e-infrastructure

CAPACITIES

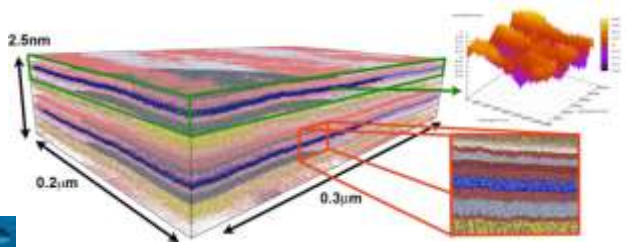
# Distributed Multiscale Computing
# The Mapper project

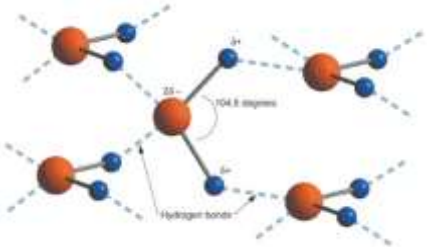
Alfons Hoekstra

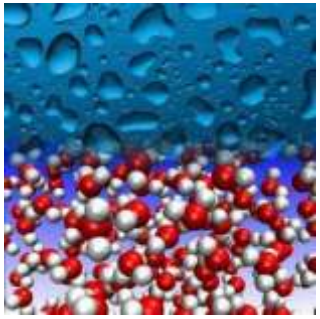
The Mapper project receives funding from the EC's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° RI-261507.

# Nature is Multiscale

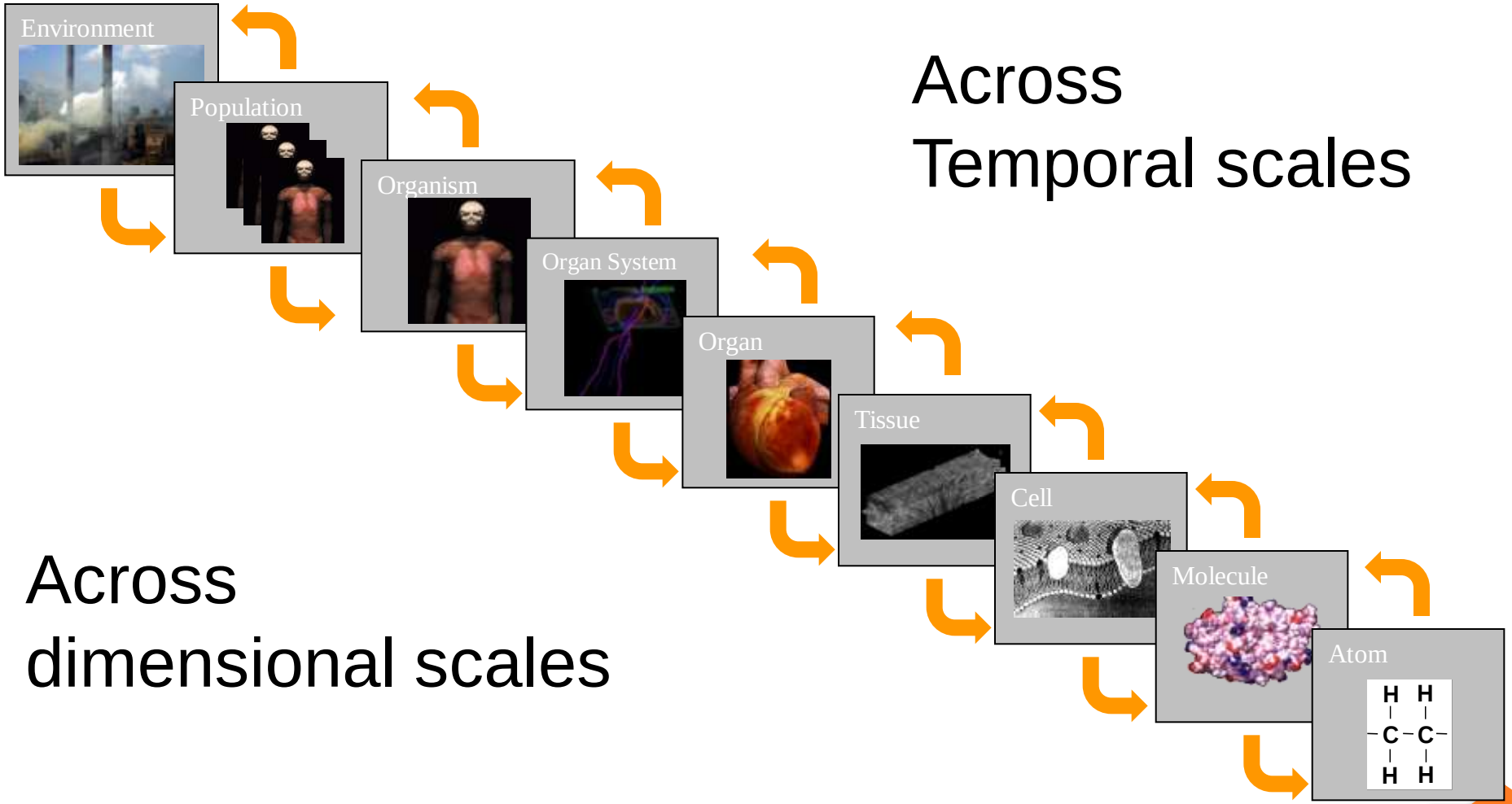
- Natural processes are multiscale
  - 1 $H_2O$ molecule
  - A large collection of $H_2O$ molecules, forming H-bonds
  - A fluid called water, and, in solid form, ice.







# Multiscale models in Biomedicine




Across Temporal scales

Across dimensional scales

A.G. Hoekstra and P.M.A. Sloot, *Multiscale Biomedical Computing*, Briefings in Bioinformatics 11, 142-152, 2010

# From Molecule to Man

(or, from DNA to Disease)

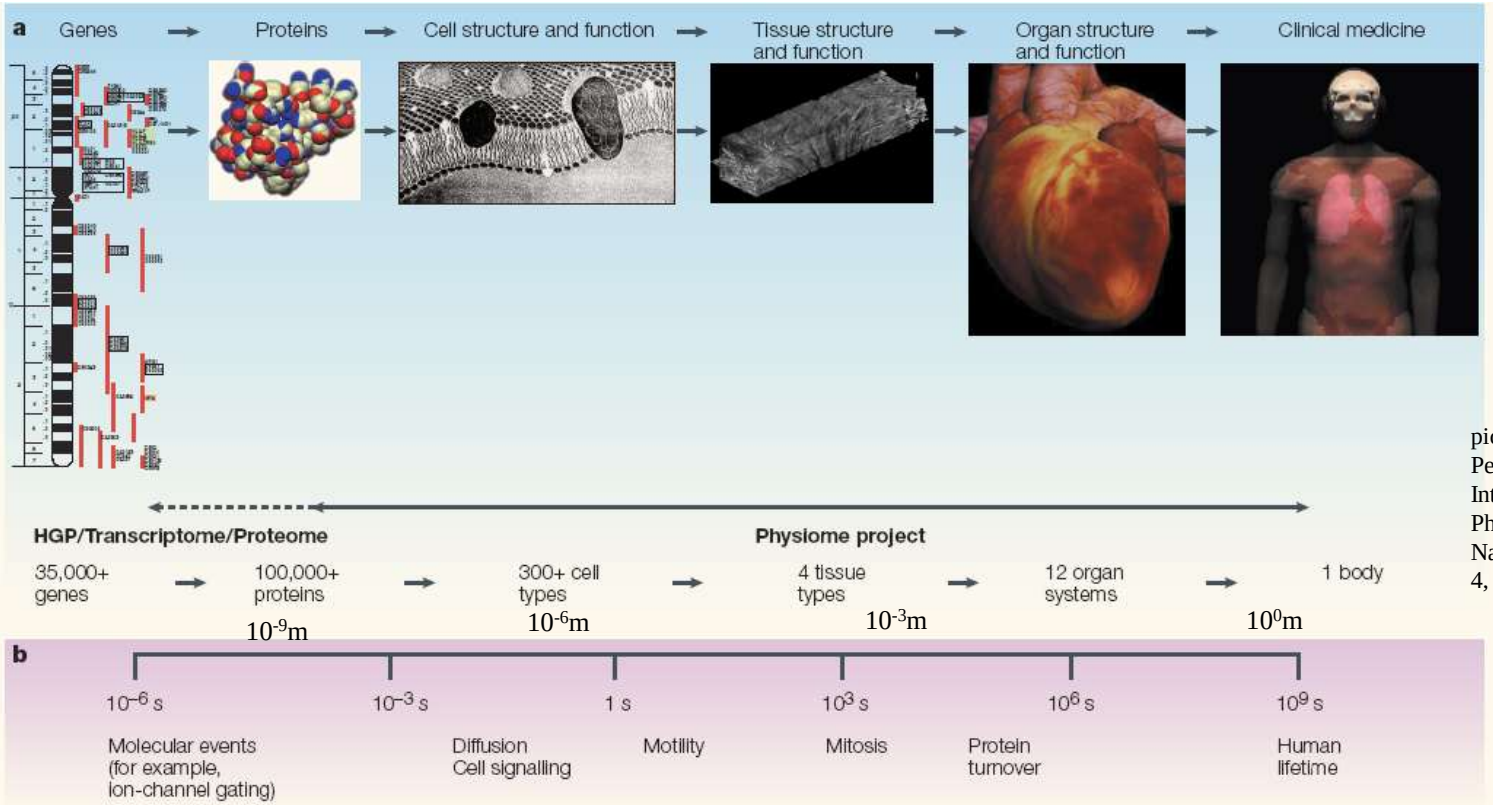



Figure 2 | **Linking molecular and cellular events with physiological function must deal with wide ranges of length scales and timescales. a** | Levels of biological organization from genes to proteins, cells, tissues, organs and finally the whole organism. The range of spatial scales — from ~1 nm for proteins to ~1 m for the whole body — requires a hierarchy of models. Different types of model are appropriate to each level, and relationships must be established between models at one level and the more detailed, but spatially or temporally limited, models at the level below. The organ-level and whole-body-level models shown are the Auckland heart and torso models, respectively[32,38]. The tissue figure is a reconstructed three-dimensional confocal image of a transmural section of rat myocardium, which is also from the Auckland Bioengineering Institute, New Zealand[28]. **b** | The range of temporal scales as shown here is even more daunting and again calls for a hierarchy of models. HGP, human genome project. Modified with permission from REF. 39 © Springer-Verlag (2002).

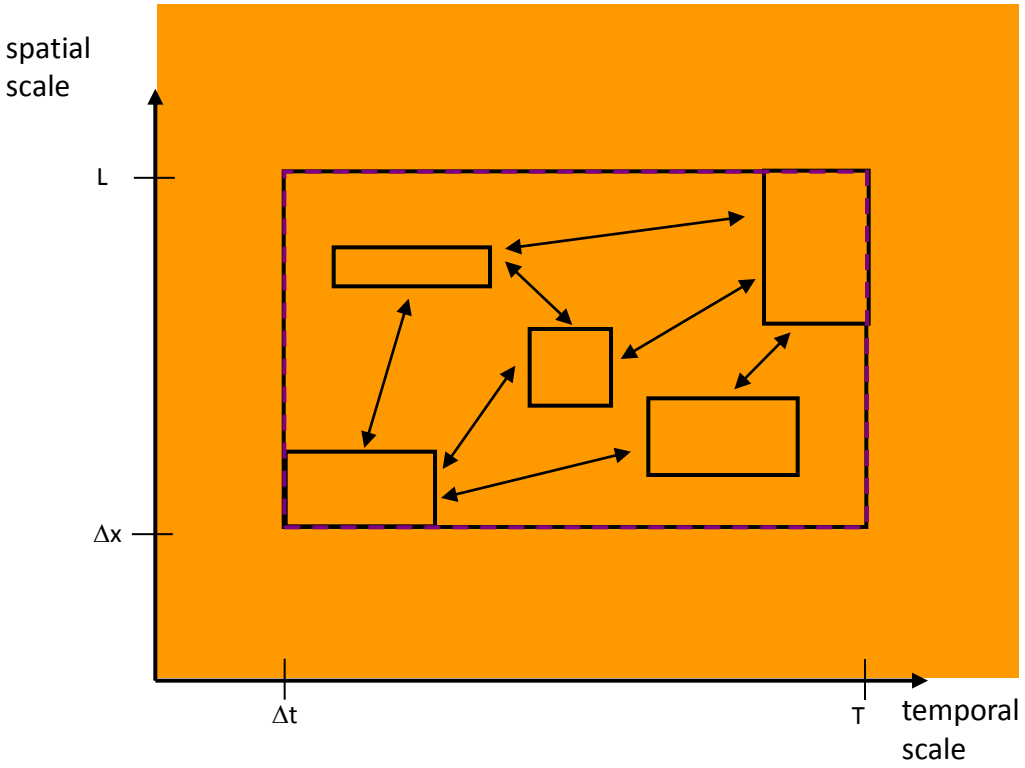
picture taken from:
Peter J. Hunter and Thomas K. Borg,
Integration from Proteins to Organs, the Physiome Project,
Nature Reviews Molecular Cell Biology, 4, 237-243, 2003

# Scale range for biomedical applications

- Temporal
  - Molecular events $O(10^{-6})$ s
  - Human life time $O(10^{9})$ s
  - A range of $\mathbf{10^{15}}$
- Spatial
  - Macro molecules $O(10^{-9})$ m
  - Size of human $O(10^{0})$ m
  - A range of $\mathbf{10^{9}}$

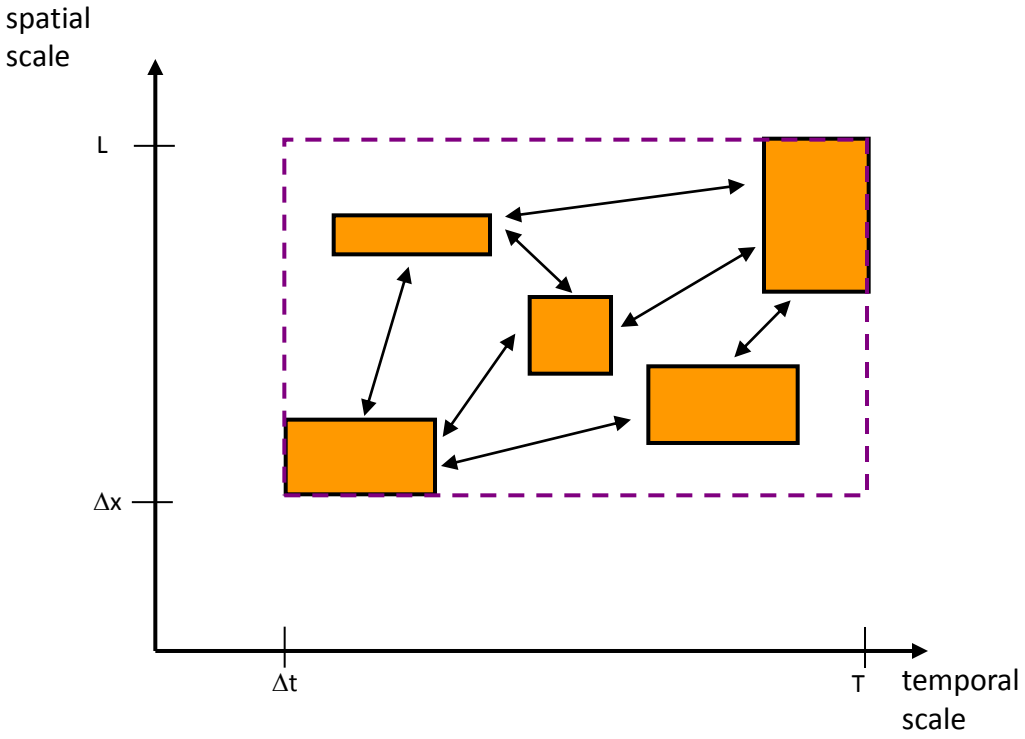
# Multi-Scale modeling

- Scale Separation Map
- Nature acts on all the scales
- We set the scales
- And then decompose the multiscale system in single scale sub-systems
- And their mutual coupling



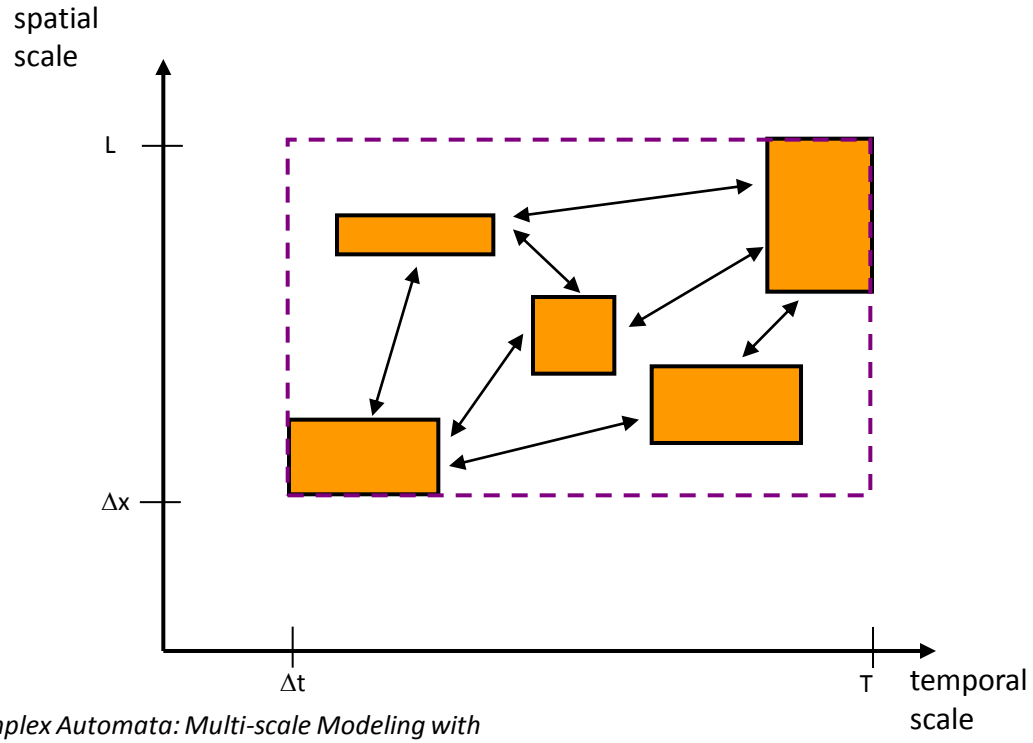
# From a Multi-Scale System to many Single-Scale Systems

- Identify the relevant scales
- Design specific models which solve each scale
- Couple the subsystems using a coupling method



# Single Scale Models

- Any model.
- Special case, Cellular Automata, leading to the paradigm of Complex Automata.




Hoekstra, A., A. Caiazzo, E. Lorenz, J.-L. Falcone, and B. Chopard, *Complex Automata: Multi-scale Modeling with Coupled Cellular Automata, in Simulating Complex Systems by Cellular Automata, A.G. Hoekstra, J. Kroc, and P.M.A. Sloot, Editors. 2010, Springer Berlin / Heidelberg. p. 29-57.*

Hoekstra, A.G., E. Lorenz, J.-L. Falcone, and B. Chopard, *Towards a Complex Automata Framework for Multi-scale Modeling. International Journal for Multiscale Computational Engineering, 2007.* ***5(6): p. 491-502.***

# Why multiscale models?

- There is simply no hope to computationally track complex natural processes at their finest spatio-temporal scales.
  - Even with the ongoing growth in computational power.

# Minimal demand for multiscale methods

$$\frac{\text{cost of multiscale solver}}{\text{cost of fine scale solver}} << 1$$

$$\text{errors in quantities of interest} < tol$$

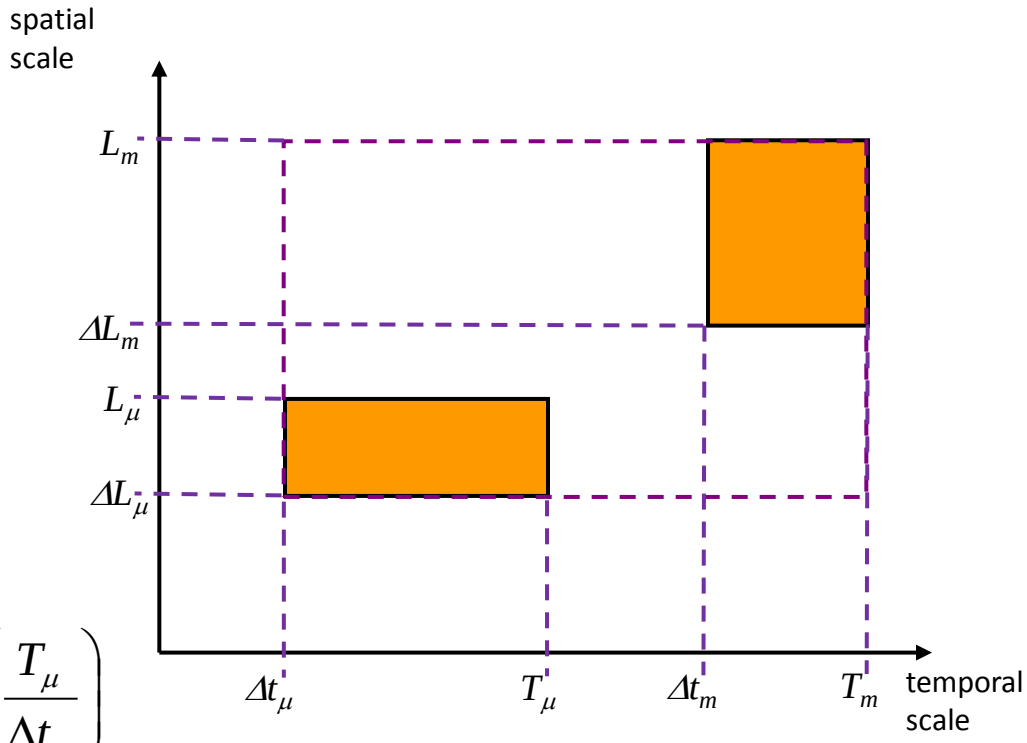
# Multiscale Speedup

- 1 microscale and one macroscale process
  - At each iteration of the macroscale, the microscale is called
- Execution time full fine scale solver

$$T_{ex}^{full} = \left(\frac{L_M}{\Delta x_\mu}\right)^D \left(\frac{T_M}{\Delta t_\mu}\right)$$

- Execution time for multiscale solver

$$T_{ex}^{multiscale} = \left(\frac{L_M}{\Delta x_M}\right)^D \left(\frac{T_M}{\Delta t_M}\right) \left(\frac{L_\mu}{\Delta x_\mu}\right)^D \left(\frac{T_\mu}{\Delta t_\mu}\right)$$




- Multiscale speedup

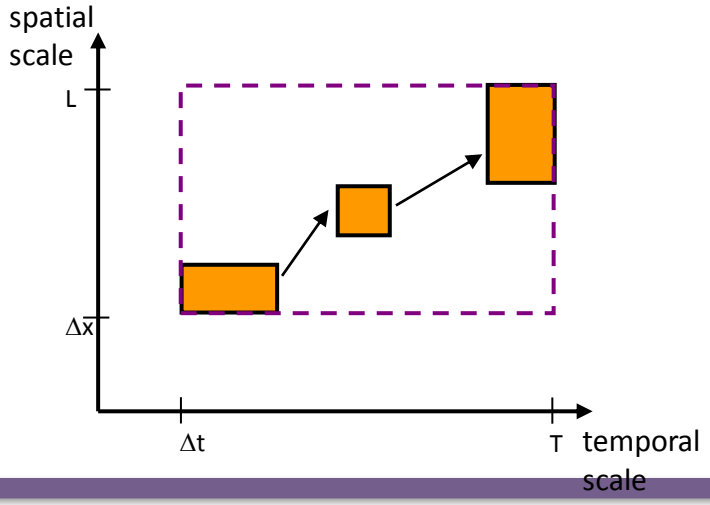
$$S^{multiscale} = \frac{T_{ex}^{full}}{T_{ex}^{multiscale}} = \left(\frac{\Delta x_M}{L_\mu}\right)^D \left(\frac{\Delta t_M}{T_\mu}\right)$$

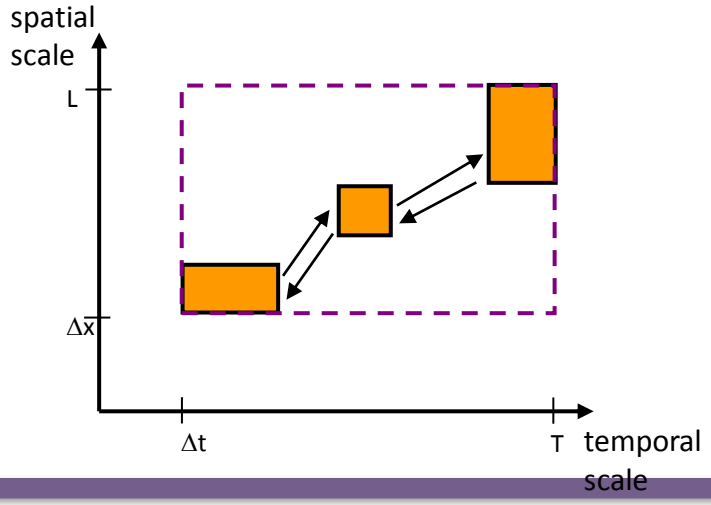
# But what about multiscale computing?

- Inherently hybrid models are best serviced by different types of computing environments
- When simulated in three dimensions, they usually require large scale computing capabilities.
- Such large scale hybrid models require a distributed computing ecosystem, where parts of the multiscale model are executed on the most appropriate computing resource.
- Distributed Multiscale Computing

# Two Multiscale Computing paradigms

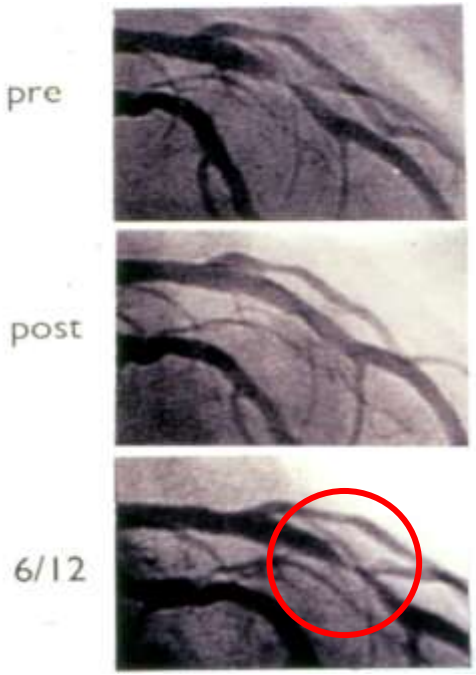
- Loosely Coupled
  - One single scale model provides input to another
  - Single scale models are executed once
  - workflows
- Tightly Coupled
  - Single scale models call each other in an iterative loop
  - Single scale models may execute many times
  - Dedicated coupling libraries are needed



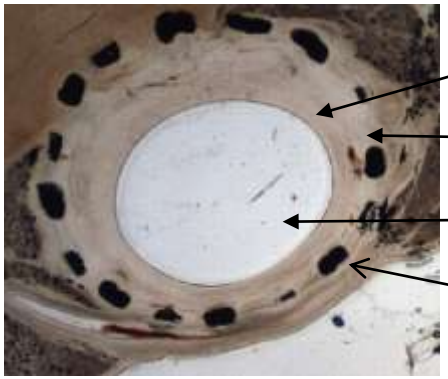



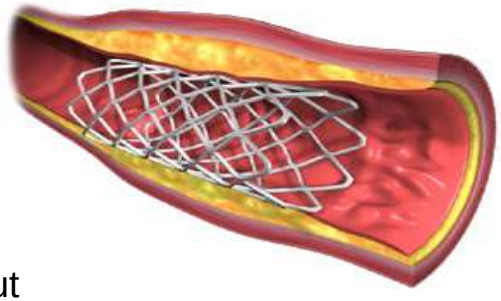
# Example 1: In-stent Restenosis

- Maladaptive response after balloon angioplasty and stenting



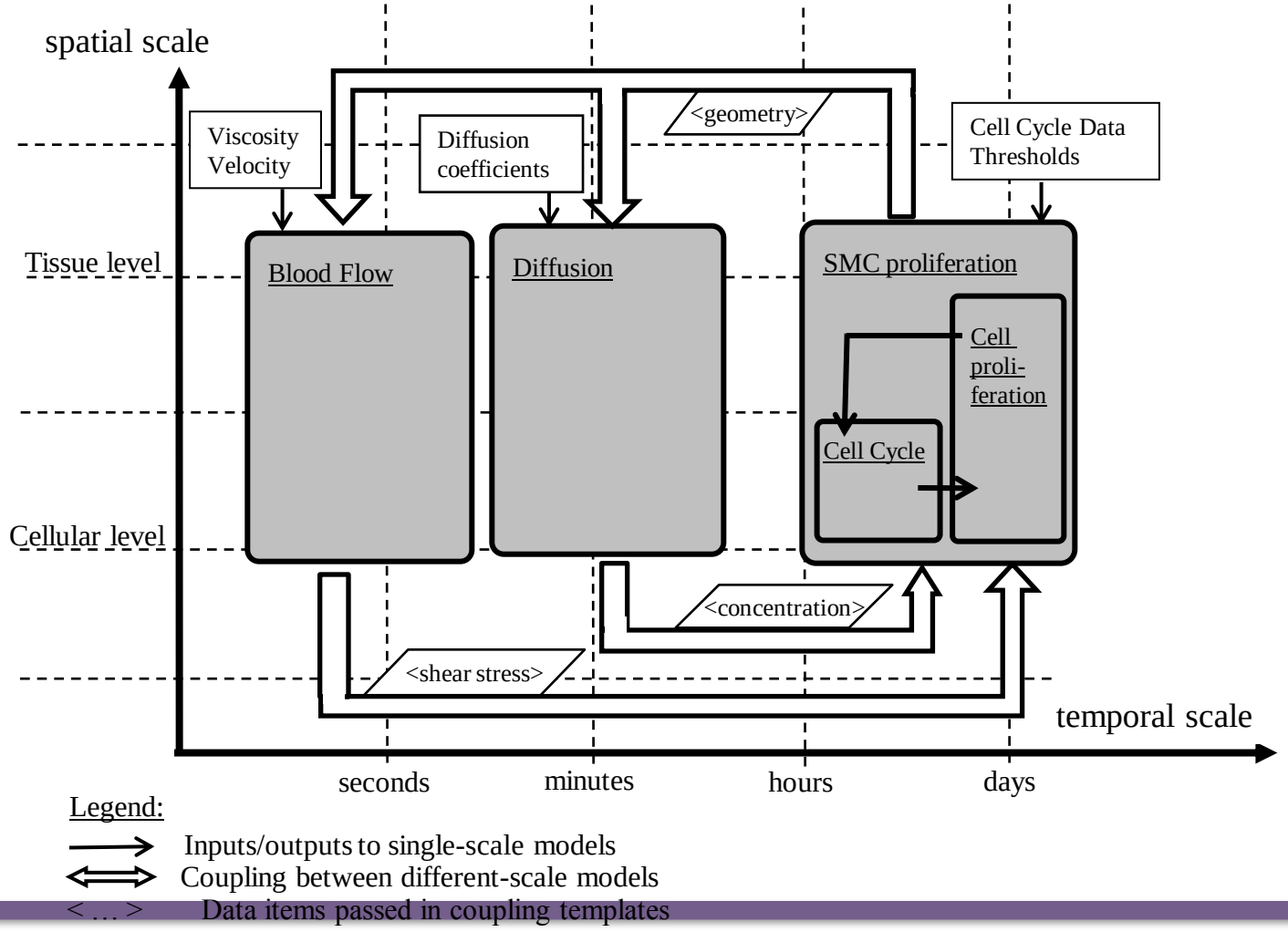

Human angiogram depicting restenosis six months post-PCI.



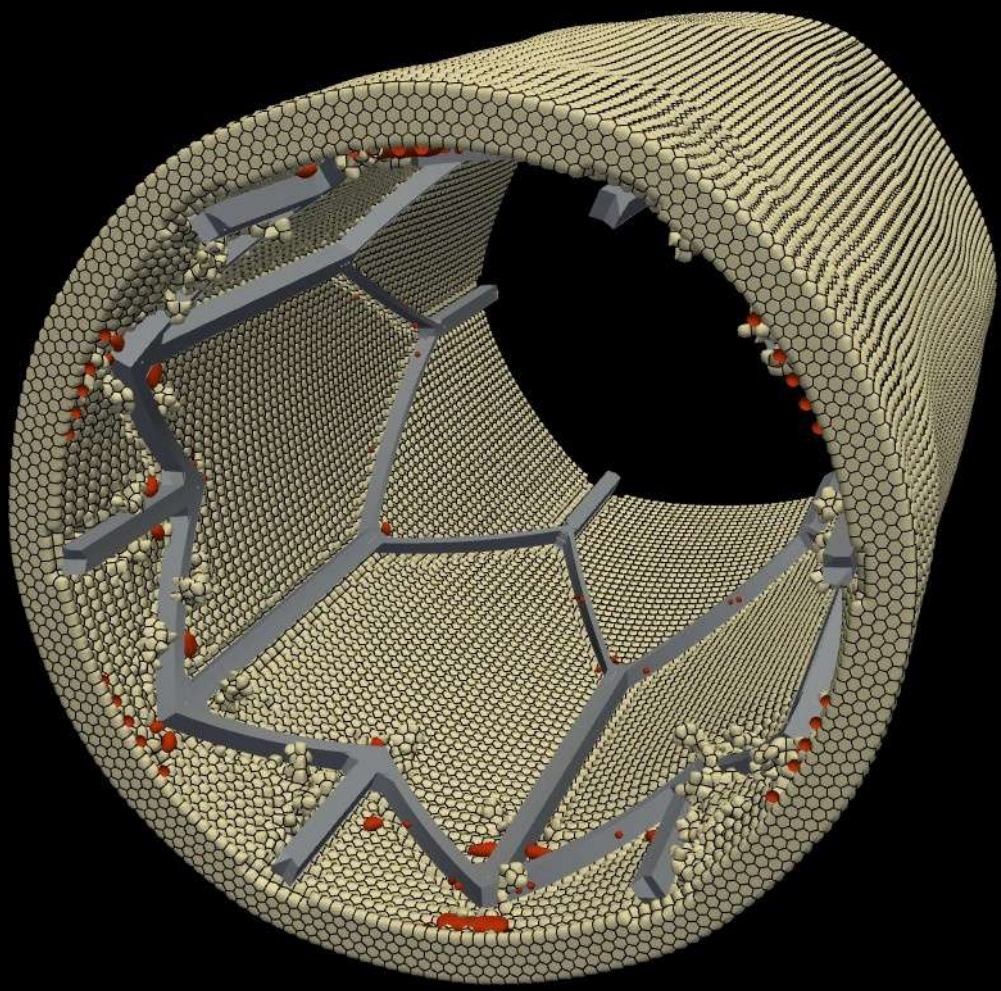
Porcine coronary artery section 28 days post stenting displaying substantial neointima.

# Simplified Scale Separation Map for ISR



# Some 3D results

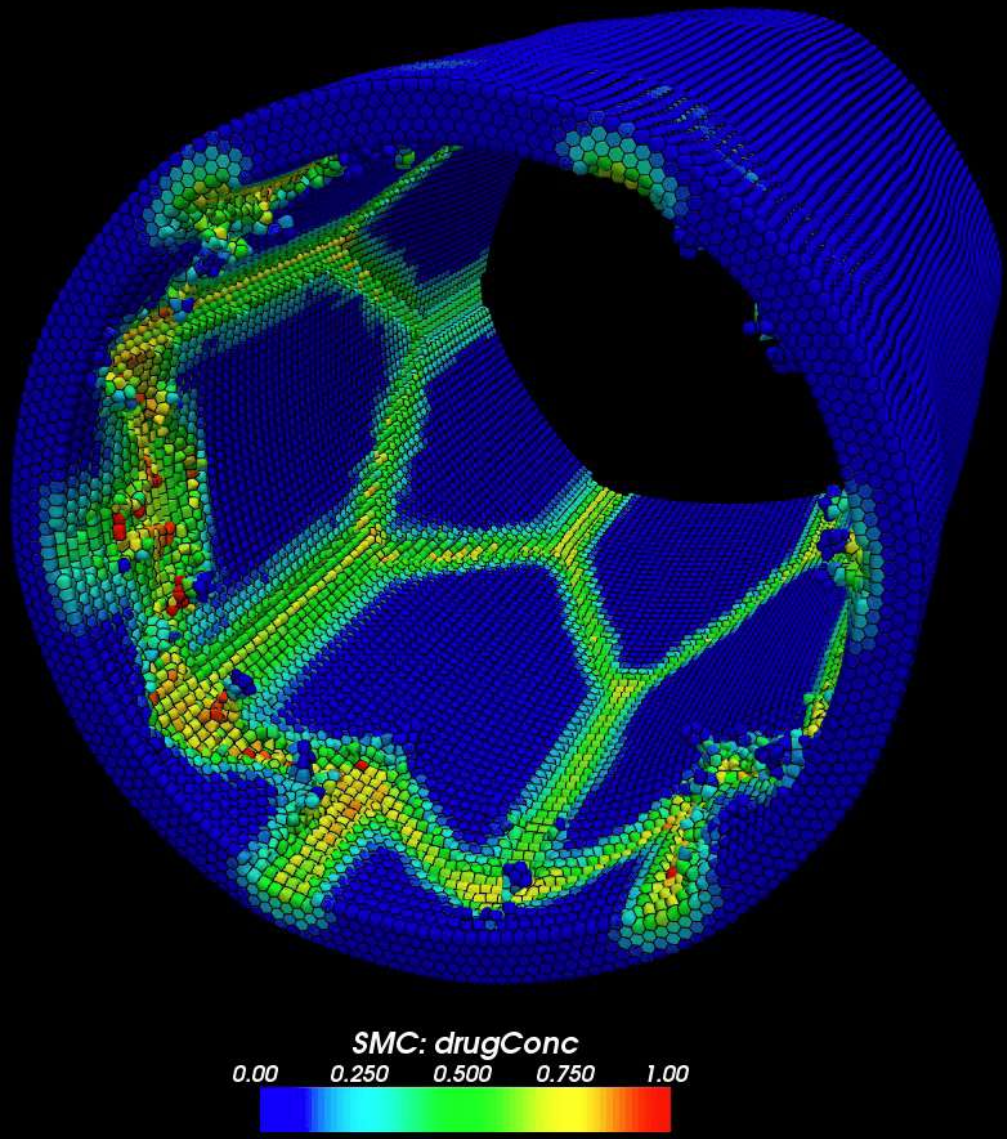
SMCs
Stent
Thrombus



Visualisations:
-- SMC Voronoi tesselation
- fill space with virtual cells
- selective edge smoothing
– Stent: hull triangulation
– Thrombus: isosurfaces

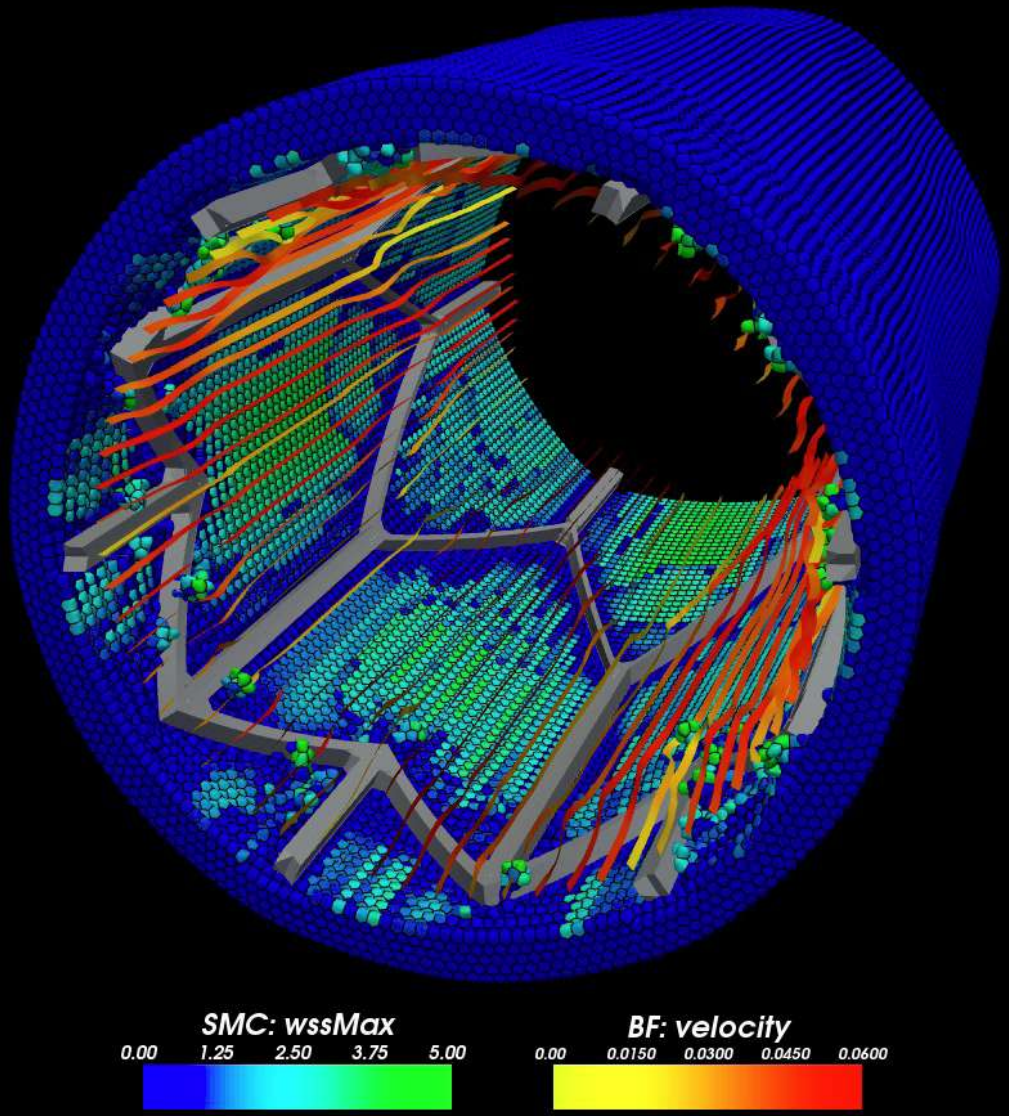
# Some 3D results

Drug concentration coloring



# Some 3D results

SMCs (WSS color scale)
Stent
Flow (Ribbons, color scale)



# Computational power needed

**Table 2: Multiscale characteristics of applications**

| Application | Loosely Coupled | Tightly Coupled | Total number of single scale models | Number of single scale models that require supercomputers |
|---|---|---|---|---|
| In-stent restenosis | | X | 5(1) | 2 |
| Coupled same-scale and multi-scale hemodynamics | | X | 3(2) | 2 |
| Multi-scale modelling of the BAXS | X | | 2(3) | 1 |
| Edge Plasma Stability | X | | 3(4) | 1 |
| Core Workflow | | X | 3-10(5) | 1-4 |
| Irrigation canals | | X | 5(6) | 1-2 |
| Clay polymers | X | | 3(7) | 2 |

(1) Blood flow, smooth muscle cell proliferation, drug diffusion, thrombus, stent deployment; Depending on state-of-the-art when starting the project; (2) HemeLB, a lattice-Boltzmann code for blood flow, NEKTAR, a FEM-based code for blood flow in large arteries, CellML models for cellular processes; (3) metabolism (Phase 1), conjugation (Phase 2) and further modification and excretion (transport) (Phase 3) of the target drug/xenobiotic/endobiotic/bile acid; (4) HELENA or equivalent plasma equilibrium code and ILSA or equivalent plasma stability code; (5) HELENA/CHEASE/EQUAL, some combination of ETAIGB/ NEOWES/ NCLASS/ GLF23/ WEILAND/ GEM, some heating modules from ICRH/NBI/ECRH/LH, some particle source modules from NEUTRALS/PELLETS, some MHD modules from SAWTEETH/NTM/ELMs (6) 1D shallow water models, 2D shallow water models, 2D Free surface flow models, 3D Free surface flow models, Sediment transport models; (7) ab initio molecular dynamics code CASTEP, atomistic molecular dynamics code LAMMPS, coarse-grained simulations also using LAMMPS;

# Multiscale APPlications on European e-infRastructures

(proposal number 261507)

## Project Overview

University of Amsterdam

University College London

University of Ulster

Poznan Supercomputing and Networking Centre

Akademia Gorniczo-Hutnicza im. Stanislawa Staszica w Krakowie

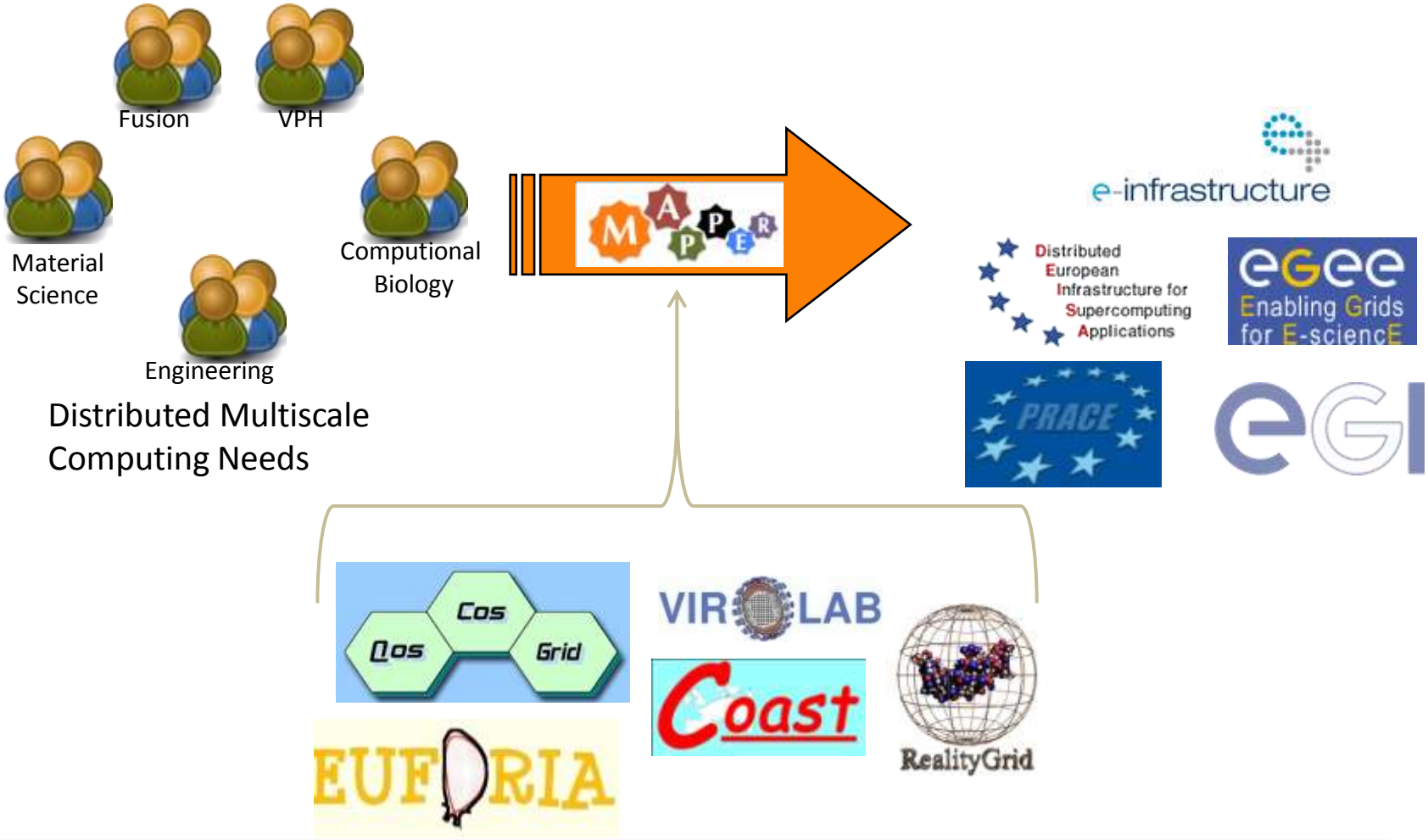
Ludwig-Maximilians-Universität München

University of Geneva

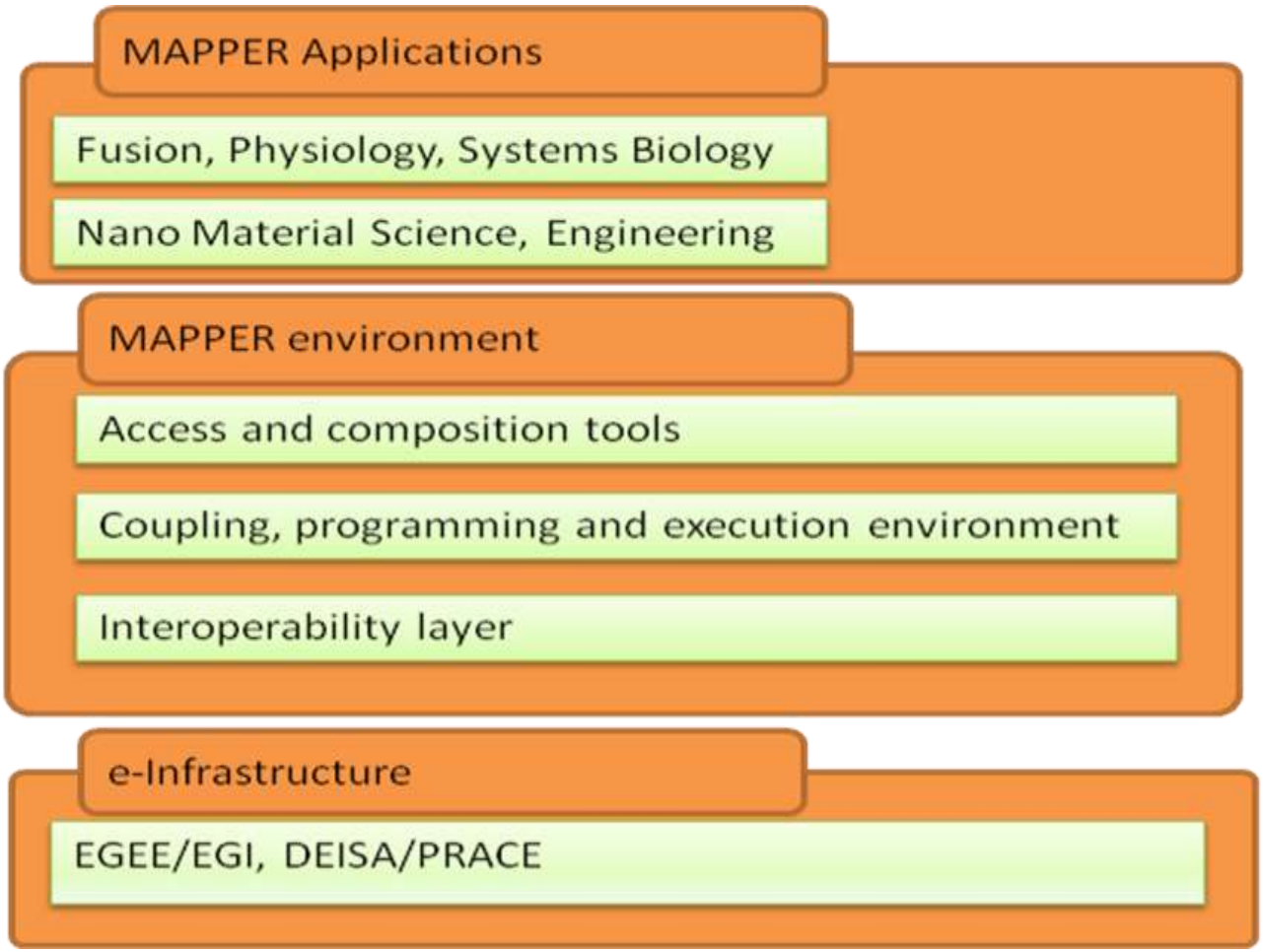
Chalmers Tekniska Högskola

Max-Planck Gesellschaft zur Foerderung der Wissenschaften E.V.

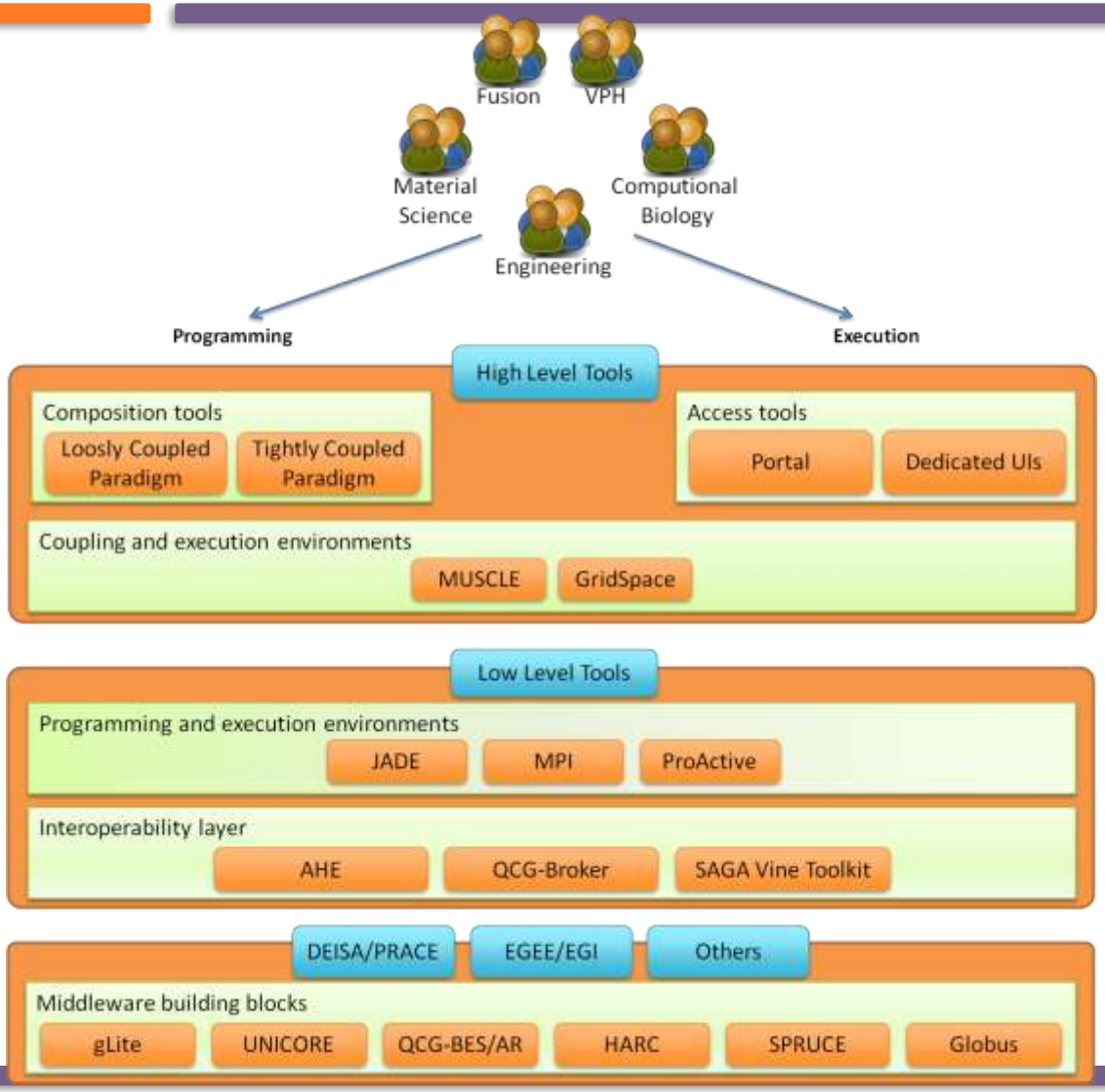
# Motivation: user needs



# Overview


MAPPER Applications
Fusion, Physiology, Systems Biology
Nano Material Science, Engineering
MAPPER environment
Access and composition tools
Coupling, programming and execution environment
Interoperability layer
e-Infrastructure
EGEE/EGI, DEISA/PRACE

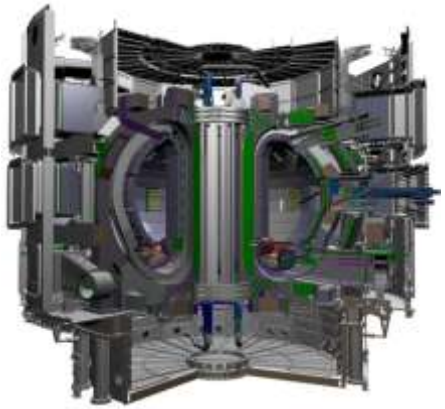
# Ambition




- Develop computational strategies, software and services
  - for *distributed multiscale simulations* across disciplines
  - exploiting existing and evolving European e-infrastructure
- Deploy a computational science infrastructure
- Deliver high quality components
  - aiming at large-scale, heterogeneous, high performance multi-disciplinary multiscale computing.
- Advance state-of-the-art in high performance computing on e-infrastructures
  - enable distributed execution of multiscale models across e-Infrastructures,

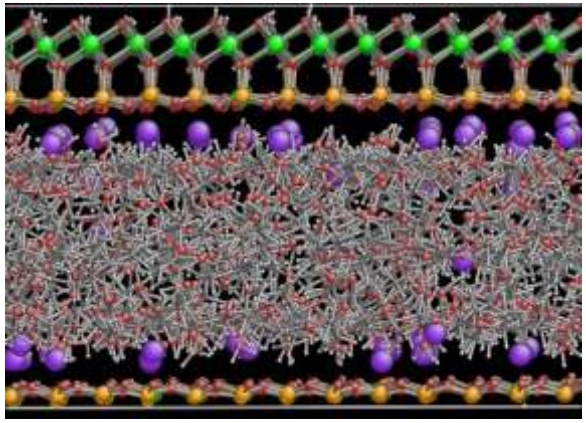
# Application Portfolio

virtual physiological human

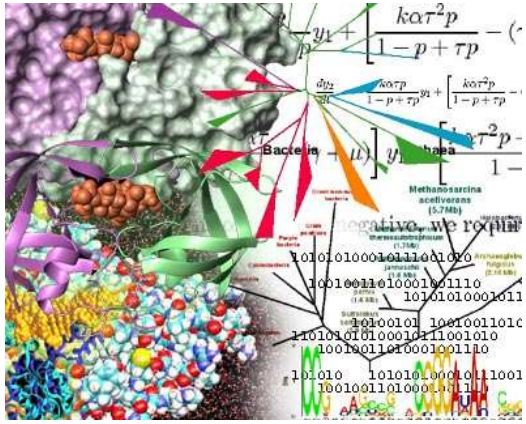


fusion

hydrology



nano material science

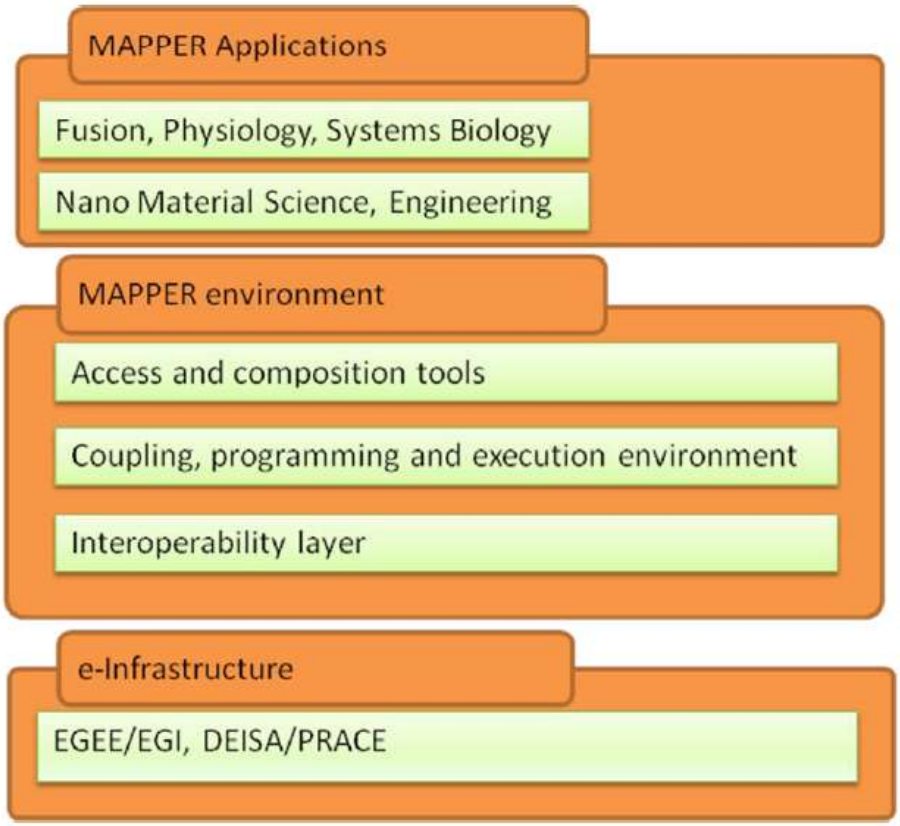


computational biology

# MAPPER Roadmap

- October 1, 2010 – start of project
- Fast track deployment – first year of project
  - Loosely and tightly coupled distributed multiscale simulations can be executed.
- Deep track deployment – second and third year
  - More demanding loosely and tightly coupled distributed multiscale simulation can be executed
  - Programming and access tools available
  - Interoperability available

# Service Activities in MAPPER



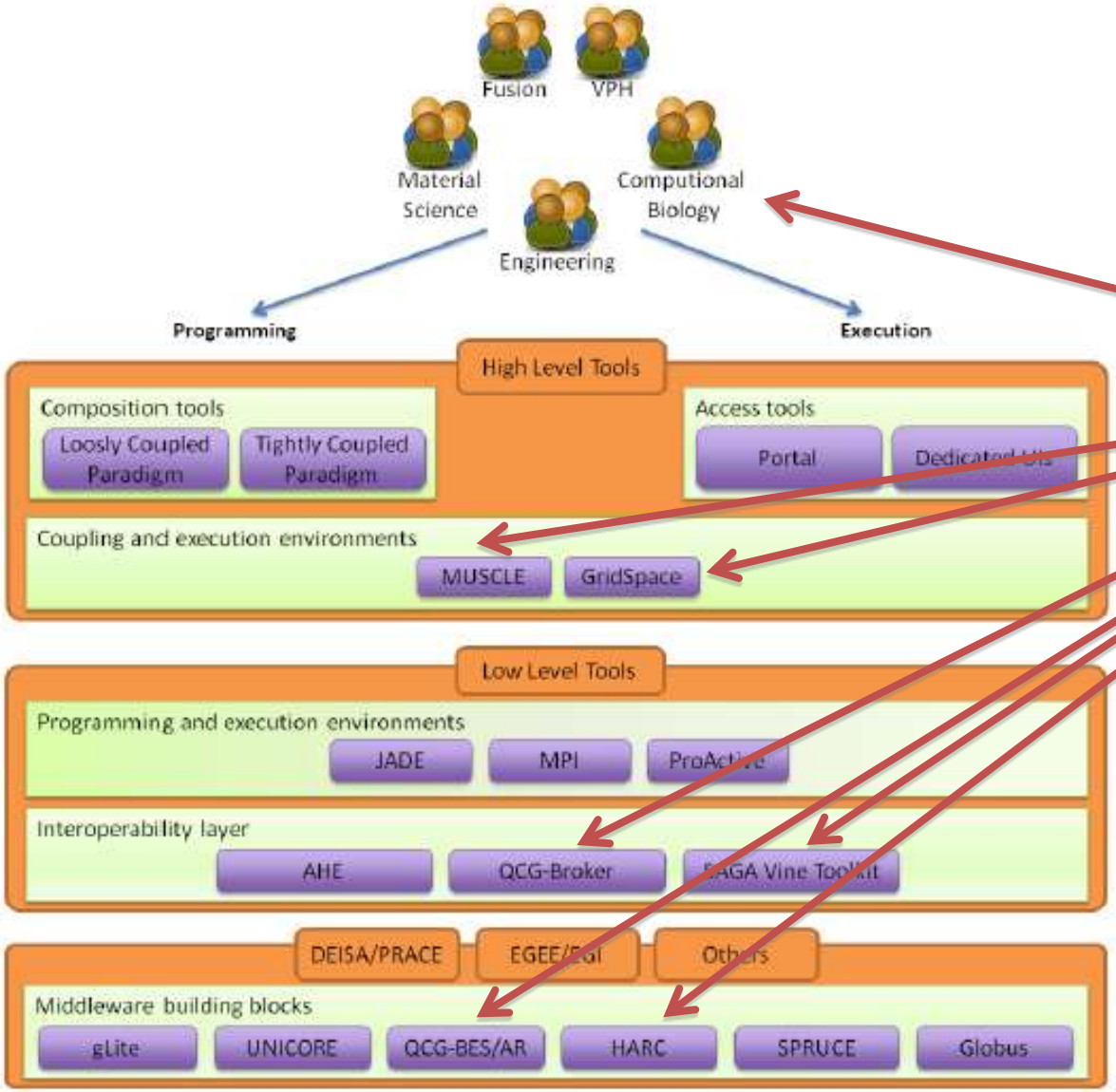

**WP7 and WP8 (JRA)**

**xMML vs. Job Profile/JSDL with extensions**

**WP4, WP5 and WP6 (SA)**

**Distributed Computing = E-Infrastructure**

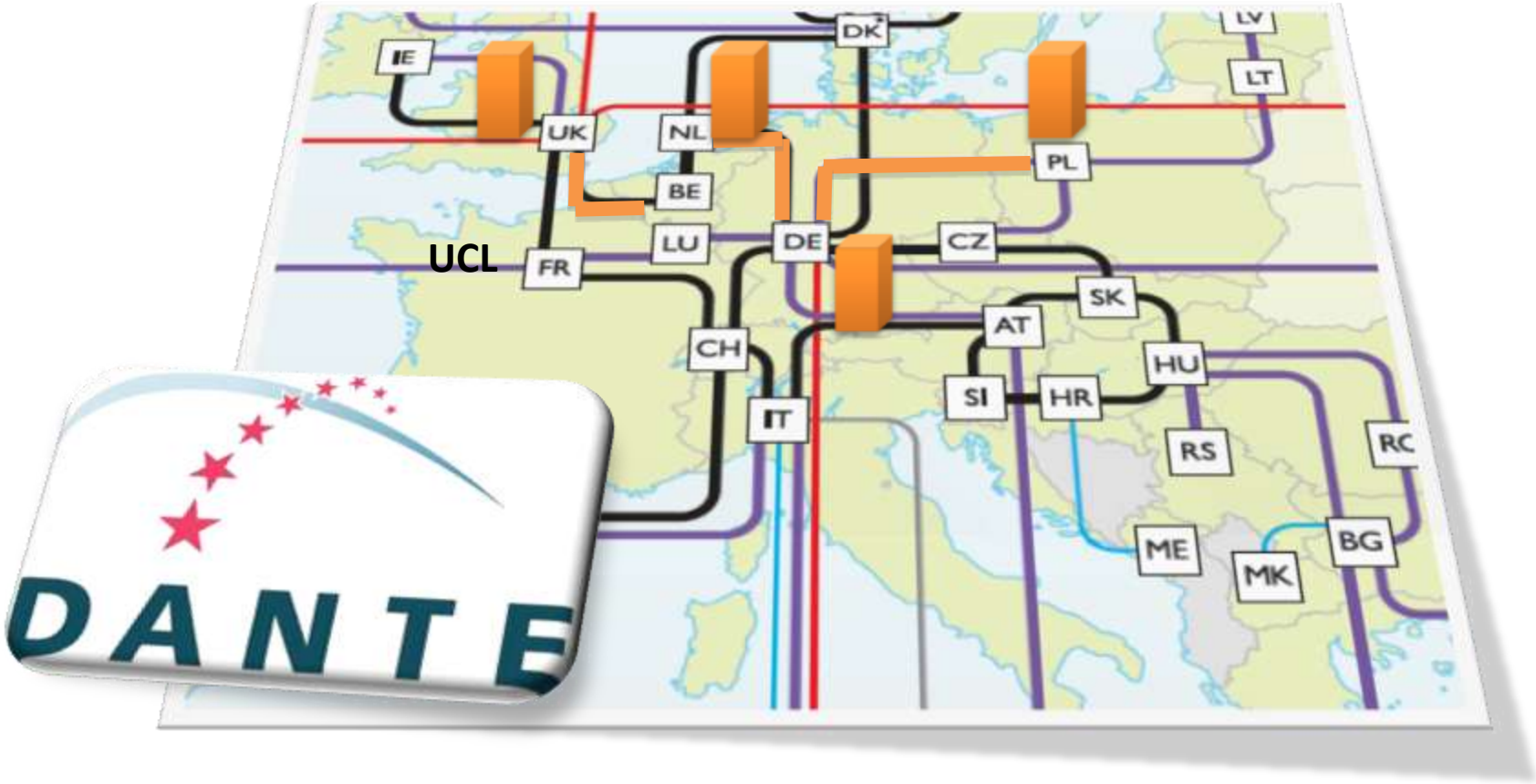
# Service Activities (WP4,5,6)




Real actions taken in SA over the last 6 months

# Computing e-Infrastructure

# Networking e-Infrastructure



# e-Infrastructure Services

- Offered:
  - Interactive access
  - Data management
  - Job execution
  - Post-processing, e.g. visualization
- Not available:
  - Workflow management and execution
  - Advanced reservation
  - Co-allocation

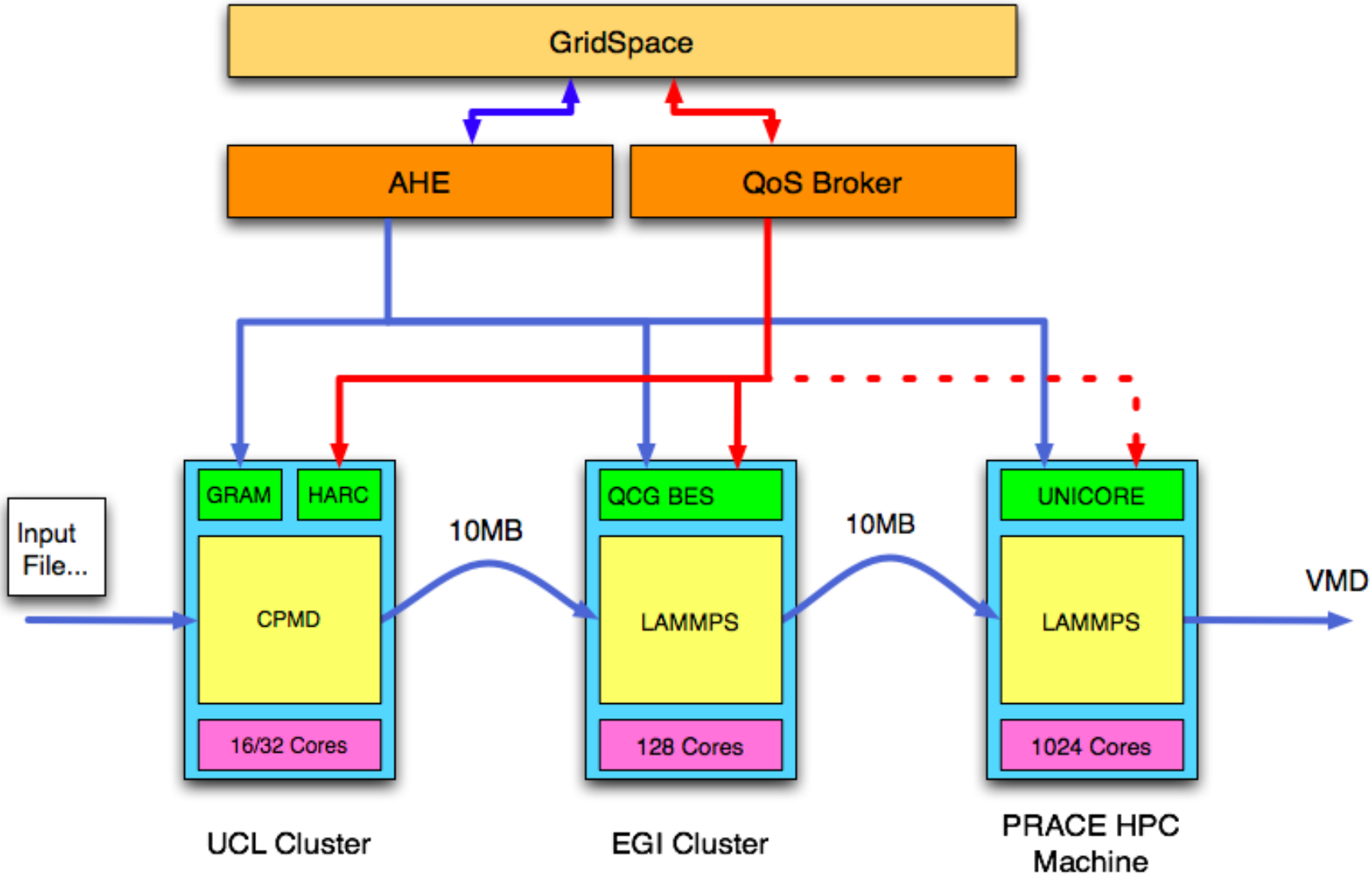
# New MAPPER components

- Fundamental blocks
  - HARC (queuing system commands via scripts)
  - QCG-BES/AR (AR extensions in DRMAA)
  - All above components are available in production !
  - MUSCLE services and other app tools (deep track)
- Interoperability services and tools
  - QosCosGrid Broker (QCG-Broker)
  - Application Hosting Environment (AHE)
  - A Simple API for Grid Applications (SAGA)
    - The obtained results will be presented this week at OGF
    - UNICORE  ver. 6.3.2, gLite ver. 3.2.0, QCG-BES/AR ver.1.1
  - Science Gateways based on Vine Toolkit

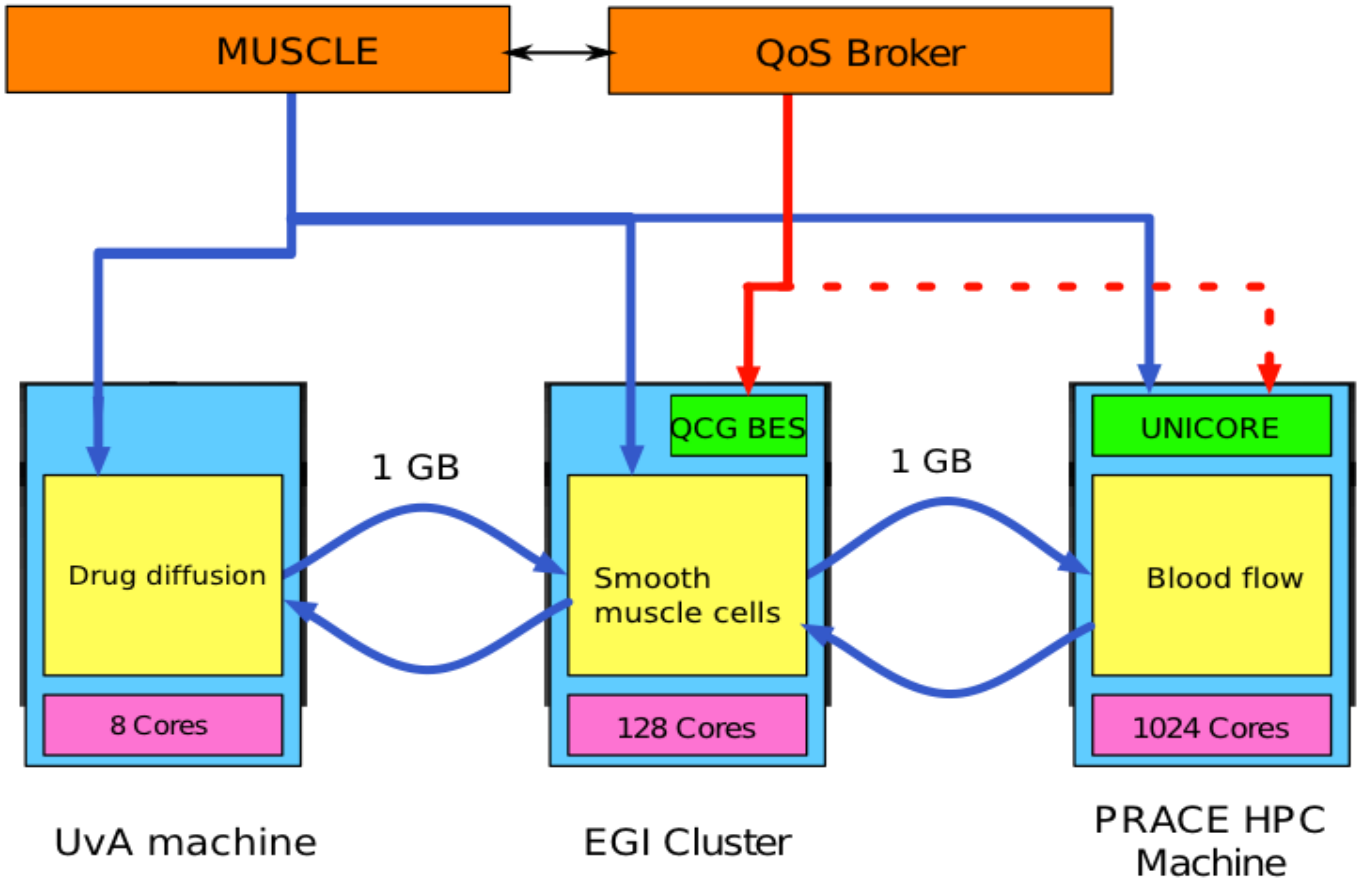
# Monitoring

- Provide information about availability and functionality of MAPPER services
- Nagios
  - Hosted at LRZ
  - Real-time service status
  - Failure notification
  - Statistics
  - Integration with EGI and PRACE monitoring

# Use case – loosely coupled



# Use case – tightly coupled



# A perfect co-allocation

# Conclusion

- Distributed Multiscale Computing
  - A relevant and important paradigm with a potential huge impact on scientific communities
- MAPPER will facilitate DMC
- Hurdles
  - Technical
  - Policies