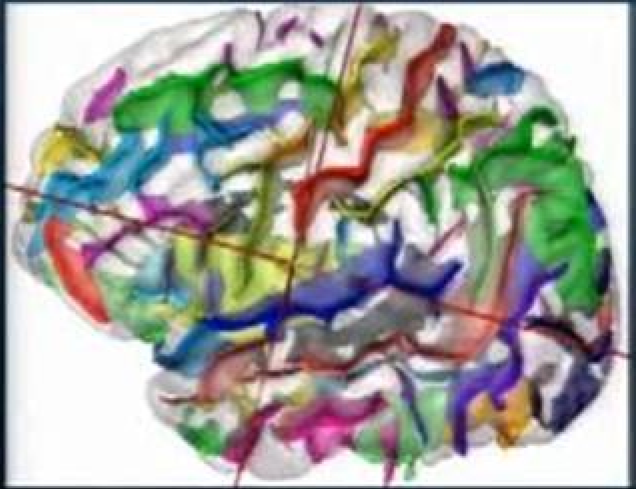
Kitware

Scaling Open Systems for Future Computational Challenges

Will Schroeder
SIAM CSE15
March 16, 2015



The Coming Era of Big Software

or

How I Learned to Stop Worrying and Love the Software Complexity Barrier

Kitware

Why Software Is Eating The World

Marc Andreessen – WSJ, 2011

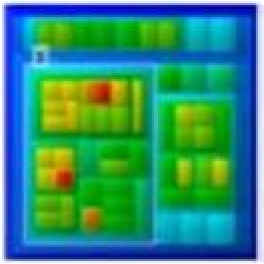
"Six decades into the computer revolution, four decades since the invention of the microprocessor, and two decades into the rise of the modern Internet, all of the technology required to transform industries through software finally works and can be widely delivered at global scale."

Agenda

- Part 1: Drivers
- Part 2: Remedies

Part 1: Drivers

- Big Data
- Growing Computational Demands
- The Parallel Computing Challenge
- Overwhelming Abundance
- Reproducibility

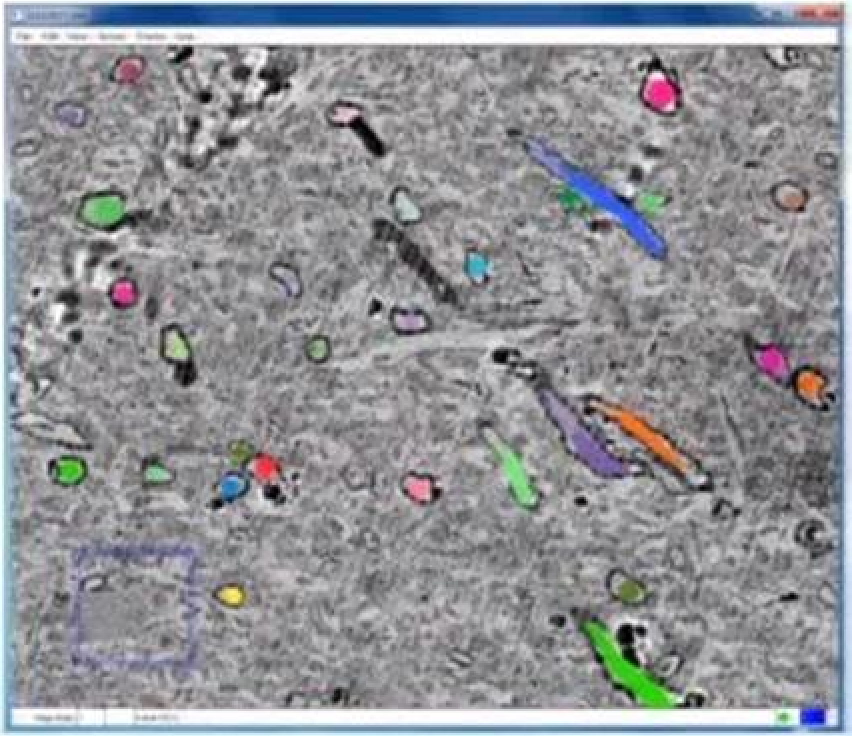


DRIVERS:

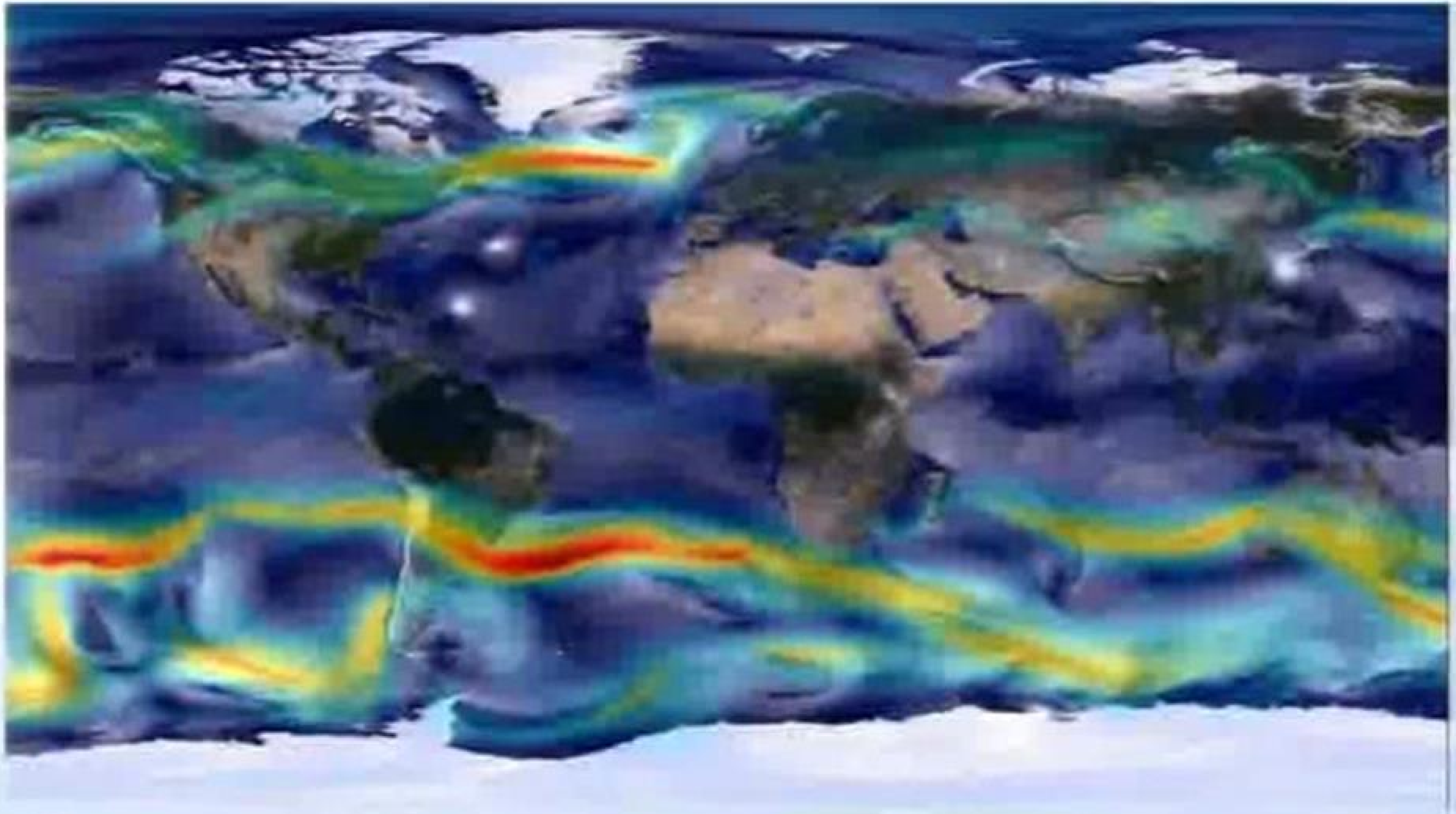
BIG DATA

Large Experimental Data

- Mouse connectome
- ~10 Nanometer resolution (Electron microscopy)
- Petabyte scale
  - per 1 cubic mm of brain tissue

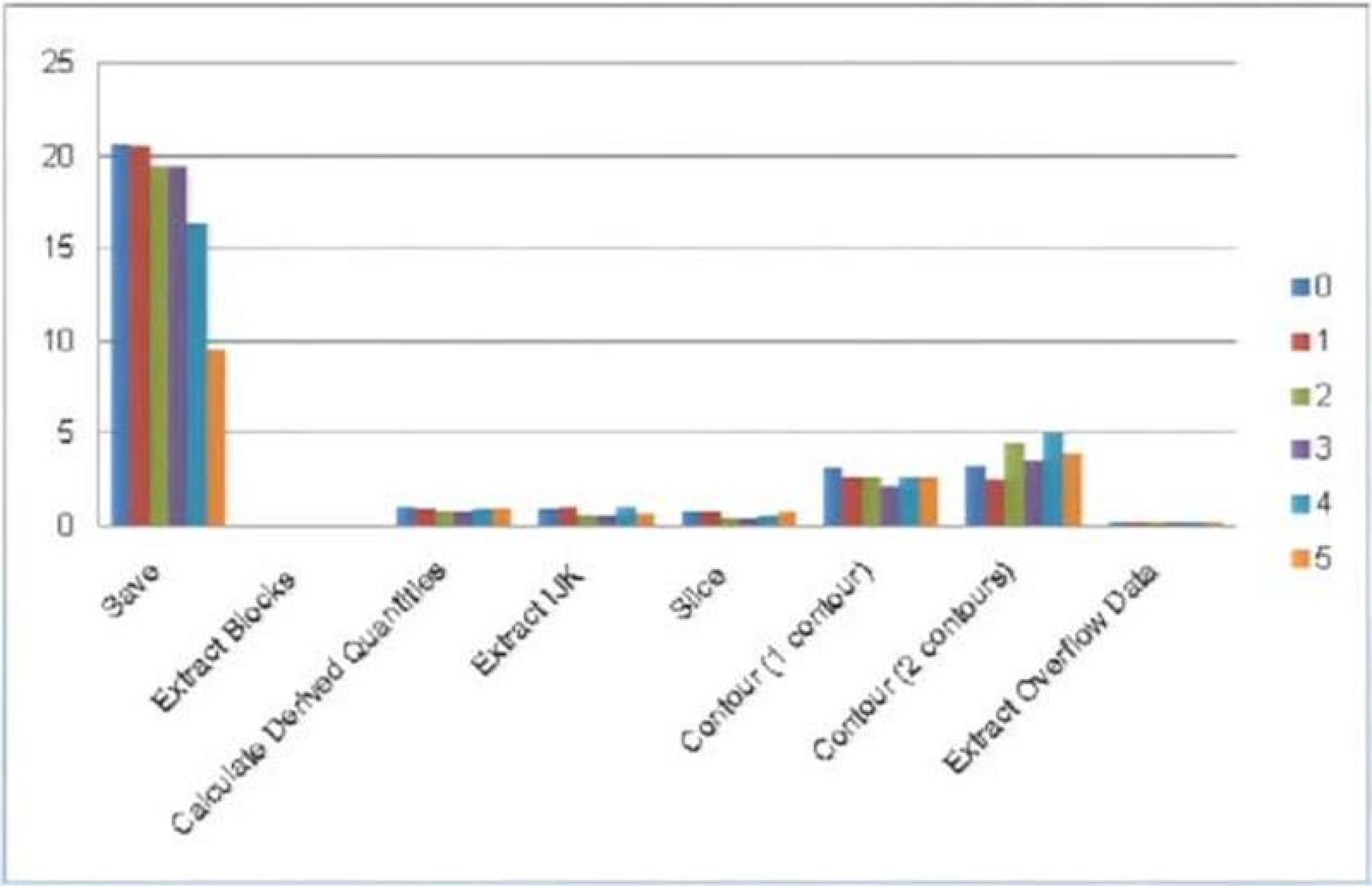


Large Computational Data



NASA

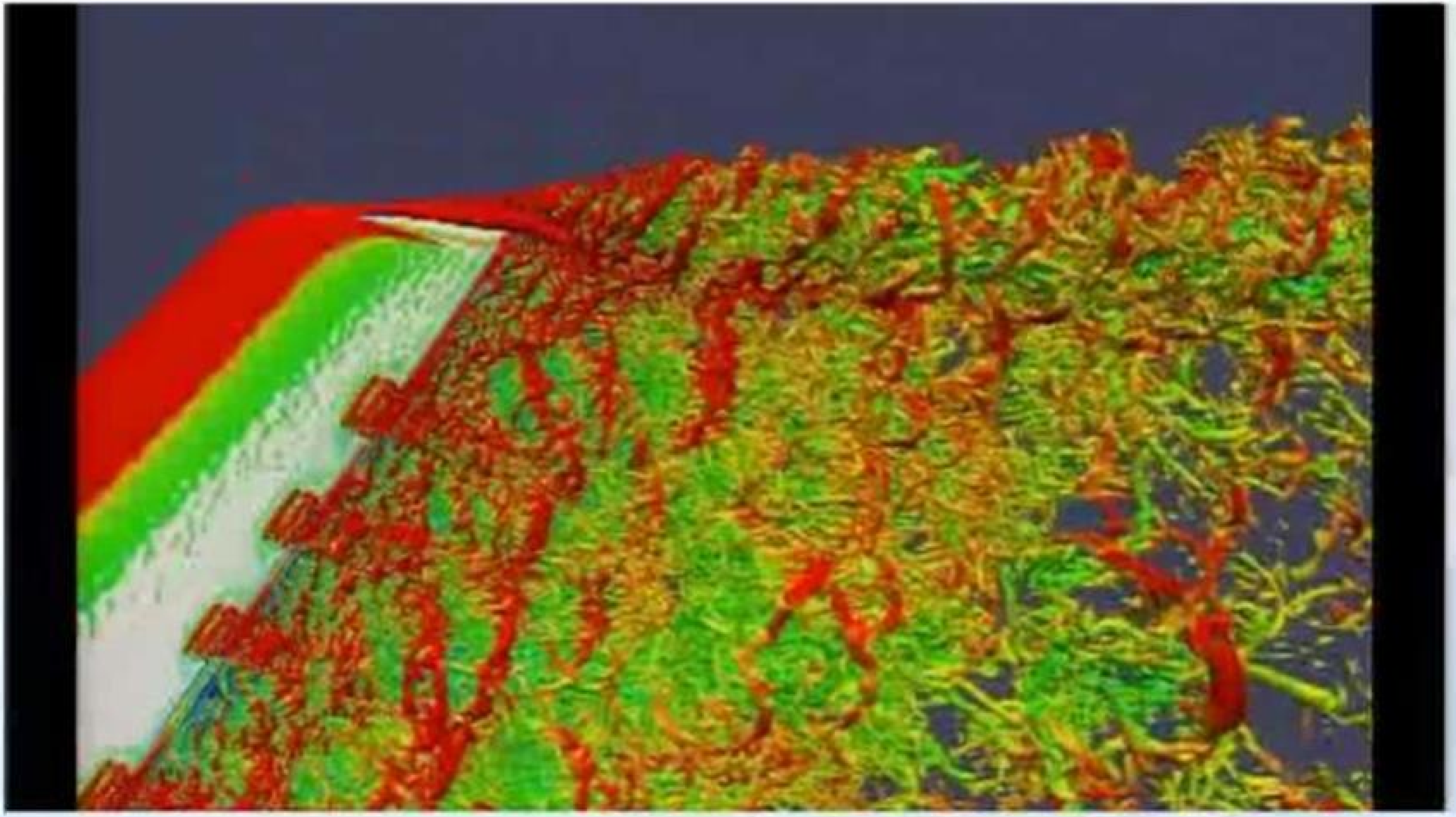
Cost of IO



Kitware

DRIVERS:

GROWING COMPUTATIONAL DEMAND

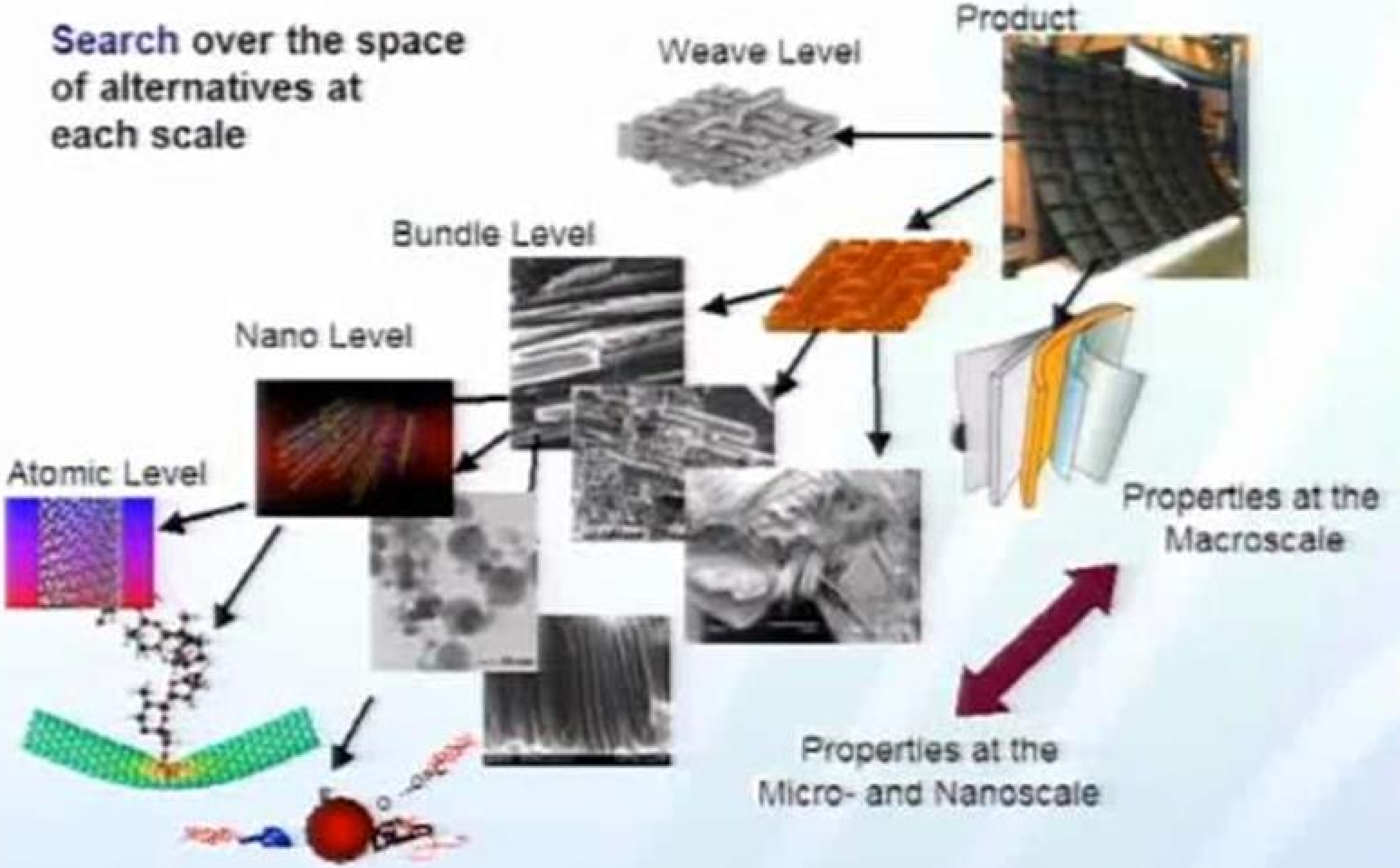


PHASTA / CATALYST Blue Gene Q / MIRA

Ken Jansen at UC Boulder, Argonne ALCF

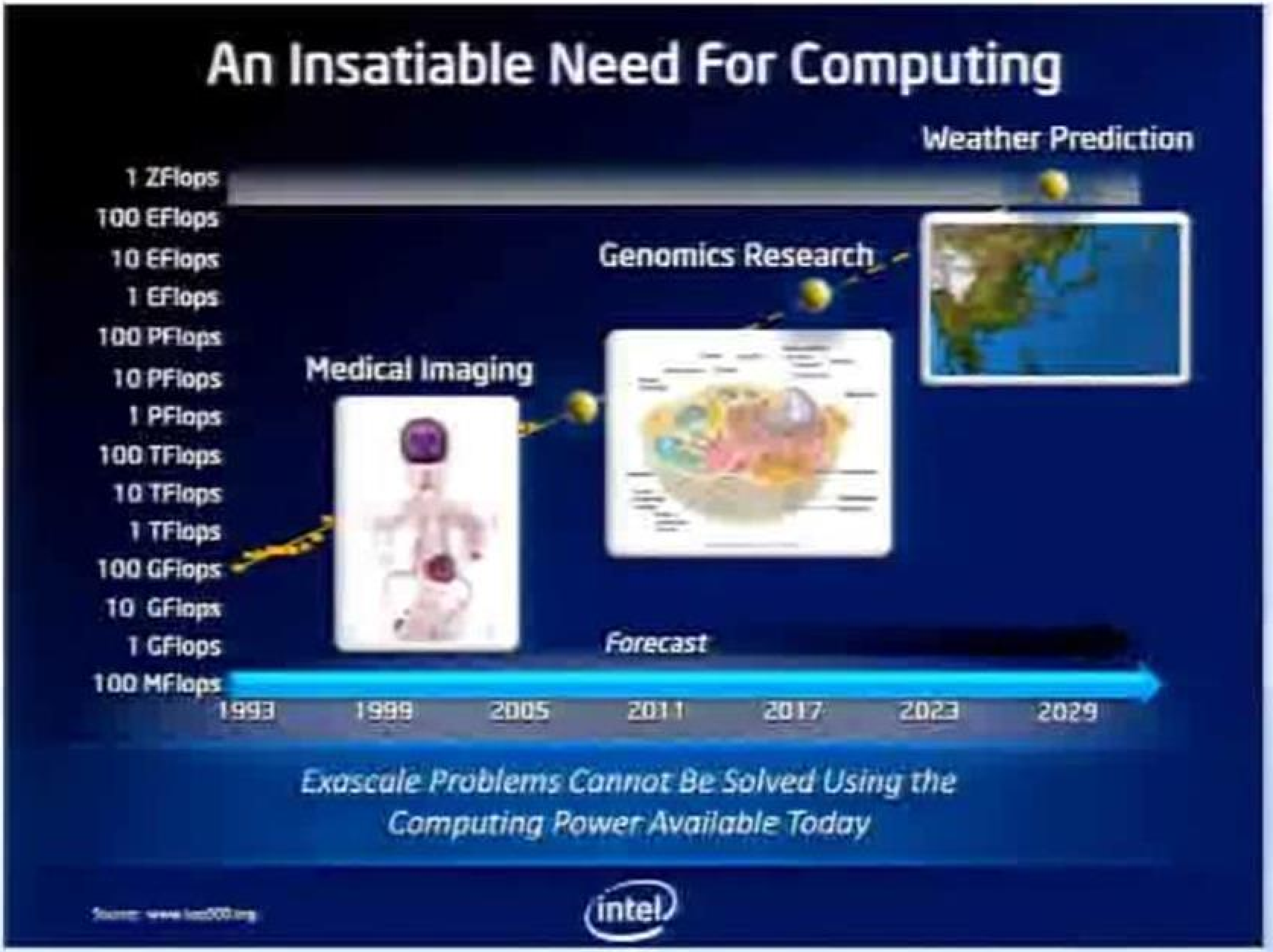
Multiscale Design

Search over the space of alternatives at each scale



Product
Weave Level
Bundle Level
Nano Level
Atomic Level
Properties at the Macroscale
Properties at the Micro- and Nanoscale

Courtesy SCOREC RPI



An Insatiable Need For Computing

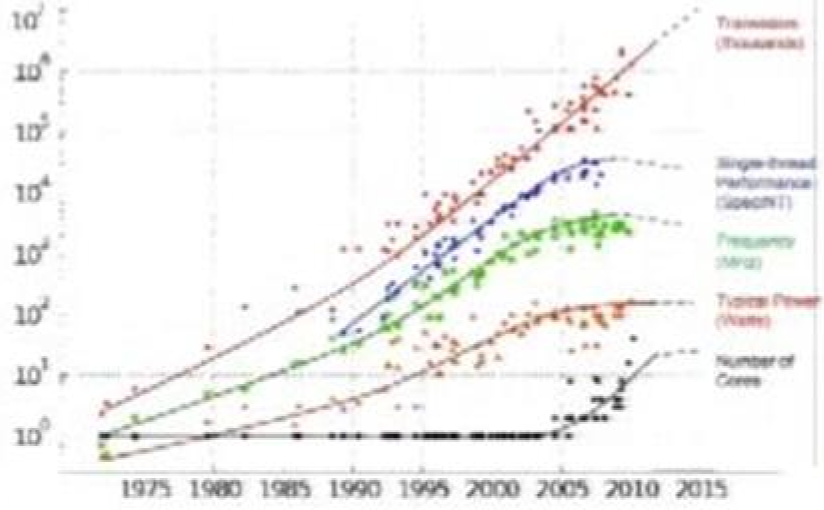
Exascale Problems Cannot Be Solved Using the Computing Power Available Today

Source: www.top500.org

DRIVERS:

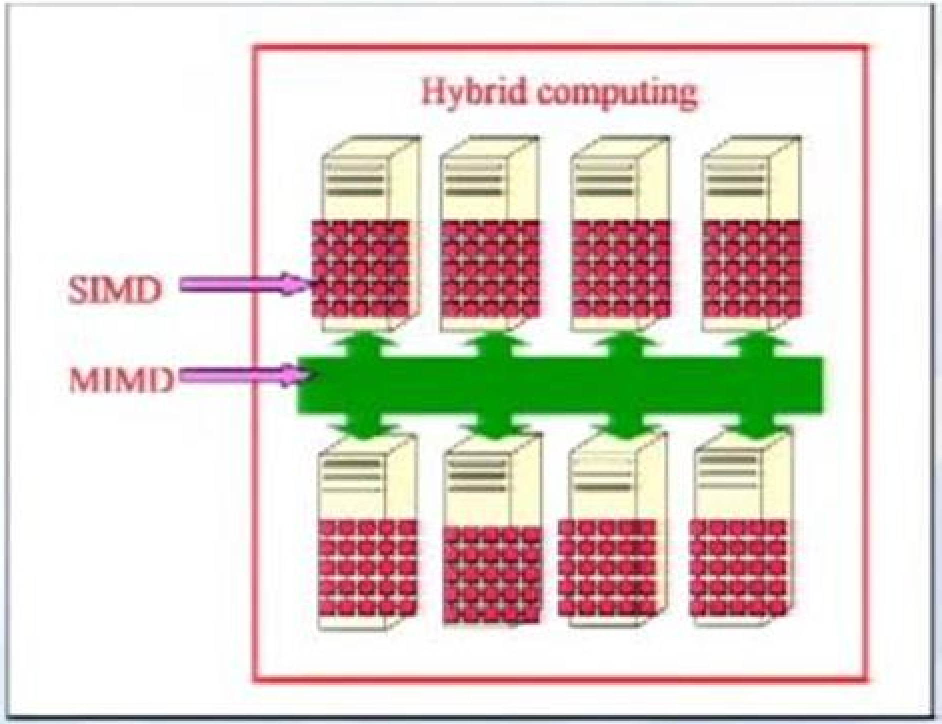
THE PARALLEL COMPUTING CHALLENGE

Result: The End of Historic Scaling



Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

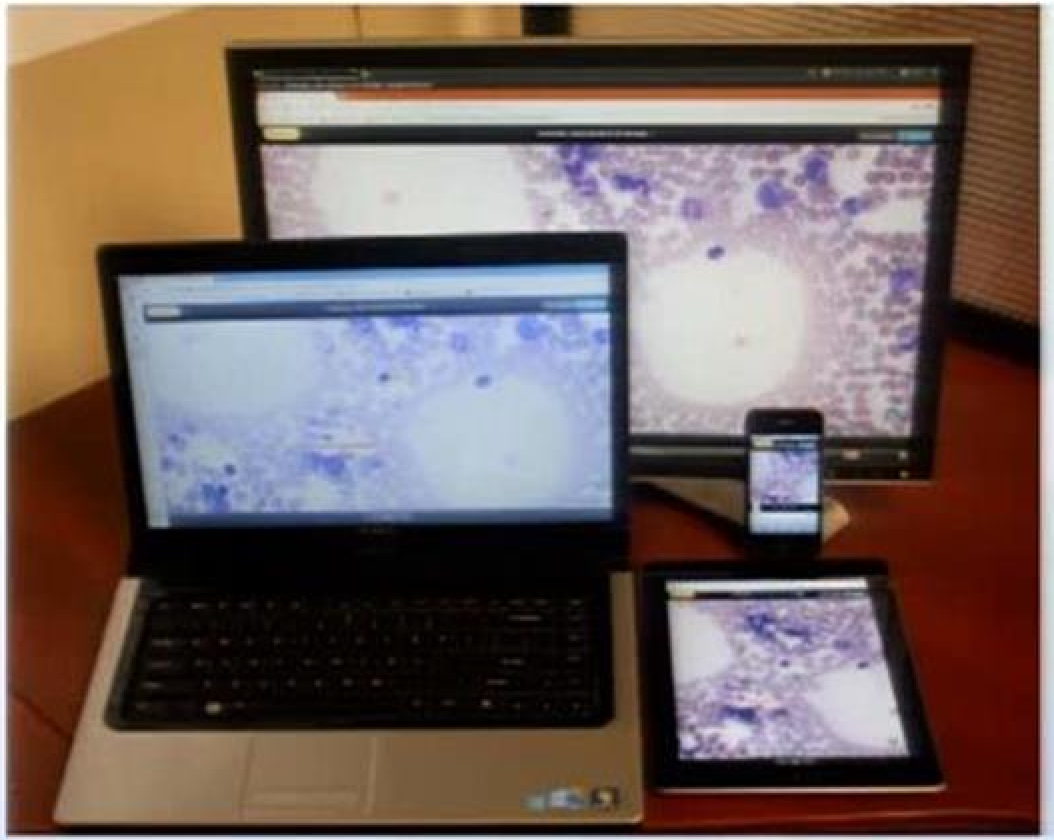
C. Moore, Data Processing in ExaScale-Class Computer Systems, Salishan, April 2011



DRIVERS:

OVERWHELMING ABUNDANCE

- MYRIAD DELIVERY PLATFORMS
- PROLIFERATION OF SOFTWARE SYSTEMS



From mobile to supercomputer

DRIVERS:

REPRODUCIBILITY

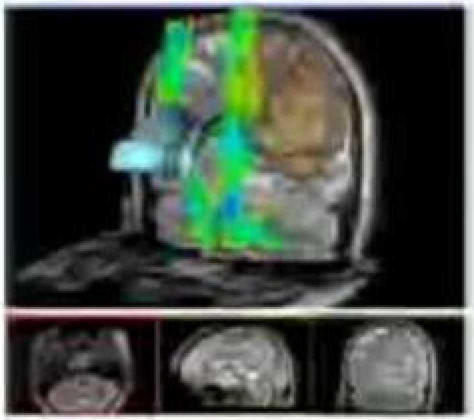
Failure of Reproducibility

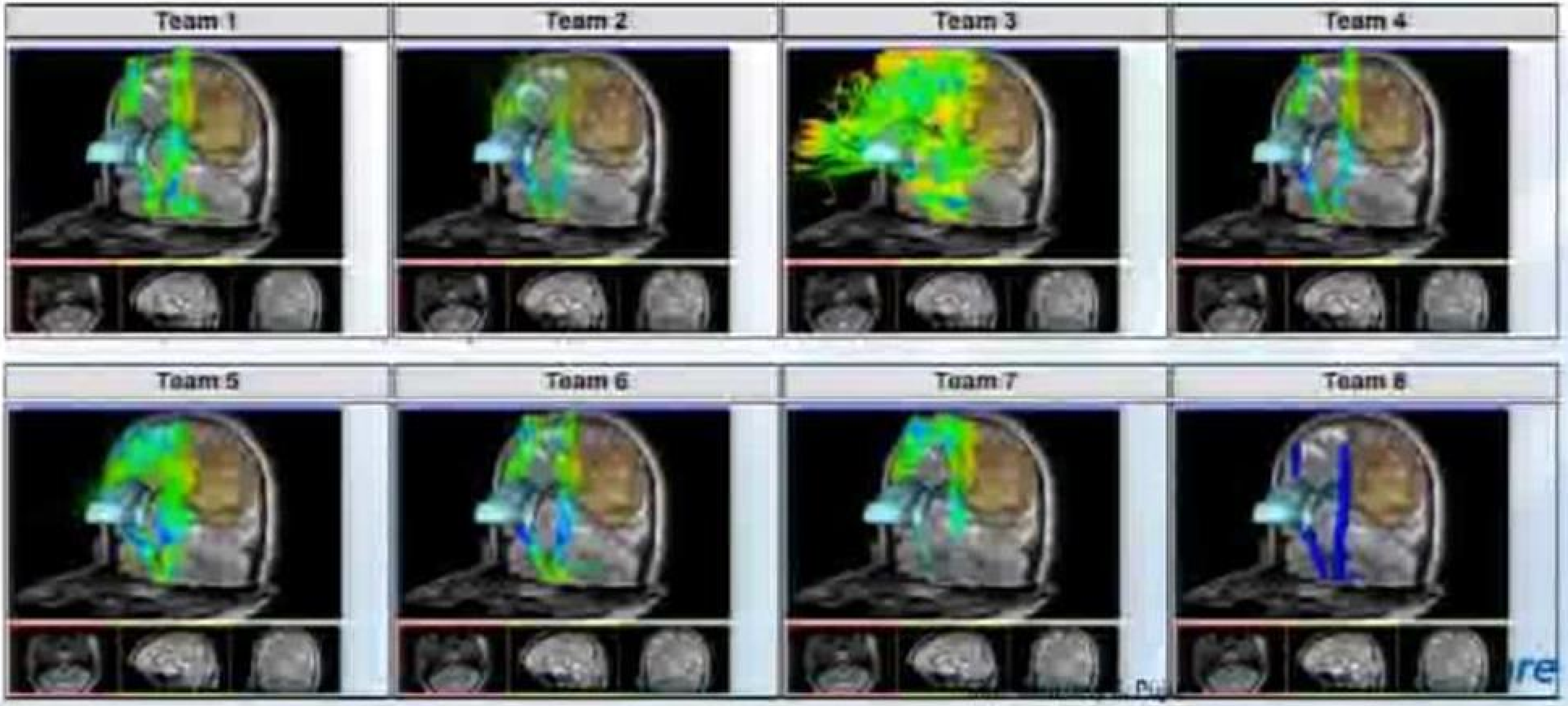
- ***Nature*** *(March 2012)*
  - Glenn Begley, former head of cancer research at pharma giant Amgen
  - Lee M. Ellis, cancer researcher at the University of Texas

Found that more than 90% of papers published in science journals describing "landmark" breakthroughs in preclinical cancer research, are not reproducible, and are thus just plain wrong.

MICCAI Workshop Results

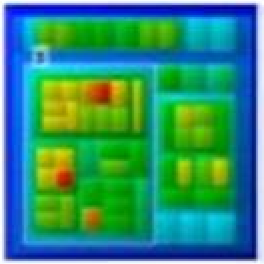
- Large **inter-algorithm** variability in finding the CST (cortico-spinal tract)
- How to compare?





Part 2: Remedies

- Integrative architectures
- Rethinking parallel computing
- Community
- New business models



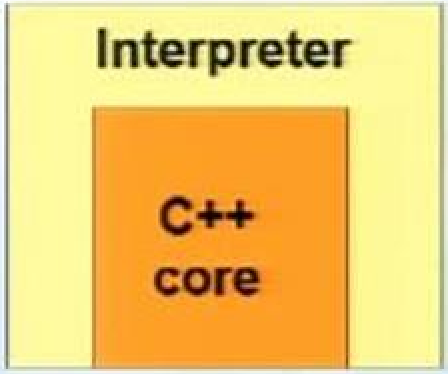
REMEDIES:

INTEGRATIVE ARCHITECTURES

- MODULES NOT MONOLITHS
- WRAPPING

VTK Architecture

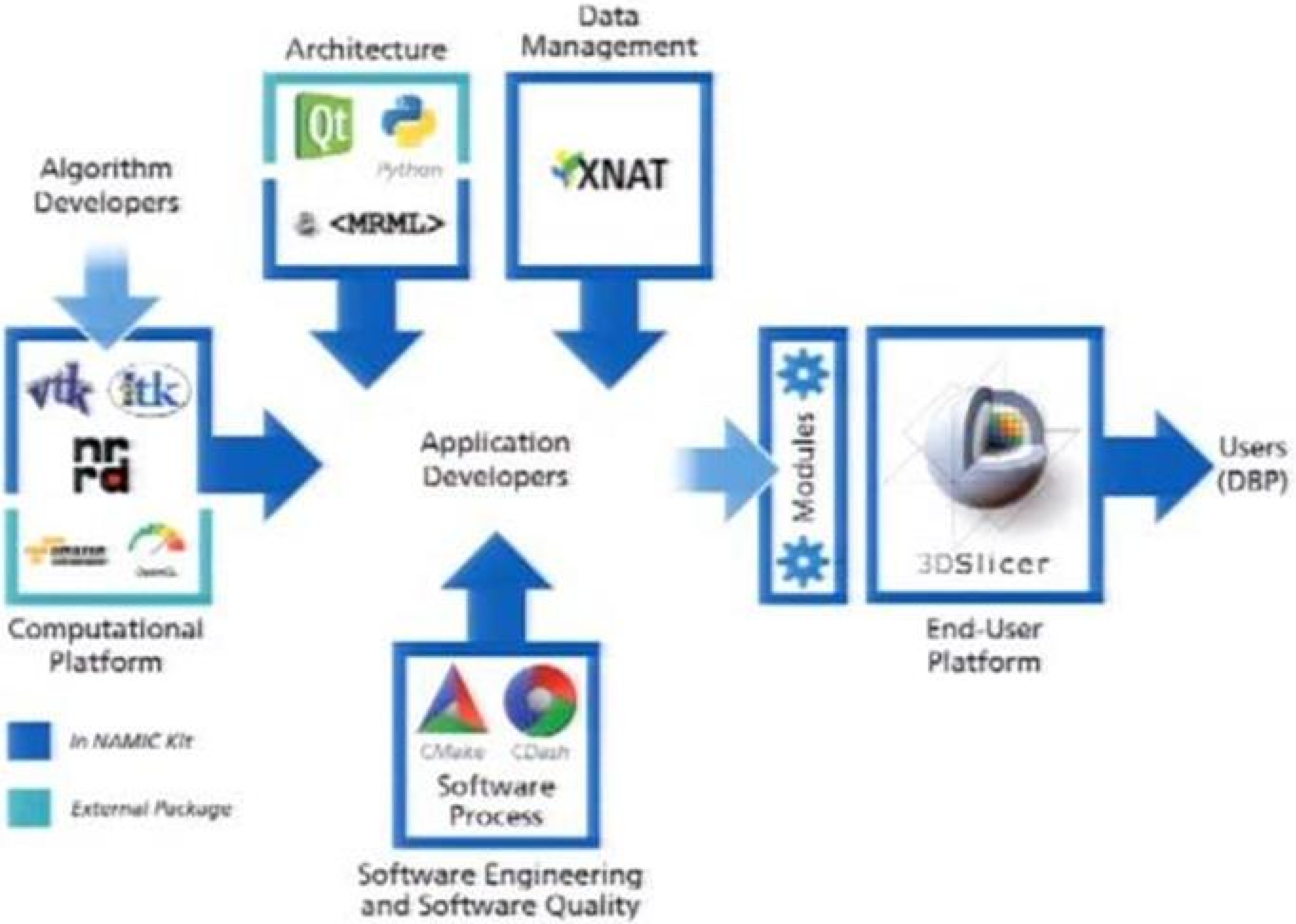
- Hybrid approach
  - Compiled C++ core (faster algorithms)
  - Interpreted applications (rapid development)
  - Interpreted layer generated automatically



National Alliance for Medical Image Computing

A National Center for Biomedical Computing

Funded under the NIH Roadmap Initiative



Architecture

Data Management

Algorithm Developers

Application Developers

Computational Platform

Software Process

Software Engineering and Software Quality

Modules

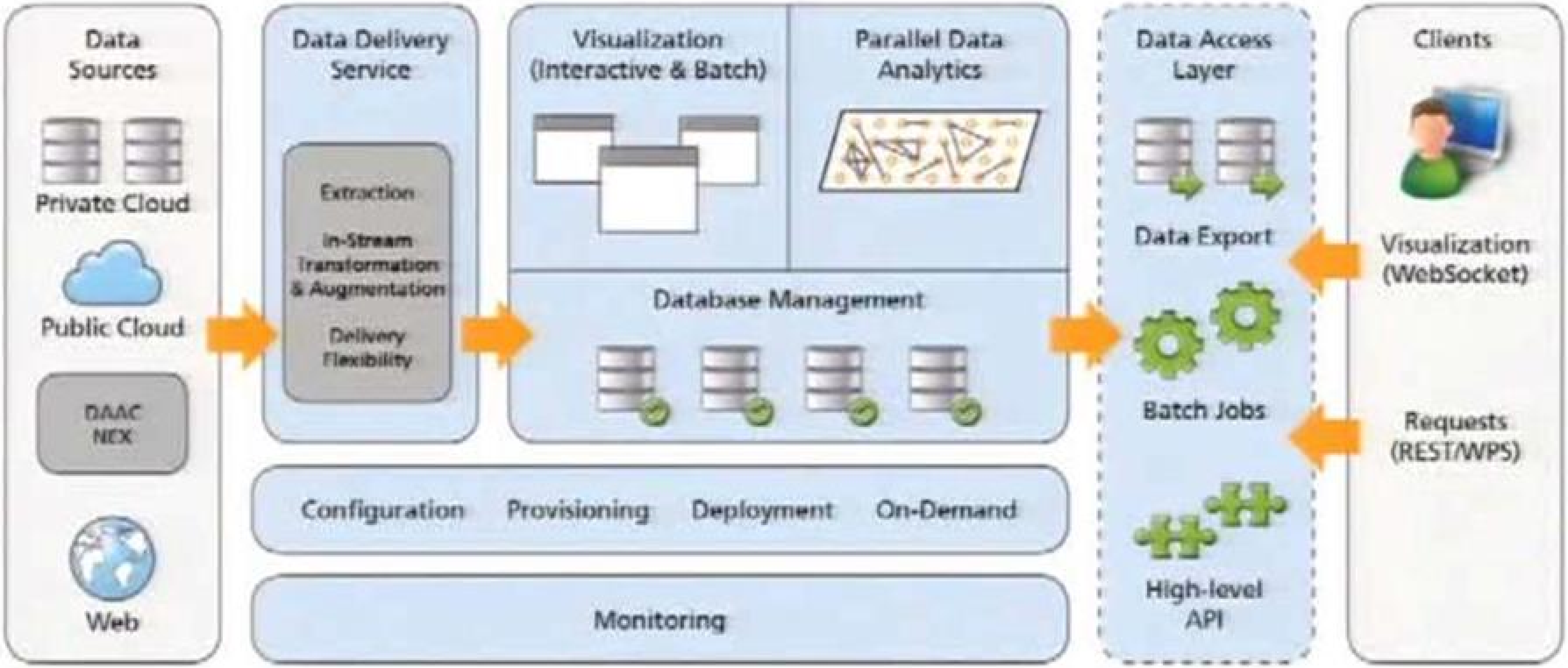
End-User Platform

Users (DBP)

🟦 In NAMIC Kit

🟦 External Package

Kitware



Data Sources

Private Cloud

Public Cloud

DAAC NEX

Web

Data Delivery Service

Extraction

In-Stream Transformation & Augmentation

Delivery Flexibility

Visualization (Interactive & Batch)

Parallel Data Analytics

Database Management

Configuration

Provisioning

Deployment

On-Demand

Monitoring

Data Access Layer

Data Export

Batch Jobs

High-level API

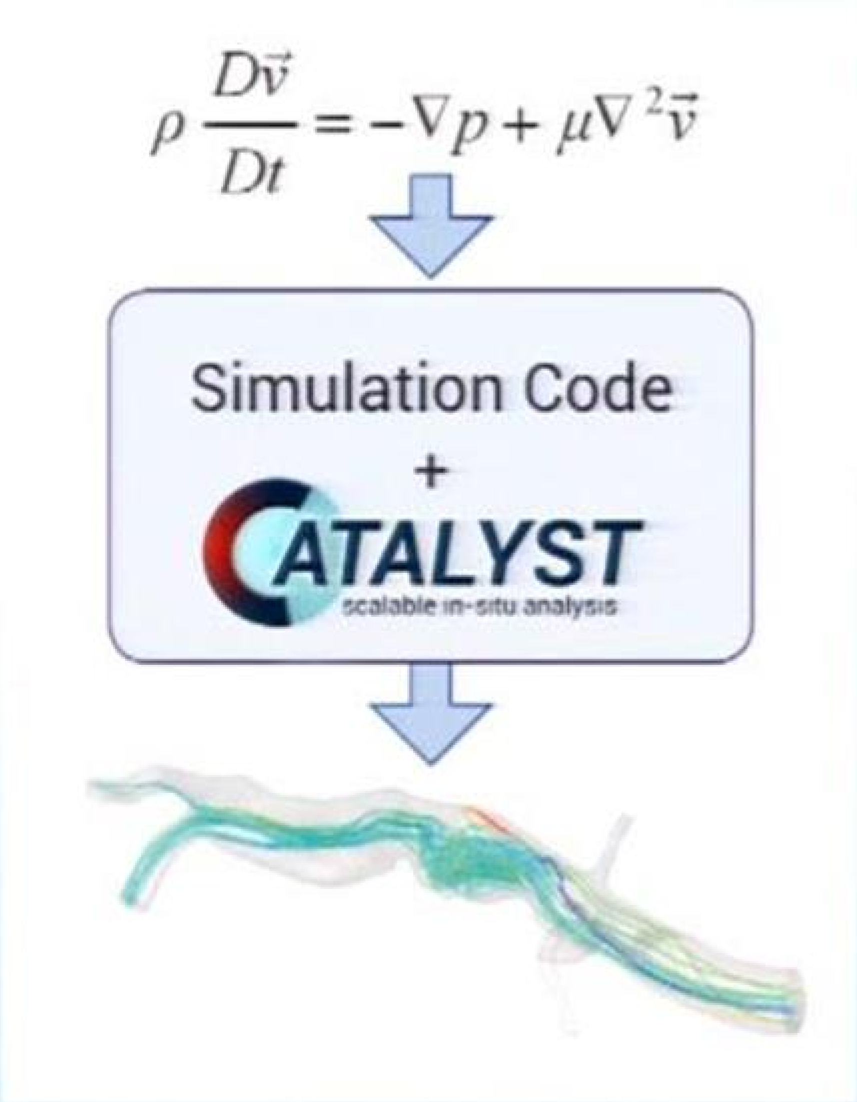
Clients

Visualization (WebSocket)

Requests (REST/WPS)

REMEDIES:

RETHINKING PARALLEL COMPUTING



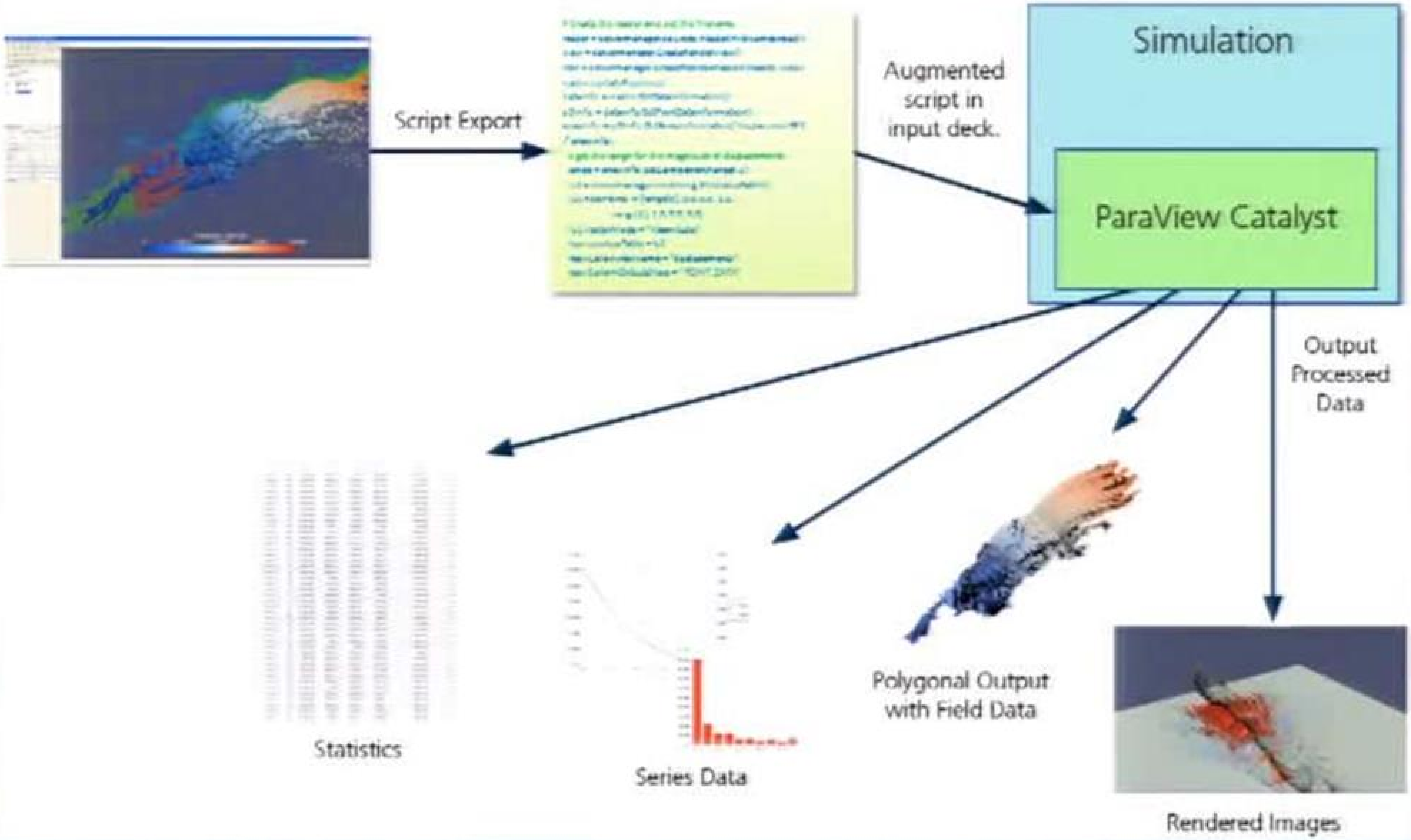
$$\rho \frac{D\vec{v}}{Dt} = -\nabla p + \mu \nabla^2 \vec{v}$$

Simulation Code

+

CATALYST

scalable in-situ analysis



Script Export

Augmented script in input deck.

Simulation

ParaView Catalyst

Output Processed Data

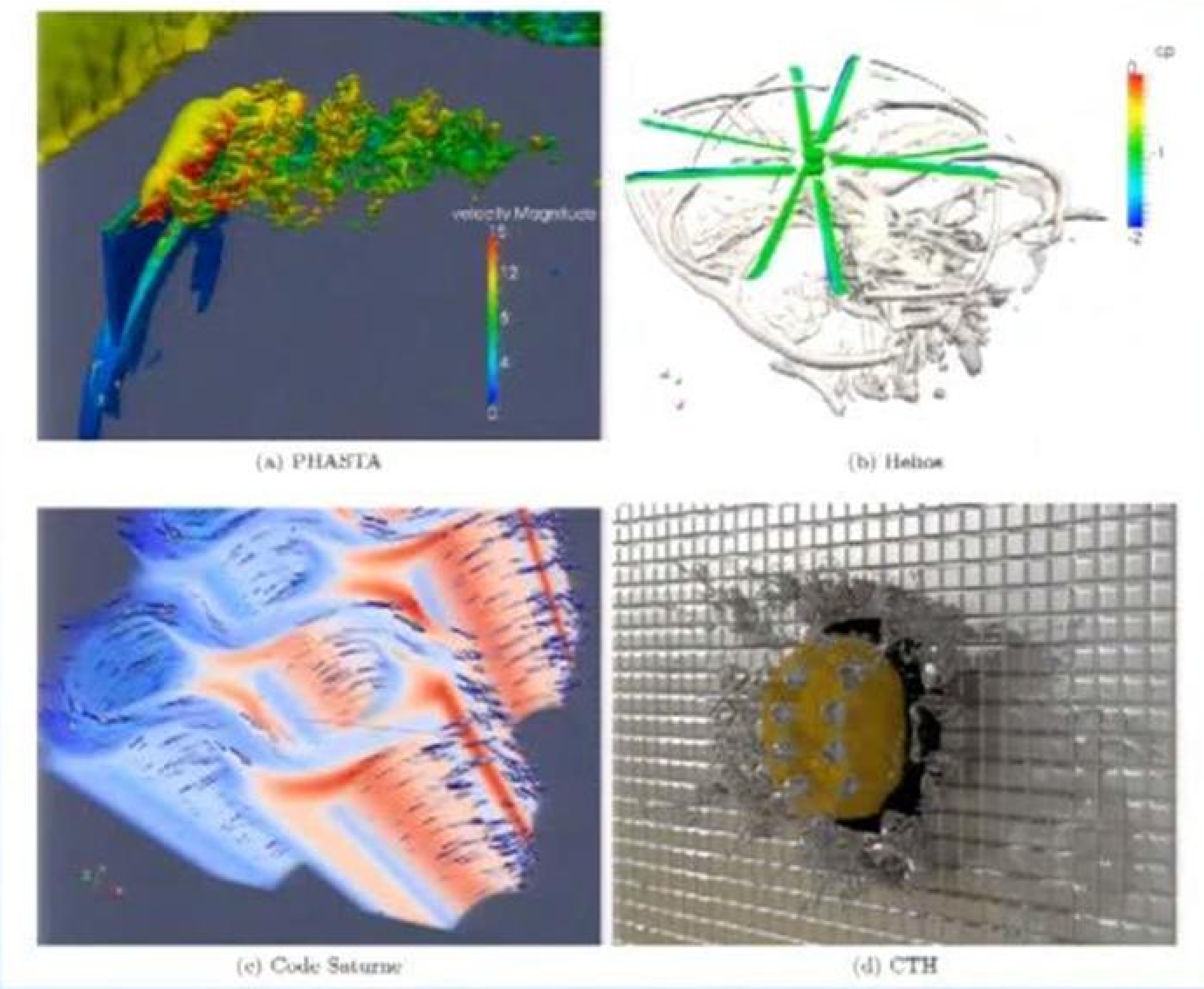
Statistics

Series Data

Polygonal Output with Field Data

Rendered Images

ParaView Catalyst Examples



(a) PHASTA

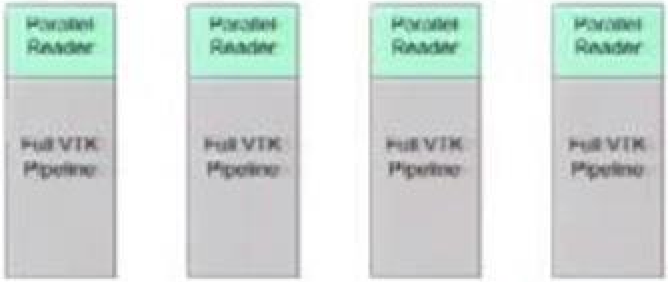
(b) Helios

(c) Code Saturne

(d) CTH

VTK: Parallel Support

- Data Parallel

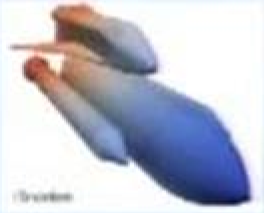


- vtkSMPTools
  - Abstract wrapper around Intel TBB, X-Kaapi, Simple, Sequential
- VTK-m

VTK-m Project Goals

- A single place for the visualization community to collaborate, contribute, and leverage massively threaded algorithms.

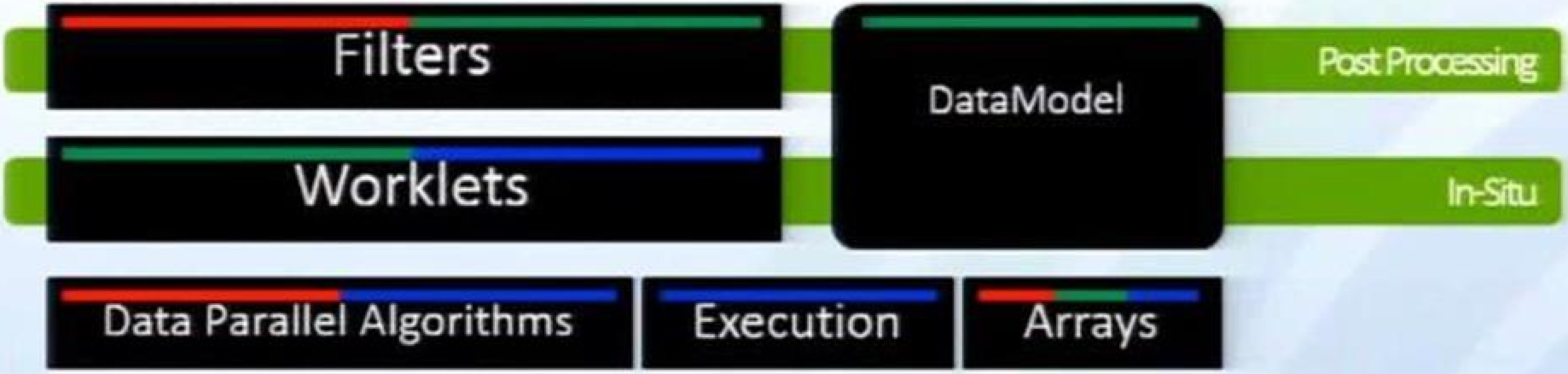
- Make it easier for simulation codes to take advantage these parallel visualization and analysis tasks on a wide range of current and next-generation hardware.



Kitware

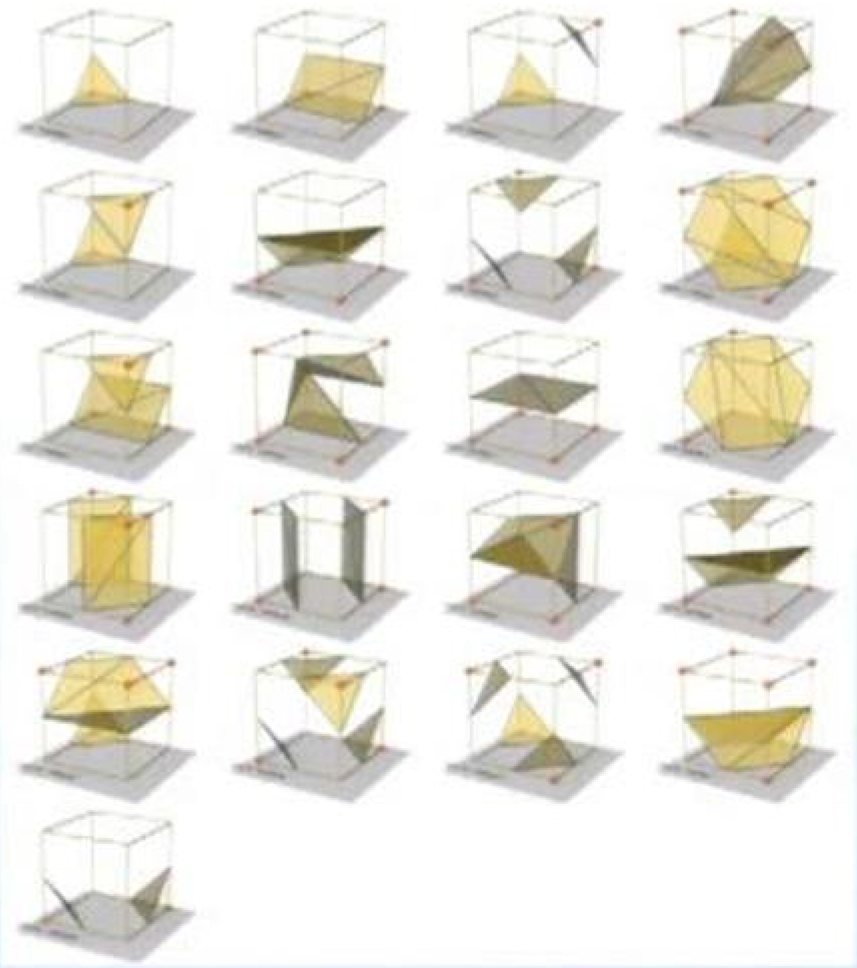
VTK-m Architecture

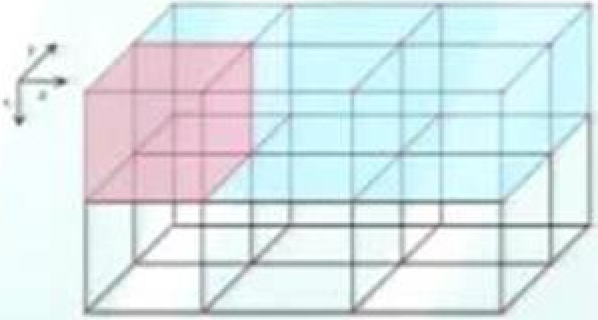
- Combines strengths of multiple projects:
  - EAVL, Oak Ridge National Laboratory
  - DAX, Sandia National Laboratory
  - PISTON, Los Alamos National Laboratory

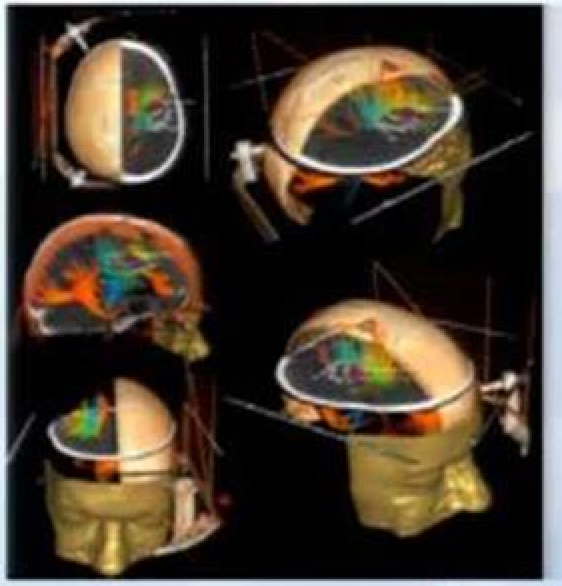


Rethinking Algorithms

- Marching Cubes

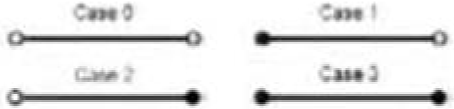




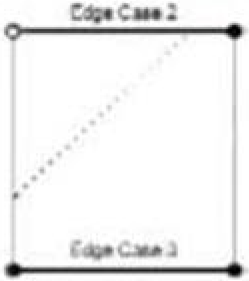


Kitware

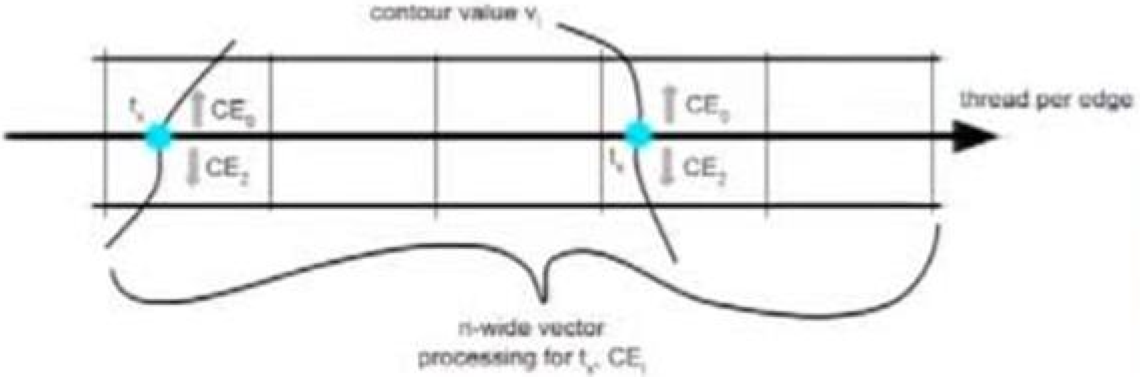
Flying Edges

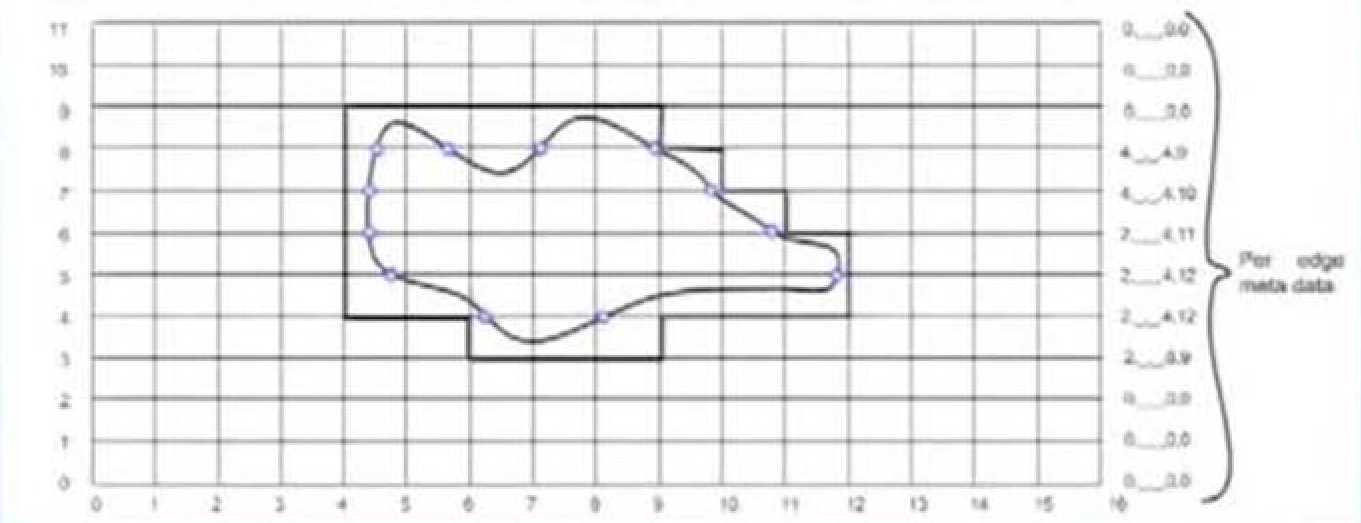


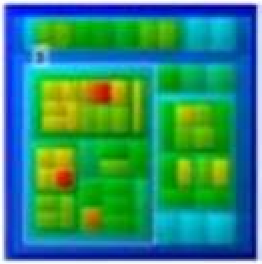
Edge-based case table. Each x-cell edge is assigned one of our different values.



Edges are combined to produce a cell case value.







REMEDIES:

COMMUNITY

- OPEN SYSTEMS
- BUILDING COMMUNITIES
- SOFTWARE PROCESS

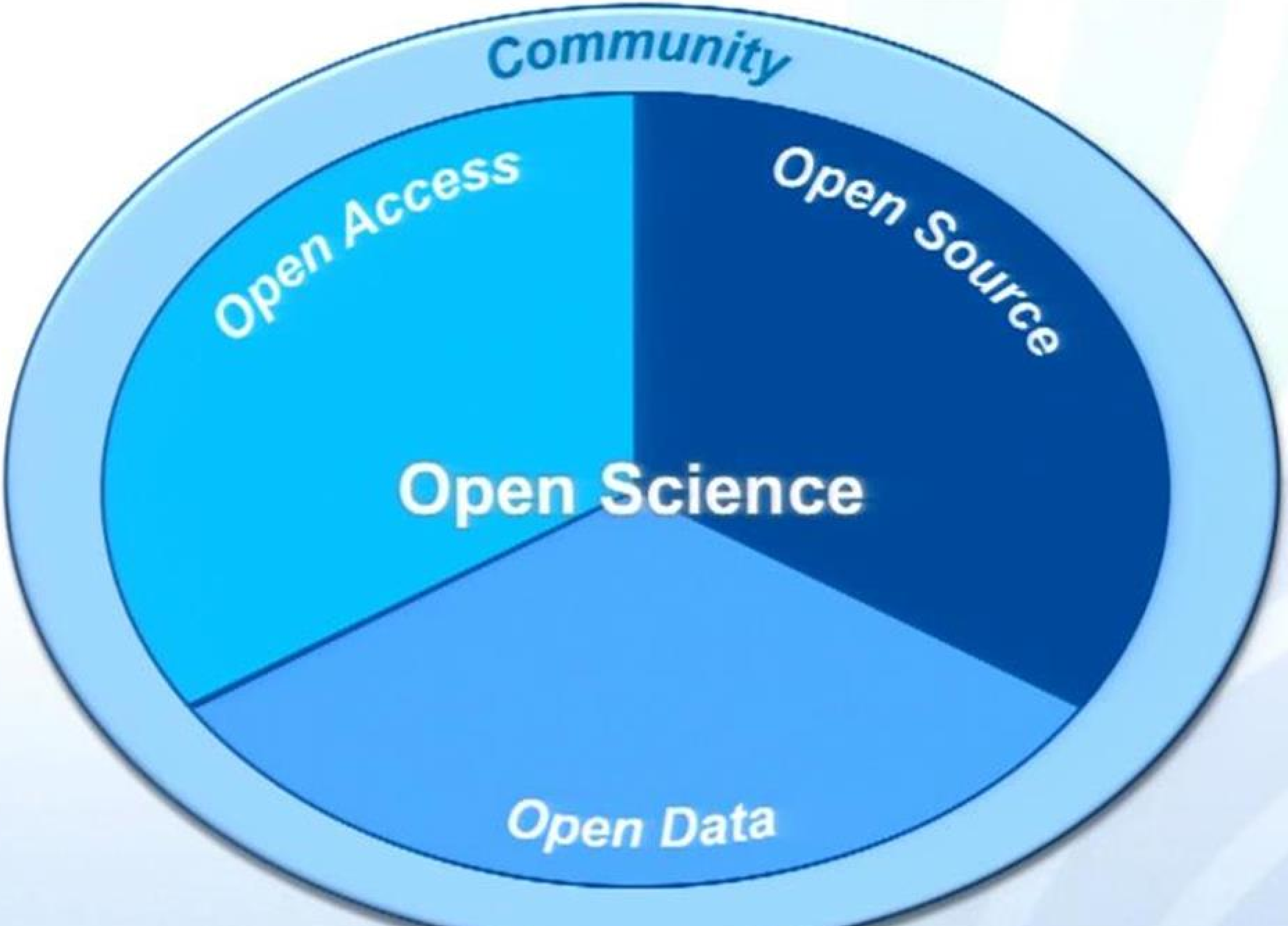
Reproducible Science

Open Access, Open Data, Open Source

Nullius in Verba

"take nobody's word for it"
Royal Society 1660

Kitware



Community

Open Access

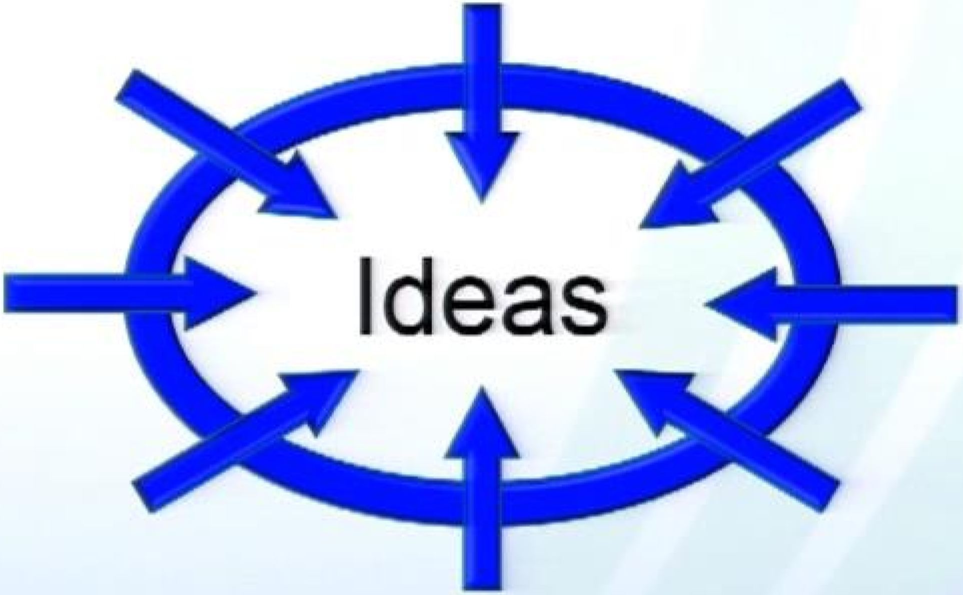
Open Source

Open Science

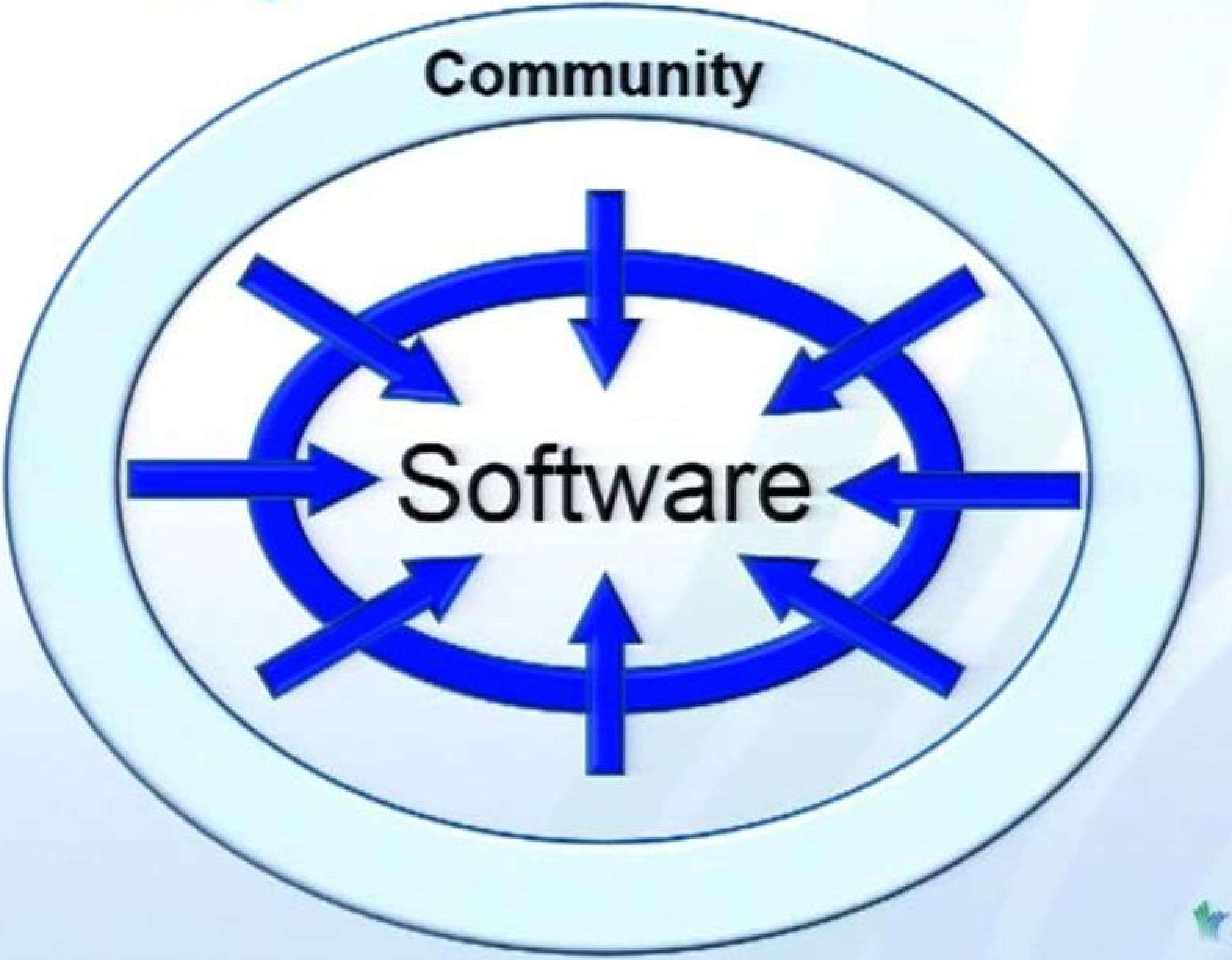
Open Data

Naïve View of Open

Innovation View of Open



Software is a Manifestation of Community



Kitware

Bigger Science Requires Bigger Communities

- Science
  - Larger teams, more disciplines, more countries

Daniel S. Katz, *NSF Program Director*
Division of Advanced Cyberinfrastructure

Kitware

Many of the issues in developing sustainable software are social, not technical

Daniel S. Katz, *NSF Program Director*
Division of Advanced Cyberinfrastructure

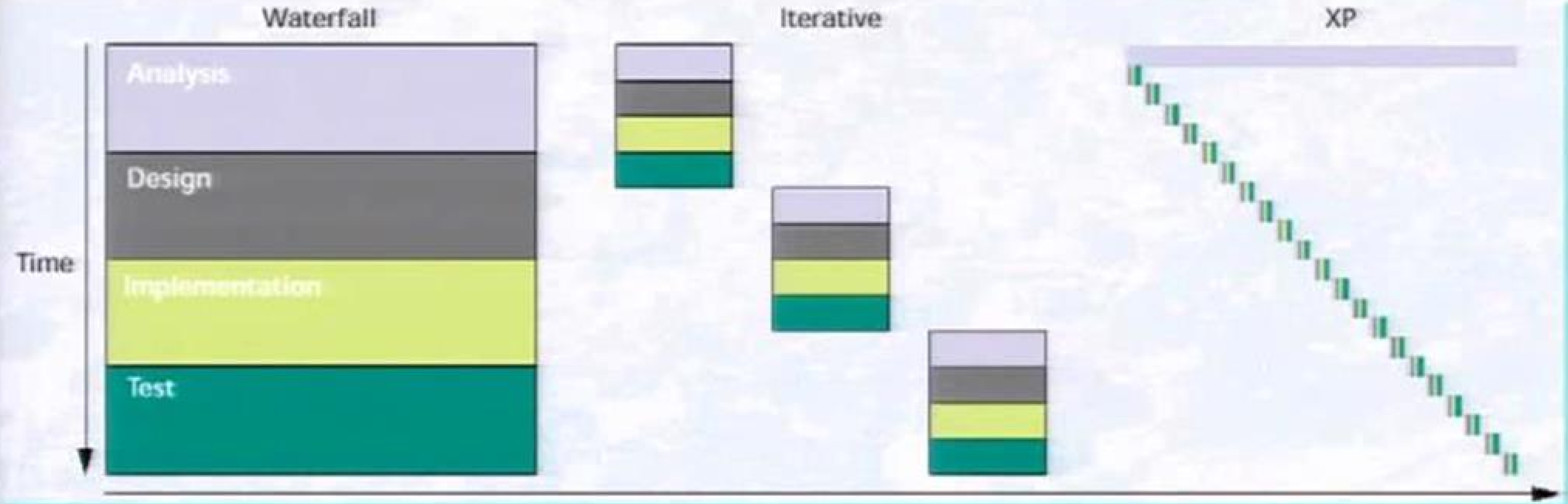
Kitware

Open Source as a Trusted Foundation

- High-quality building blocks
  - Can be deployed rapidly
- Transparent
  - See what's going on, debug easier
  - Interface to open systems more readily
- Avoid vendor lock-in
  - Enable competitive bidding

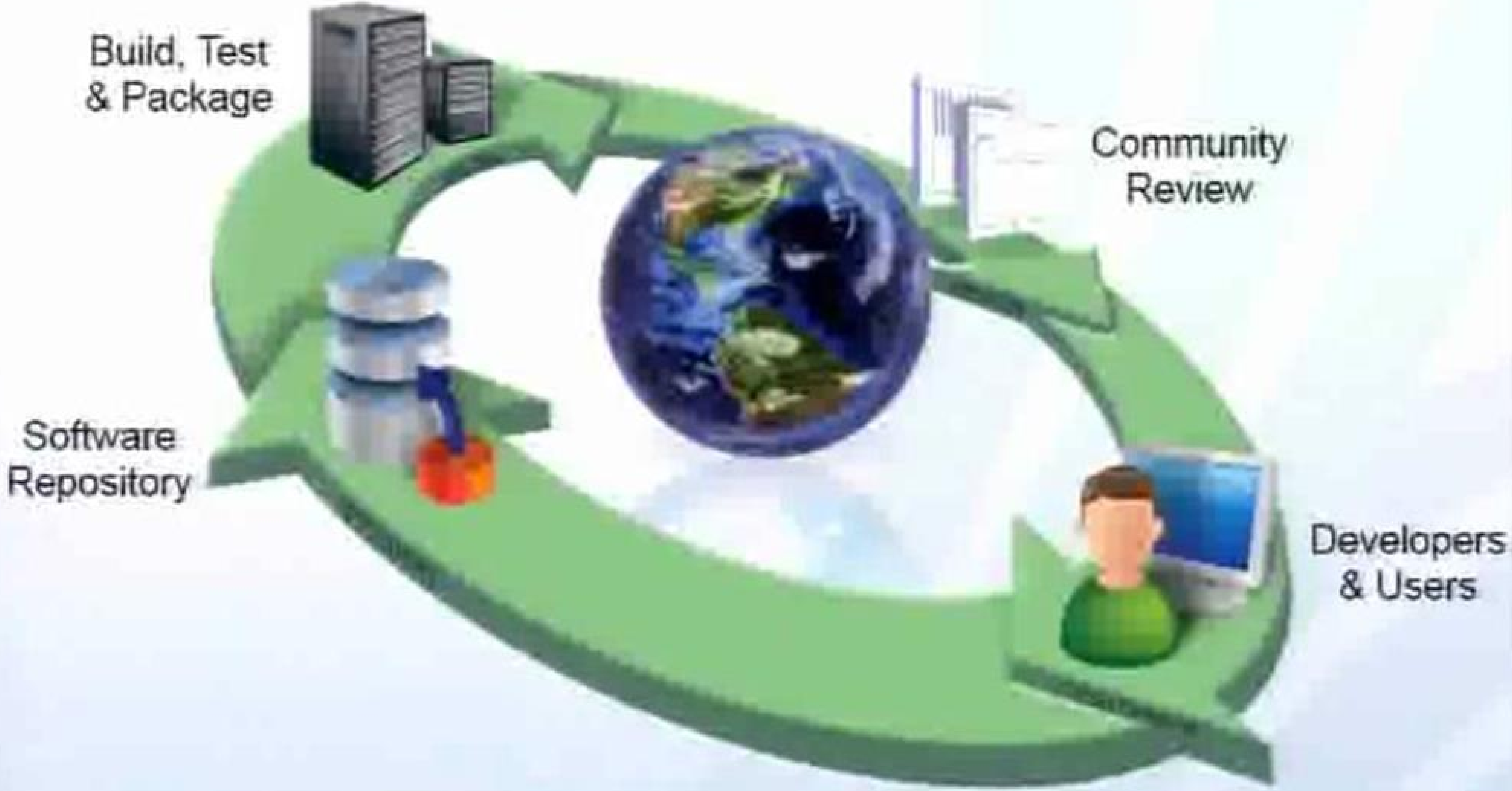
Open Software Process

- Who does waterfall anymore?
- Now: Agile, test-driven development, etc.

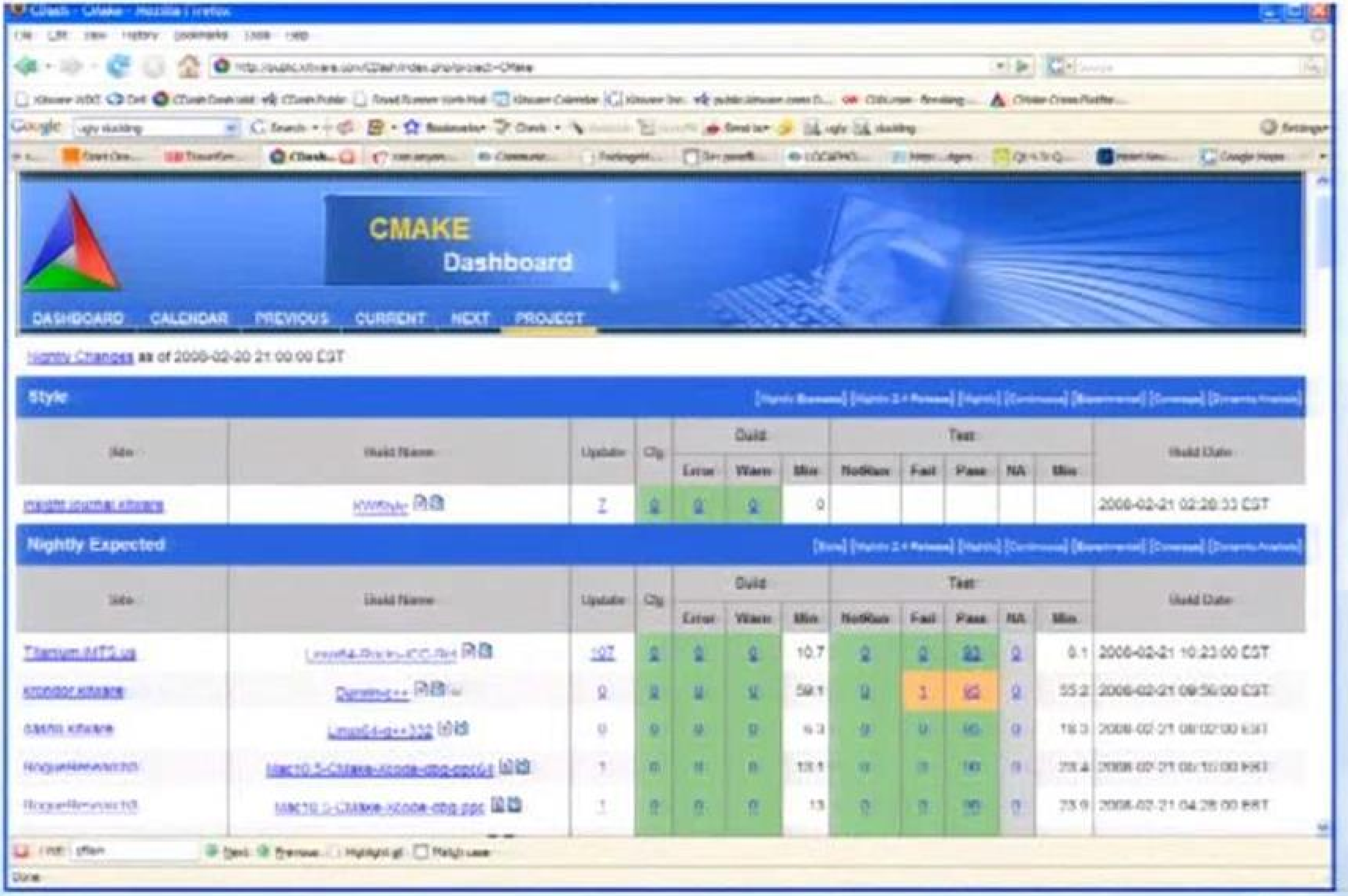


Kitware

Software Quality Process



CDash Dashboard www.cdash.org



Nightly Changes as of 2008-02-20 21:00:00 EST

Style

Site	Build Name	Update	Cfg	Build			Test					Build Date
				Error	Warn	Min	NotRun	Fail	Pass	NA	Min	
[illegible]	KWStyle	7	0	0	0	0						2008-02-21 02:28:33 EST

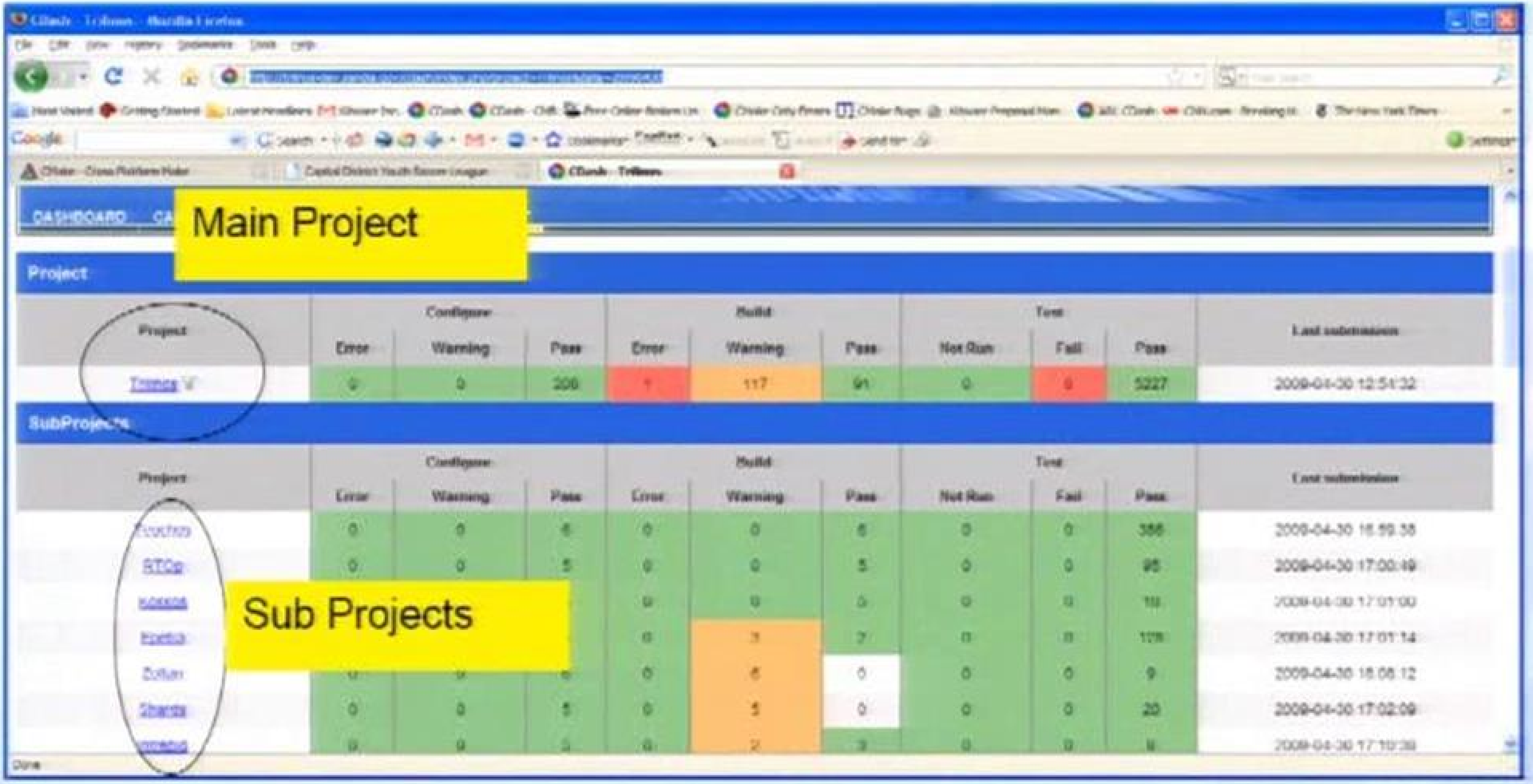
Nightly Expected

Site	Build Name	Update	Cfg	Build			Test					Build Date
				Error	Warn	Min	NotRun	Fail	Pass	NA	Min	
Titanium.MTS.ua	[illegible]	107	0	0	0	10.7	0	0	93	0	0.1	2008-02-21 10:23:00 EST
[illegible].kitware	Dynamic++	0	0	0	0	58.1	0	1	92	0	55.2	2008-02-21 09:56:00 EST
[illegible].kitware	Linux64-g++332	0	0	0	0	6.3	0	0	93	0	18.3	2008-02-21 08:02:00 EST
[illegible]	Mac10.5-CMake-Xcode-dbg-ppc64	1	0	0	0	13.1	0	0	93	0	29.4	2008-02-21 [illegible]
[illegible]	Mac10.5-CMake-Xcode-dbg-ppc	1	0	0	0	13	0	0	93	0	23.9	2008-02-21 04:28:00 EST

Done

Kitware

Trilinos (Multi-Package Dashboard) http://trilinos-dev.sandia.gov/cdash/index.php



Main Project

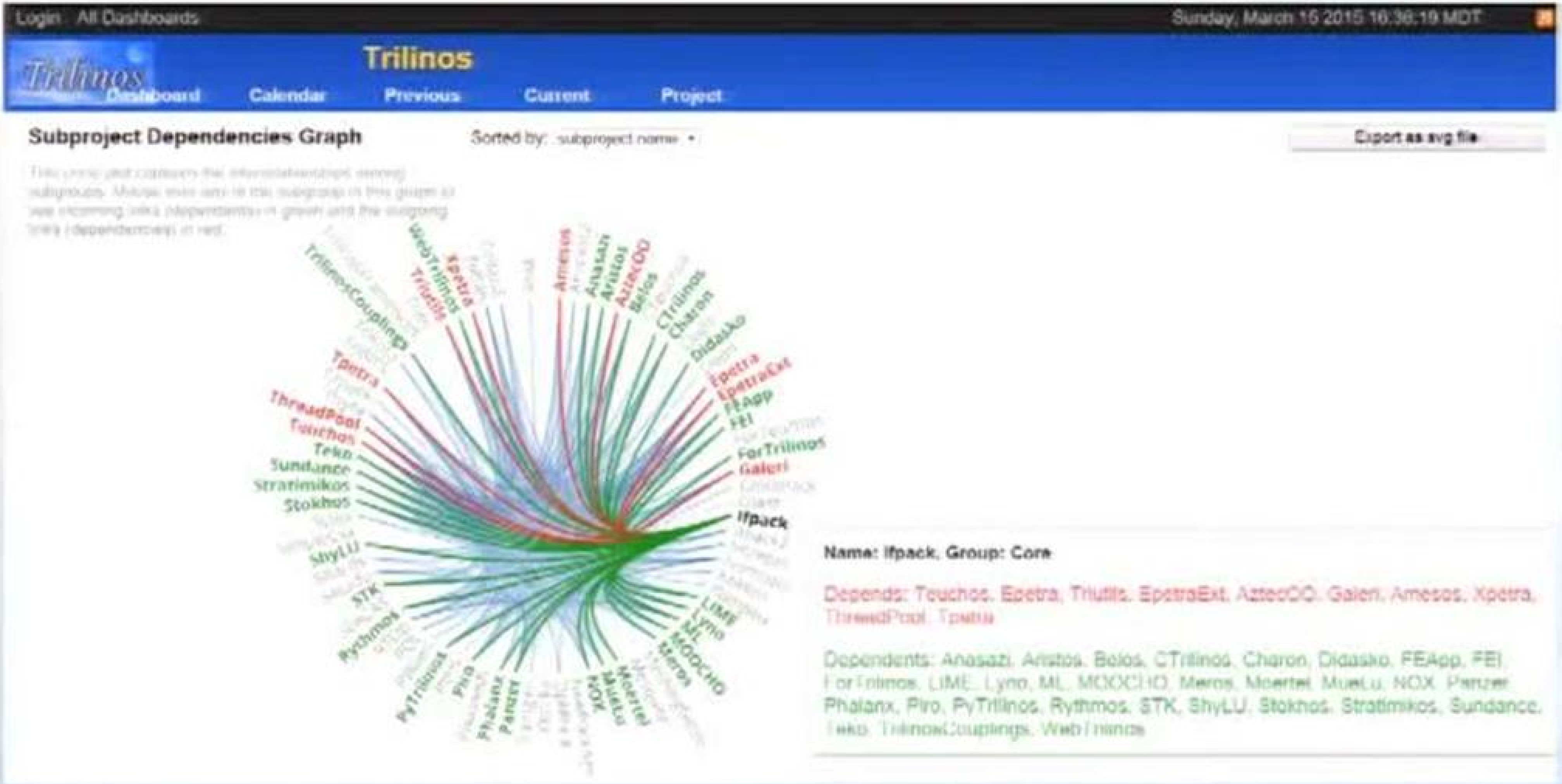
Project

Project	Configure			Build			Test			Last submission
	Error	Warning	Pass	Error	Warning	Pass	Not Run	Fail	Pass	
Trilinos	0	0	206	1	117	91	0	9	5227	2009-04-30 12:54:32

SubProjects

Project	Configure			Build			Test			Last submission
	Error	Warning	Pass	Error	Warning	Pass	Not Run	Fail	Pass	
Teuchos	0	0	6	0	0	6	0	0	386	2009-04-30 16:59:38
RTOp	0	0	5	0	0	5	0	0	95	2009-04-30 17:00:49
Kokkos	[illegible]	[illegible]	[illegible]	0	0	5	0	0	19	2009-04-30 17:01:00
Epetra	[illegible]	[illegible]	[illegible]	0	3	7	0	0	128	2009-04-30 17:01:14
Zoltan	0	0	6	0	6	0	0	0	9	2009-04-30 18:08:12
Shards	0	0	5	0	5	0	0	0	20	2009-04-30 17:02:09
Amesos	0	0	5	0	2	3	0	0	8	2009-04-30 17:10:39

Sub Projects



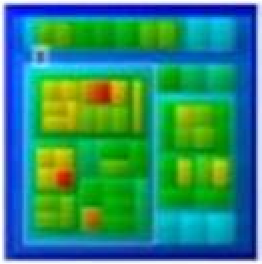
Subproject Dependencies Graph

This circle plot captures the interrelationships among subgroups. Mouse over any of the subgroup in this graph to see incoming links (dependents) in green and the outgoing links (dependencies) in red.

Name: Ifpack, Group: Core

Depends: Teuchos, Epetra, Triutils, EpetraExt, AztecOO, Galeri, Amesos, Xpetra, ThreadPool, Tpetra

Dependents: Anasazi, Aristos, Belos, CTrilinos, Charon, Didasko, FEApp, FEI, ForTrilinos, LIME, Lyno, ML, MOOCHO, Meros, Moertel, MueLu, NOX, Panzer, Phalanx, Piro, PyTrilinos, Rythmos, STK, ShyLU, Stokhos, Stratimikos, Sundance, Teko, TrilinosCouplings, WebTrilinos



REMEDIES:

NEW BUSINESS MODELS

- PLATFORM STRATEGY
- DITCH THE IP
- FUNDING STRATEGIES

Platform Strategy

ParaView

Parallel Visualization Application

ParaView Desktop

ParaView Python

ParaView HPC

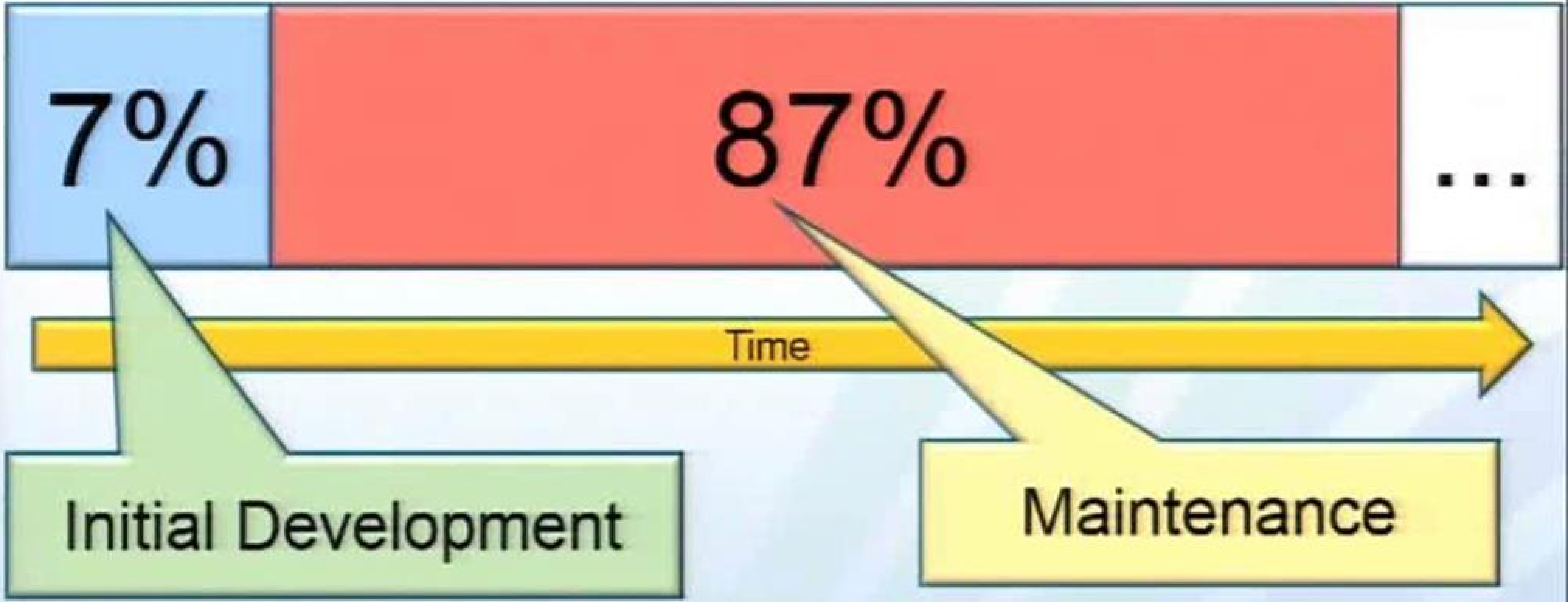
ParaViewWeb

ParaView In Situ

ParaView Immersive

VeloView

The Cost of Software



7%

87%

...

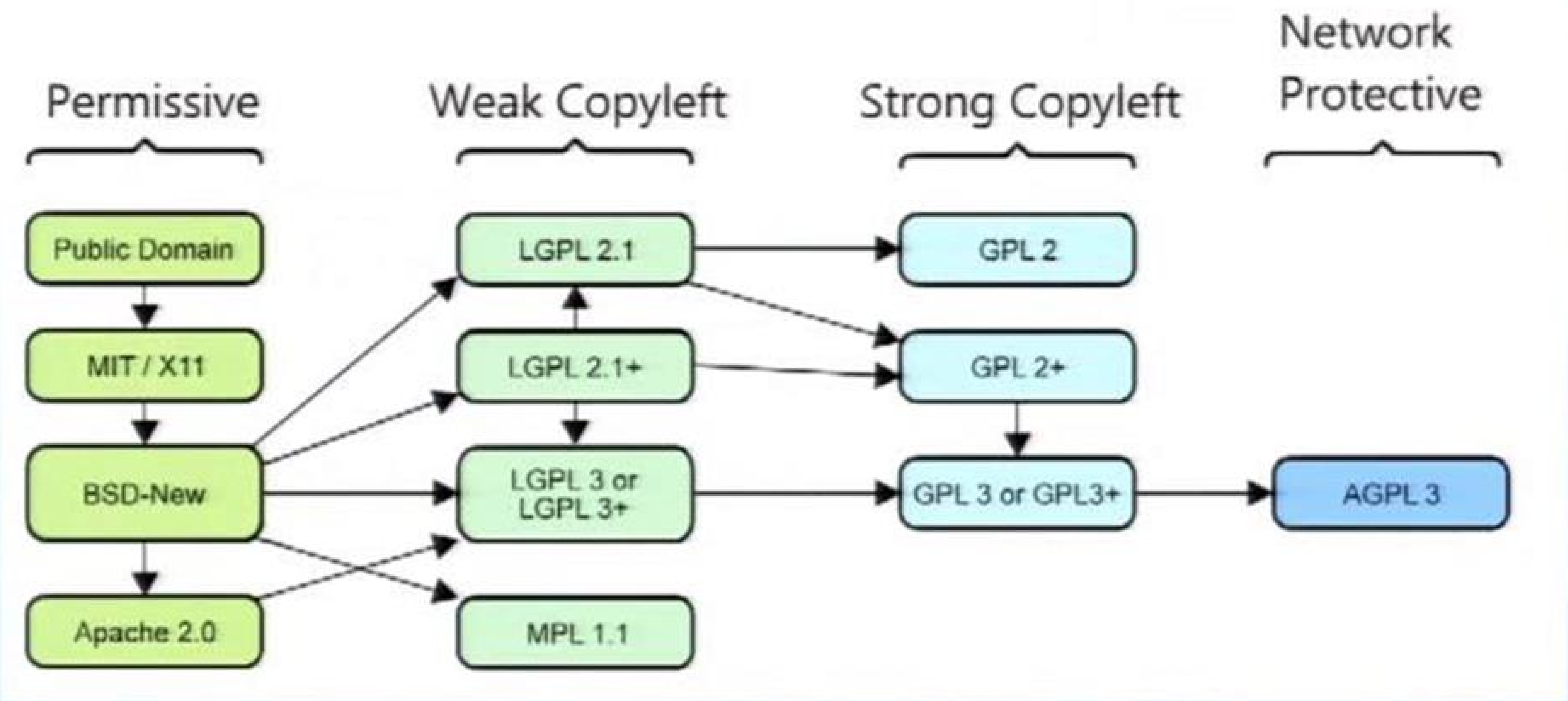
Time

Initial Development

Maintenance

Use the Right OS License

Permissive licenses harvest more contributions



Permissive

Public Domain

MIT / X11

BSD-New

Apache 2.0

Weak Copyleft

LGPL 2.1

LGPL 2.1+

LGPL 3 or LGPL 3+

MPL 1.1

Strong Copyleft

GPL 2

GPL 2+

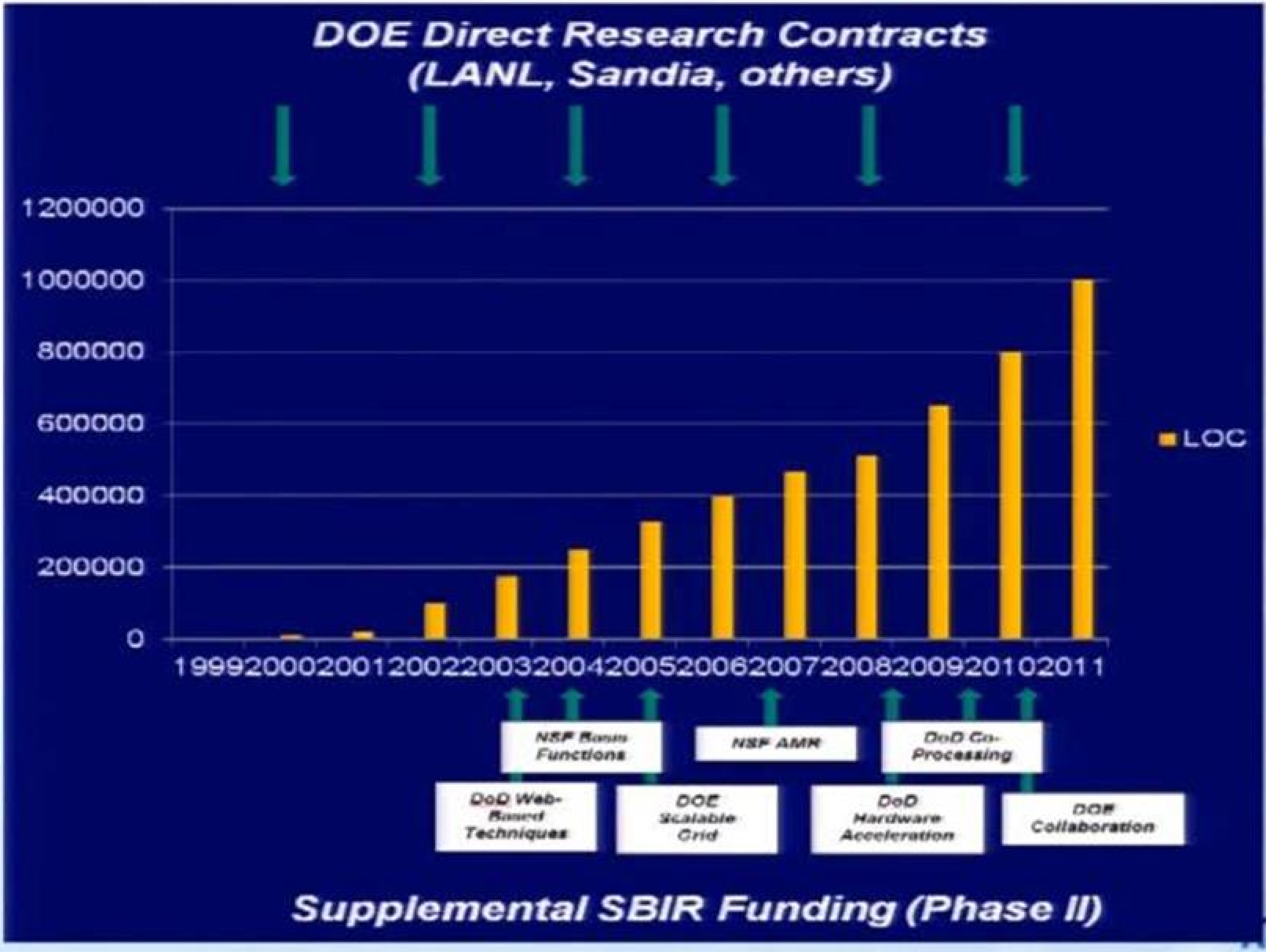
GPL 3 or GPL3+

Network Protective

AGPL 3

Kitware

ParaView / VTK Funding Stream



DOE Direct Research Contracts (LANL, Sandia, others)

NSF Basis Functions

NSF AMR

DoD Co-Processing

DoD Web-Based Techniques

DOE Scalable Grid

DoD Hardware Acceleration

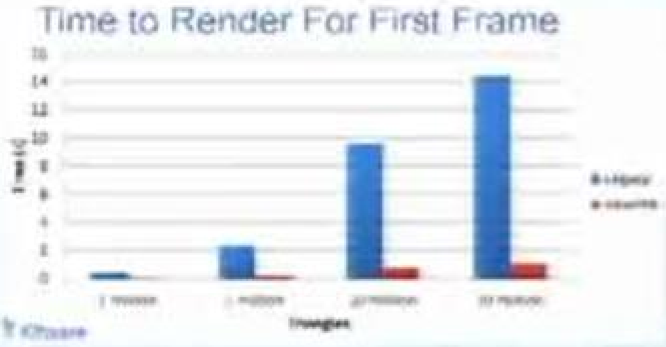
DOE Collaboration

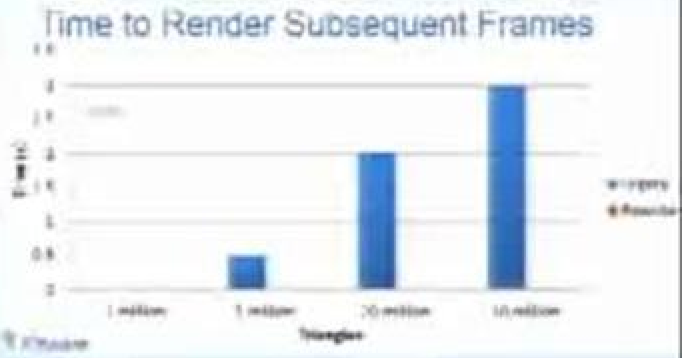
Supplemental SBIR Funding (Phase II)

Funding Productive Work

- Platforms

- Infrastructure





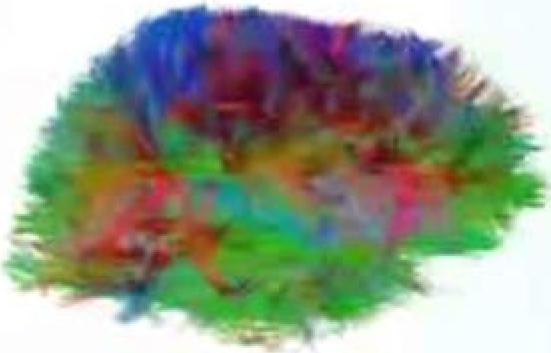
- Community

Kitware

Learning to Love Software Complexity

- Integrative architectures
  - Embrace contributions from everywhere
- Rethinking parallel computing
  - Build simple, core operations
  - A golden algorithmic age
- Community
  - Open your systems and solve tough problems
  - Build communities and make new friends
  - Software process to ensure a stable foundation
- New business models
  - Shield users from churn via the platform strategy
  - Ditch the IP: stop building toll gates
  - Funding strategies: build for the long term

Kitware



Questions?

will.schroeder@kitware.com