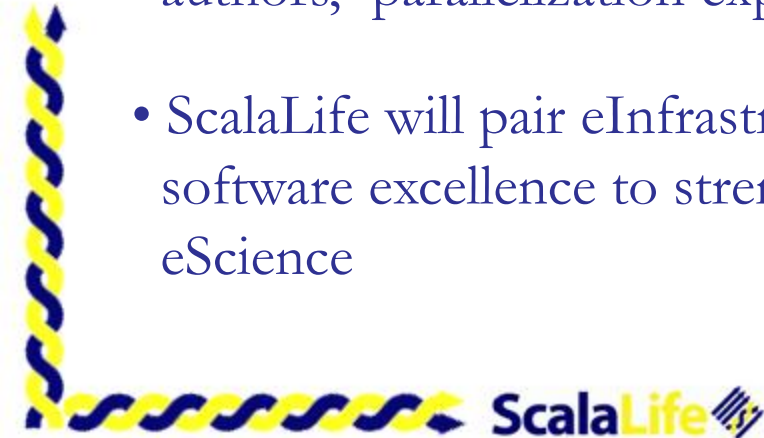
ScalaLife Project Director



- Simulation/computation is now essential in life science
- Very complex & large problems are being dealt with:
 - genomes,
 - drug design,
 - systems biology
- This puts high demands on **standardized open source software** and requires to **exploit parallelism** both on the algorithm and problem level
- ScalaLife will deliver:
 - Hierarchical parallelization of key European open source codes
 - Ensemble & high-throughput techniques for multi-core and streaming architectures to parallelize in the problem domain
 - Establish open software standards for data storage and exchange
 - Implement, document, and maintain such techniques in pilot European open-source codes – using licenses that allow code reuse
 - Create a Competence Centre to strengthen Europe as a software provider and enable the community to exploit European e-Infrastructures to their full extent. *“How should my problem be parallelized to use 1M cores, and on what hardware will it run best?”*

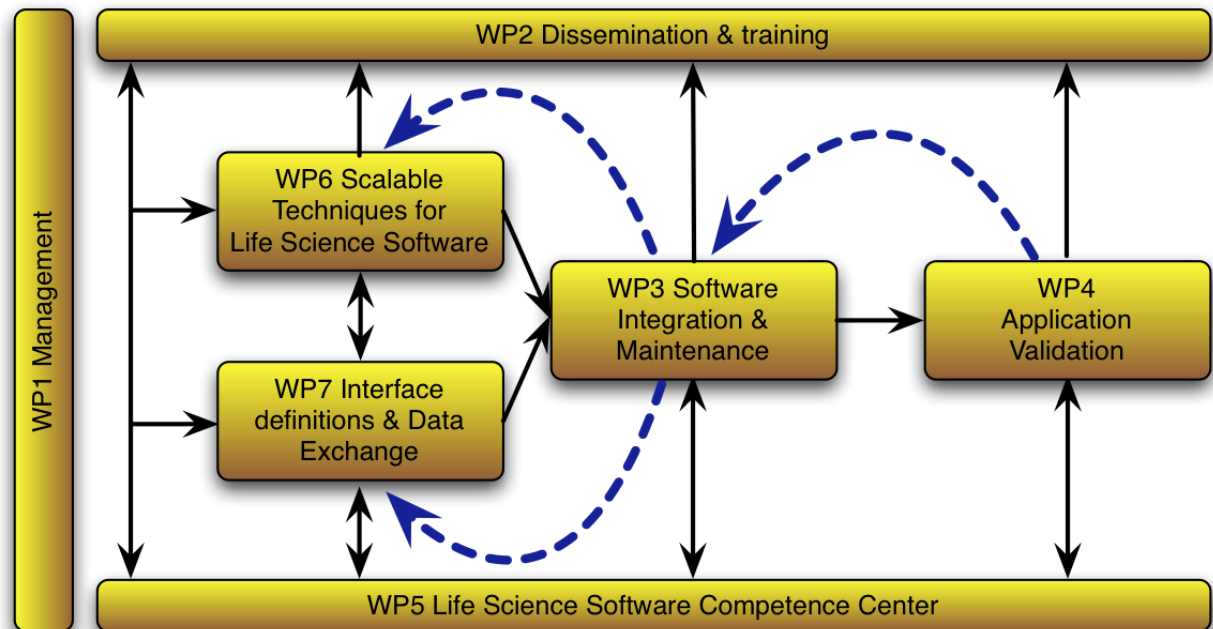


- The establishment of sustained European eInfrastructures through EGI and DEISA/PRACE asks for a concerted approach to software
- Many Life Science codes have a long history of development and large amounts of previously invested work
 - Ad-hoc efforts by individual researchers no longer sustainable
 - Standard one-size-fits-all models don't work for 2M lines of code
 - Adaptations to hardware characteristics are essential
- It is critical to joint efforts between life science software authors, parallelization experts, and infrastructure providers
- ScalaLife will pair eInfrastructure excellence with networks of software excellence to strengthen Europe's leadership in eScience



The ScalaLife Approach

- WP6 will analyze the properties of the pilot applications and develop scalable techniques
- WP7 will optimize the interfaces and data formats used for life science applications
- WP3 will integrate new concepts into the pilot applications and maintain and release them
- WP4 will validate the results by analyzing their impact on end applications running on European e-infrastructures
- WP2 will widely disseminate the results and train members of the European life science software communities
- WP5 will collect and make widely available the knowledge and best practices developed by the project to the community and provide expert services



- The ScalaLife Consortium includes the software authors involved to make sure results get used long-term, and to cover different types of parallelization requirements
 - KTH (Dalton, Gromacs), IRB (Discrete), LRZ (Bioinformatics)
- Parallelization experts find bottlenecks and new ideas, eScience infrastructure centers take responsibility for validation and user training
 - KTH (PDC), BSC, LRZ, OeRC
- Industrial partners that are experts on new hardware (GPUs)
 - Synective Labs

