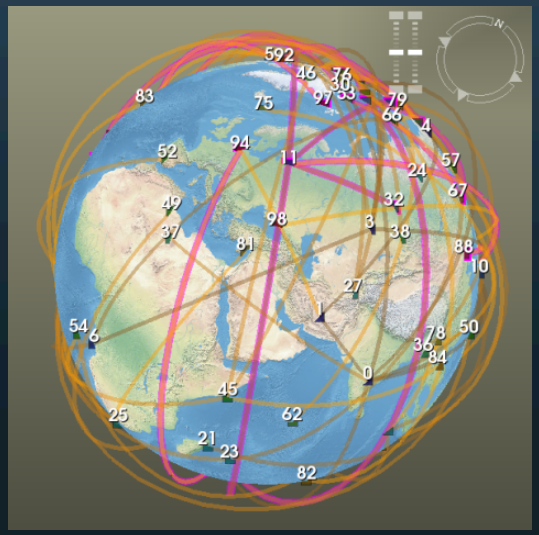
Kitware

Scaling for the Future

How to Practice Open Science

Will Schroeder
co-Founder, President

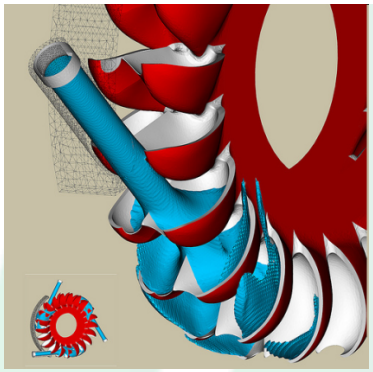
Scientific Software Days 2012

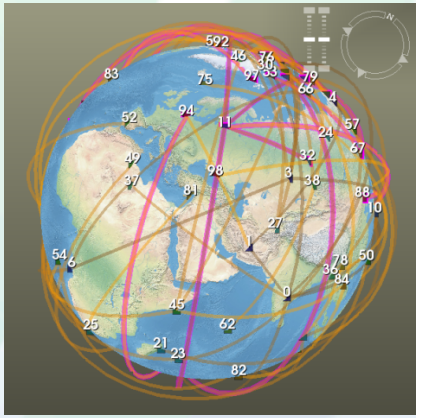


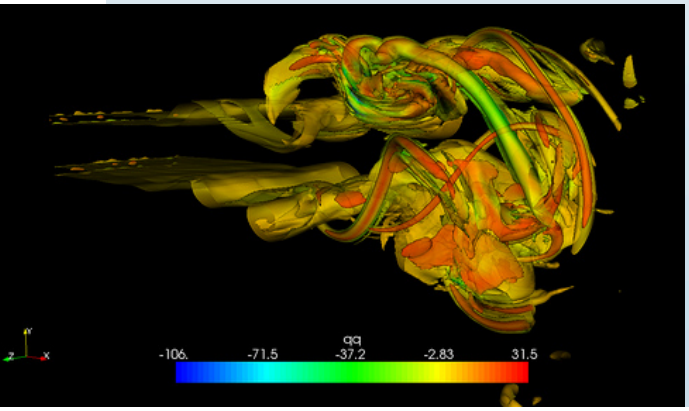
Kitware, Inc.

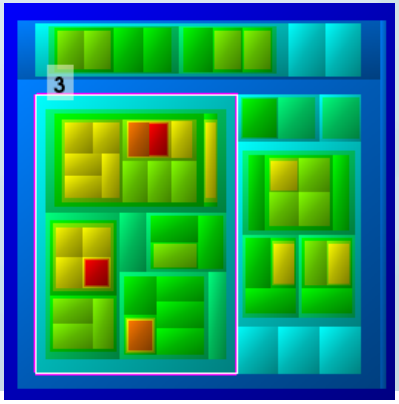
Open Source
Scientific Computing Software

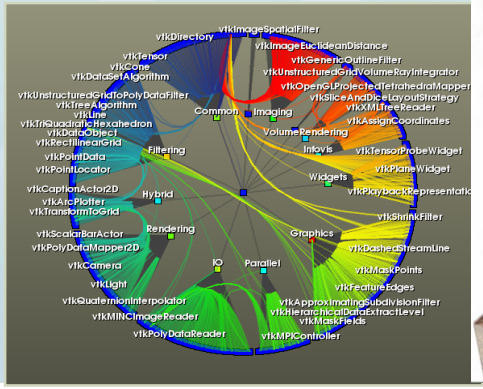
Software Services





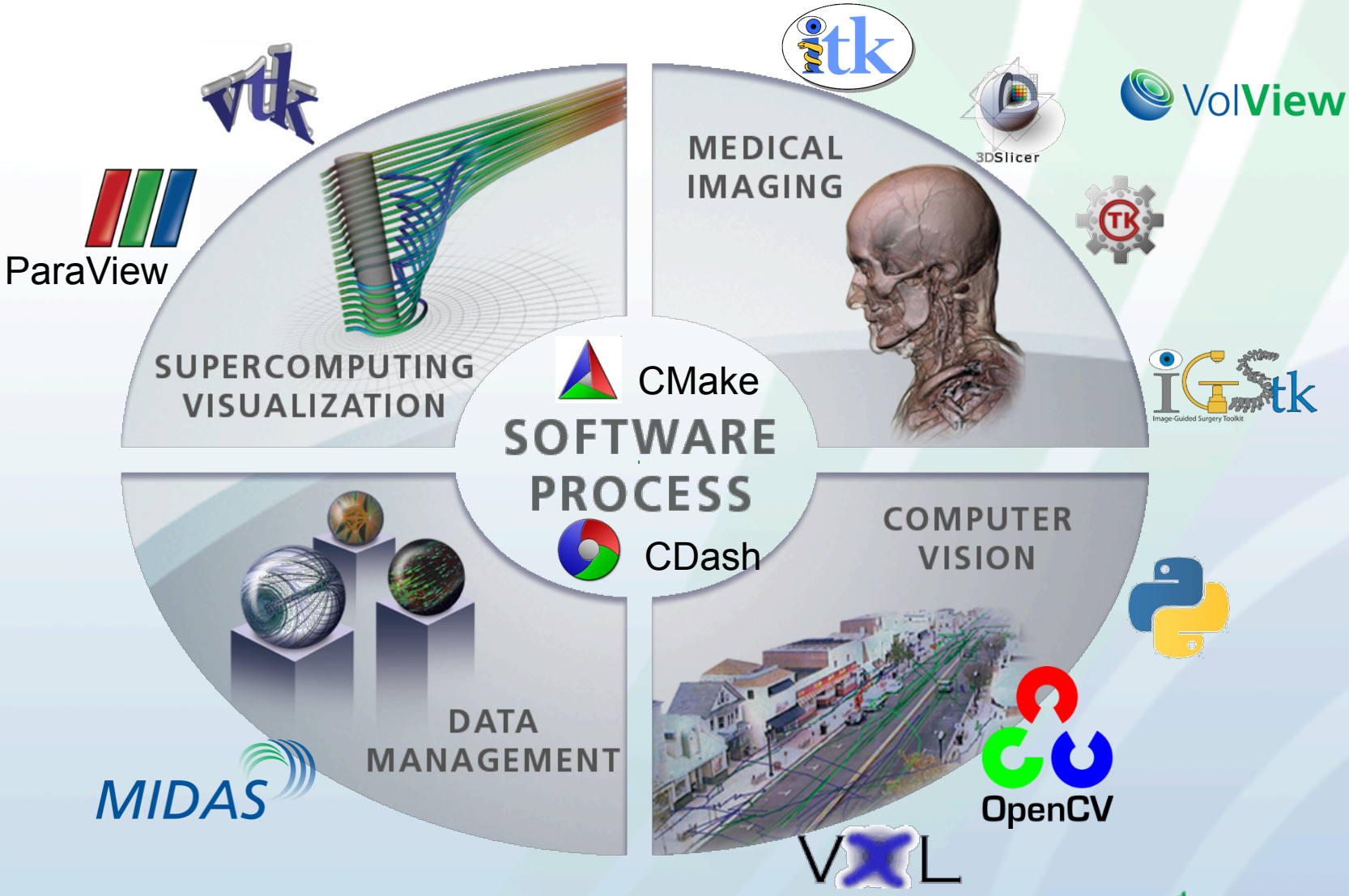




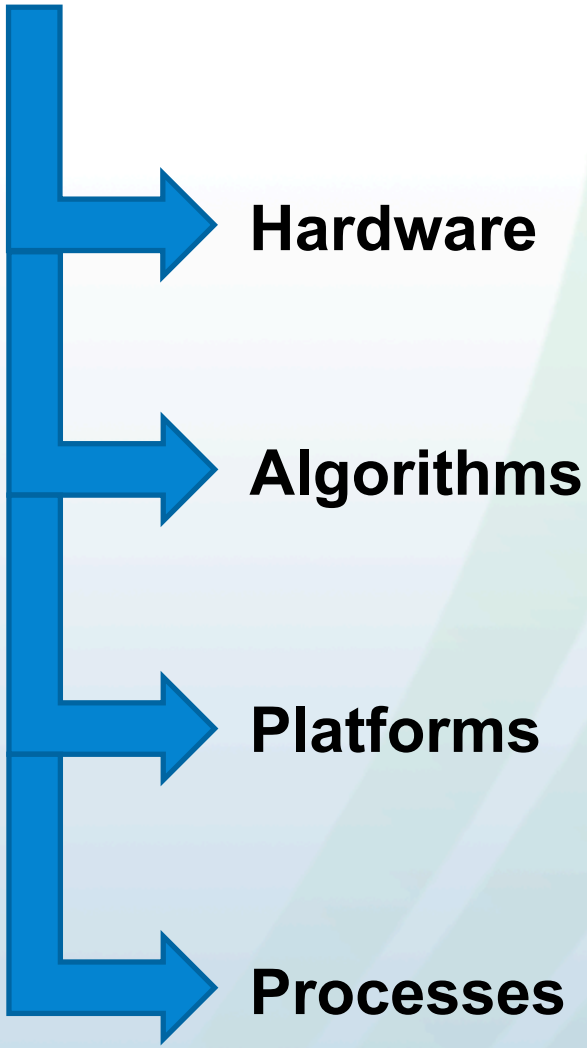


Kitware

Kitware

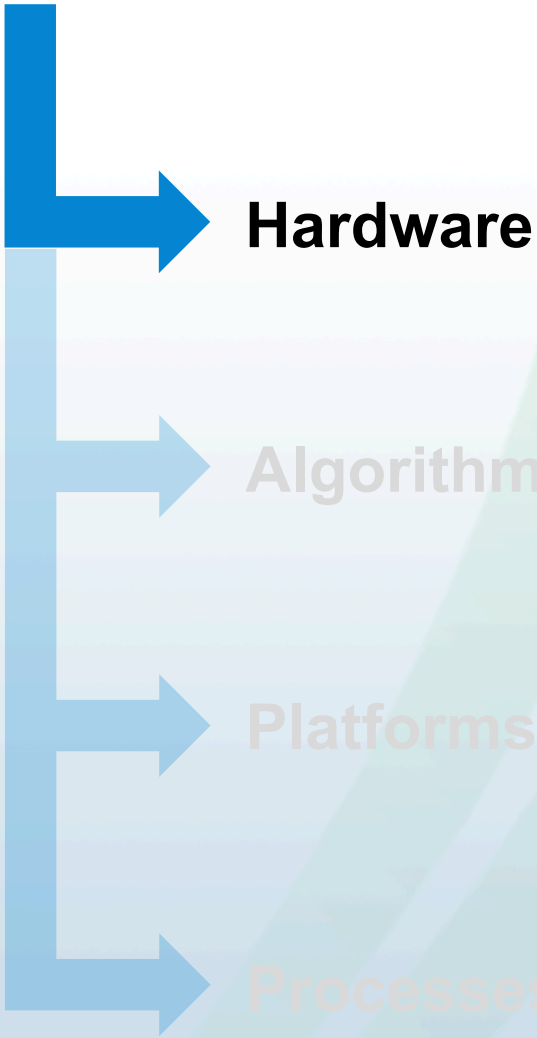

vtk
itk
3DSlicer
VolView
ParaView
MEDICAL IMAGING
SUPERCOMPUTING VISUALIZATION
CMake
SOFTWARE PROCESS
CDash
COMPUTER VISION
DATA MANAGEMENT
MIDAS
OpenCV
VXL
Kitware

SCALE



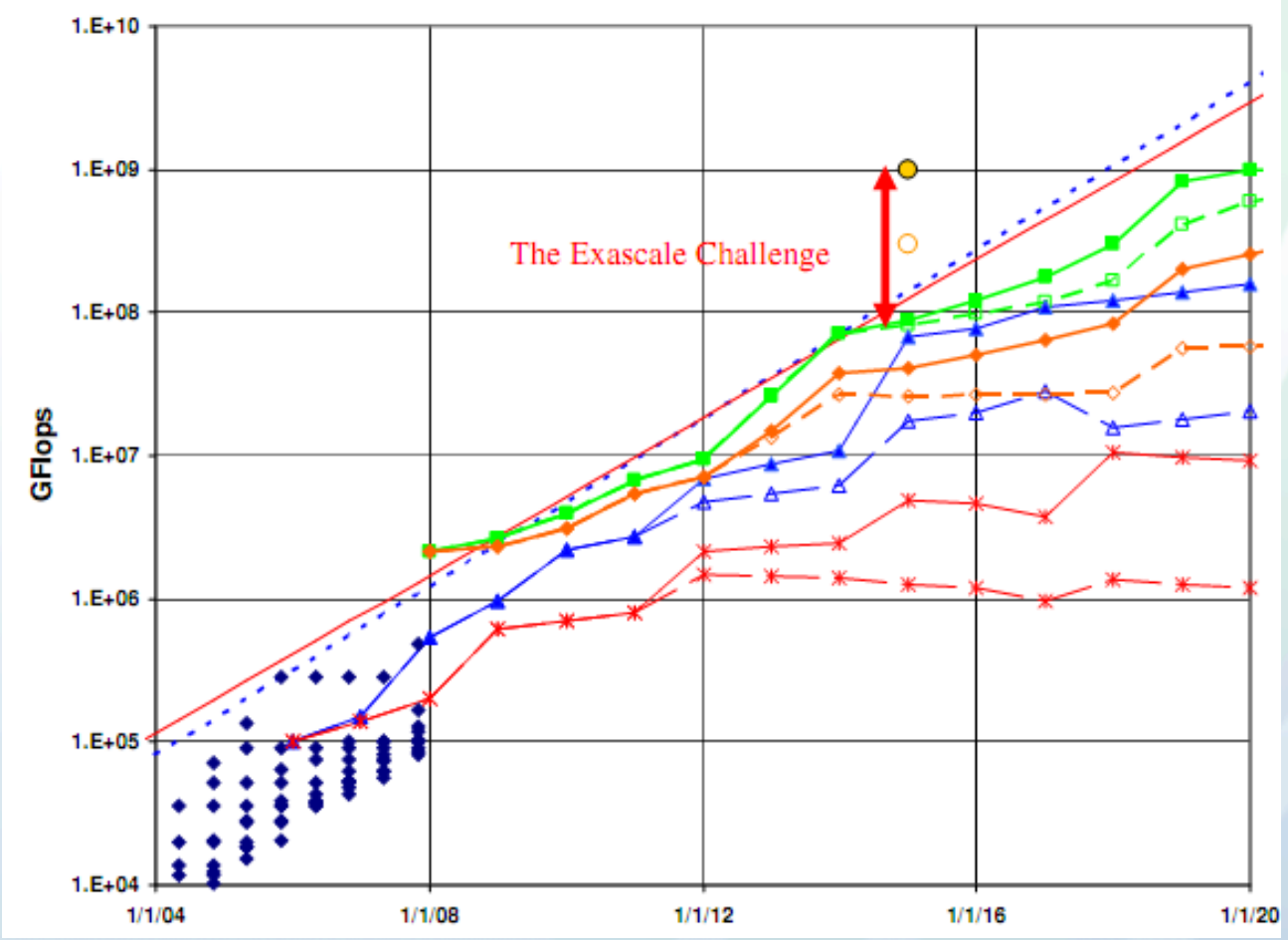
Kitware

SCALE



Kitware

Power Challenge



Kitware

The Traditional Visualization Workflow is Breaking Down

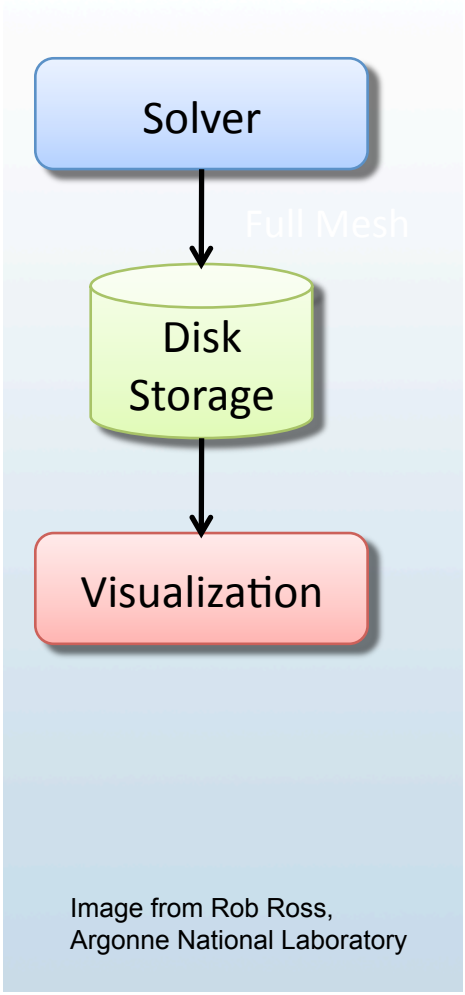
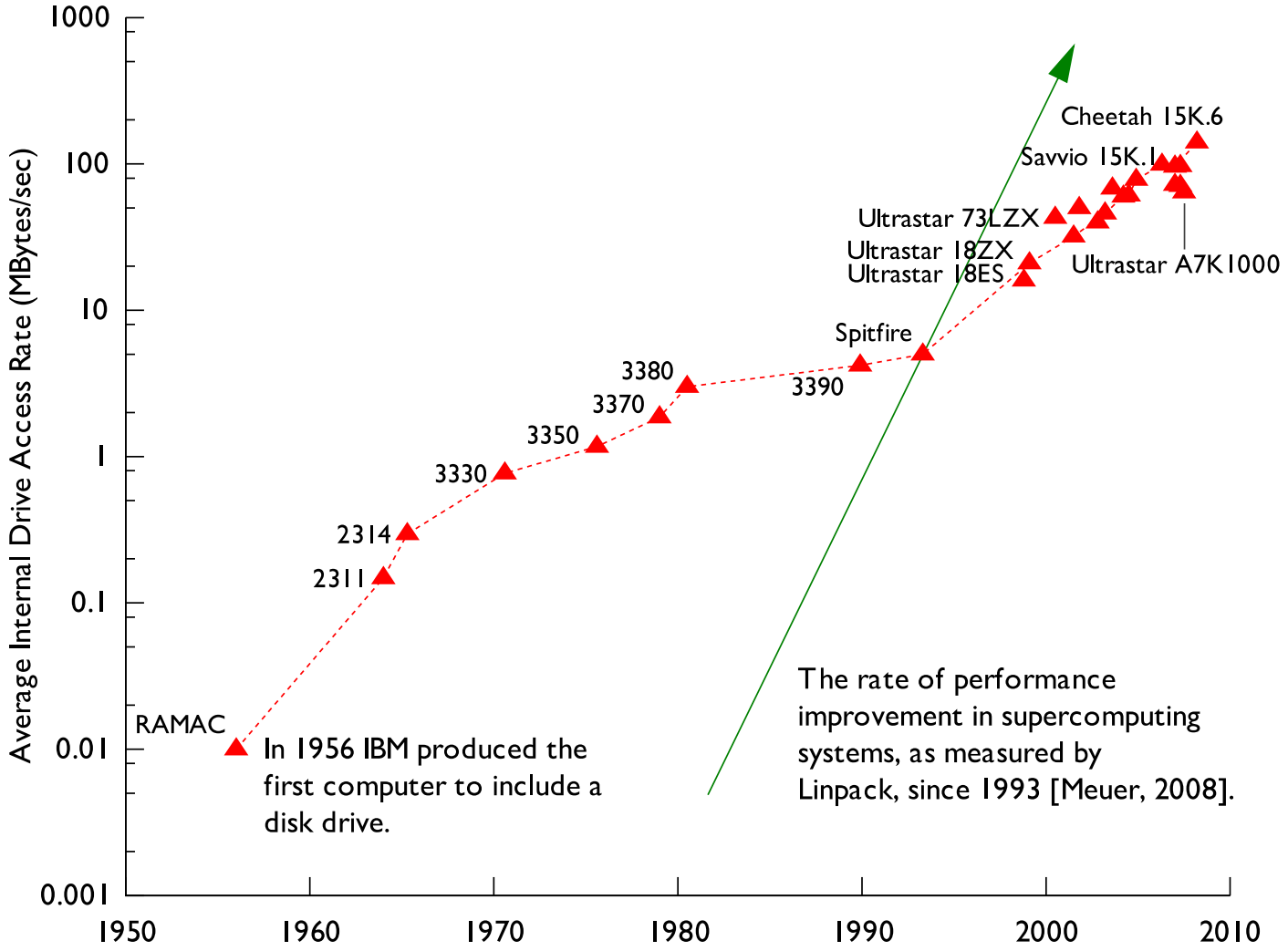


Image from Rob Ross, Argonne National Laboratory



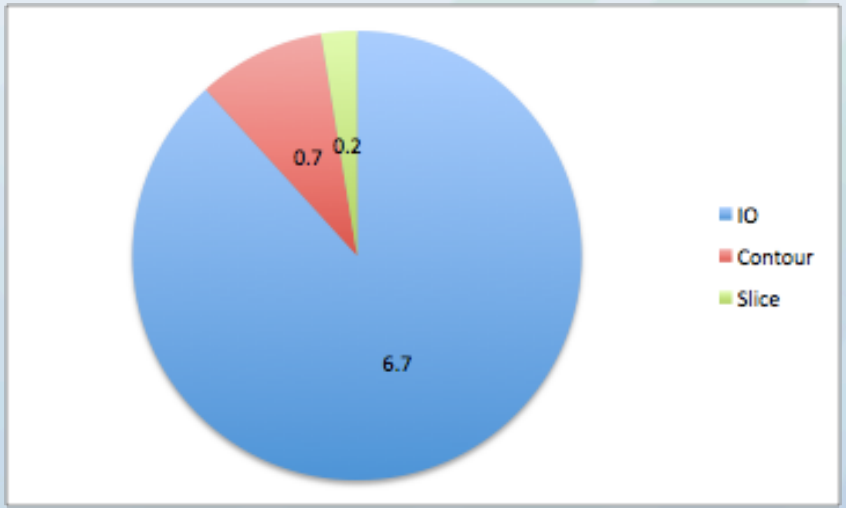
The Problem with IO: A Small Example

Simulation

- 40 million finite elements simulation
- File size: 3.2GB per time step
- 1000 time steps
- 100 time steps written to disk

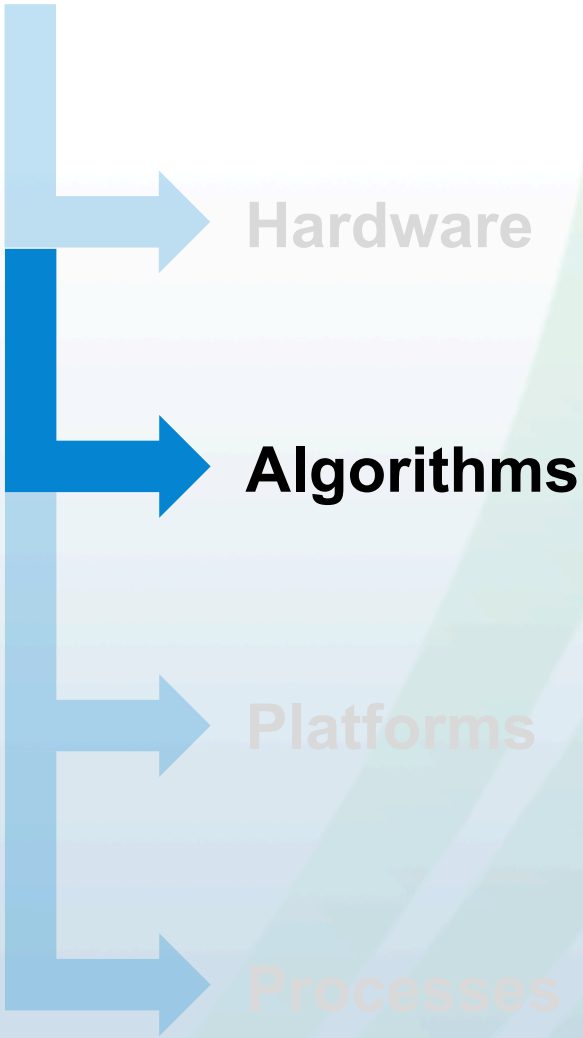
Visualization

- ParaView
- Quad-core Mac Pro with 12 GB memory
- IO: 240 secs
- Contour: 25 secs
- Slice: 7 secs

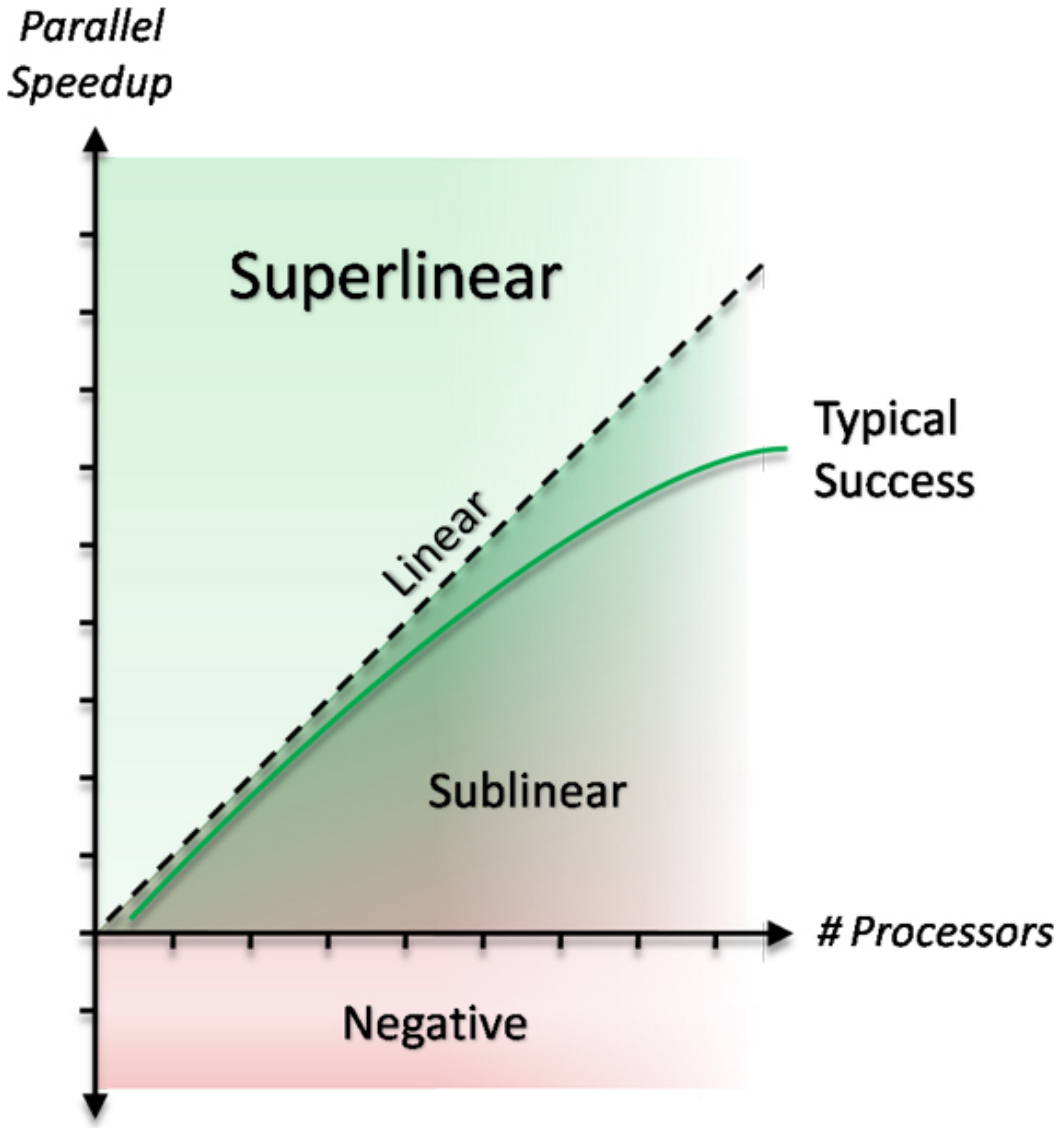


Kitware

SCALE

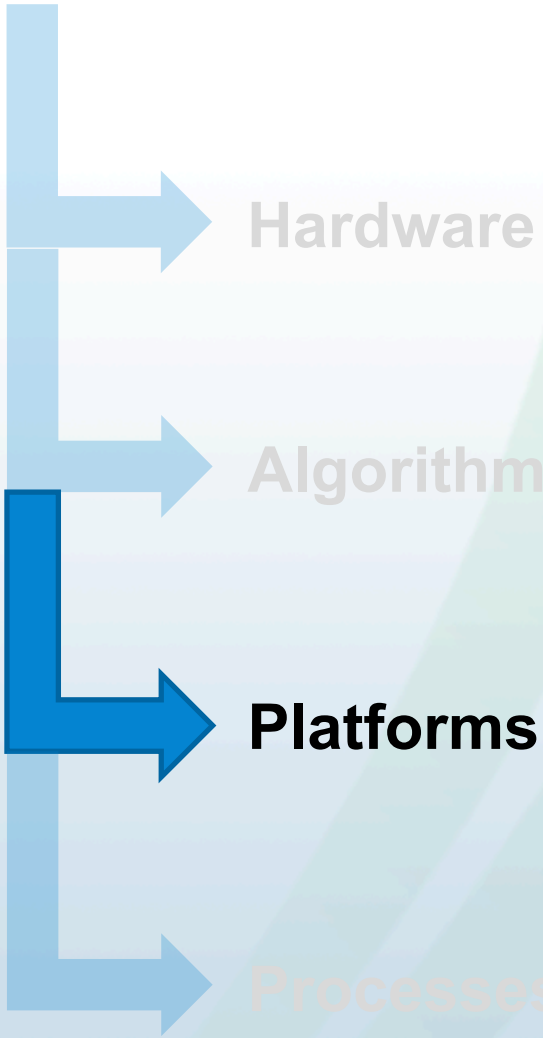


Kitware

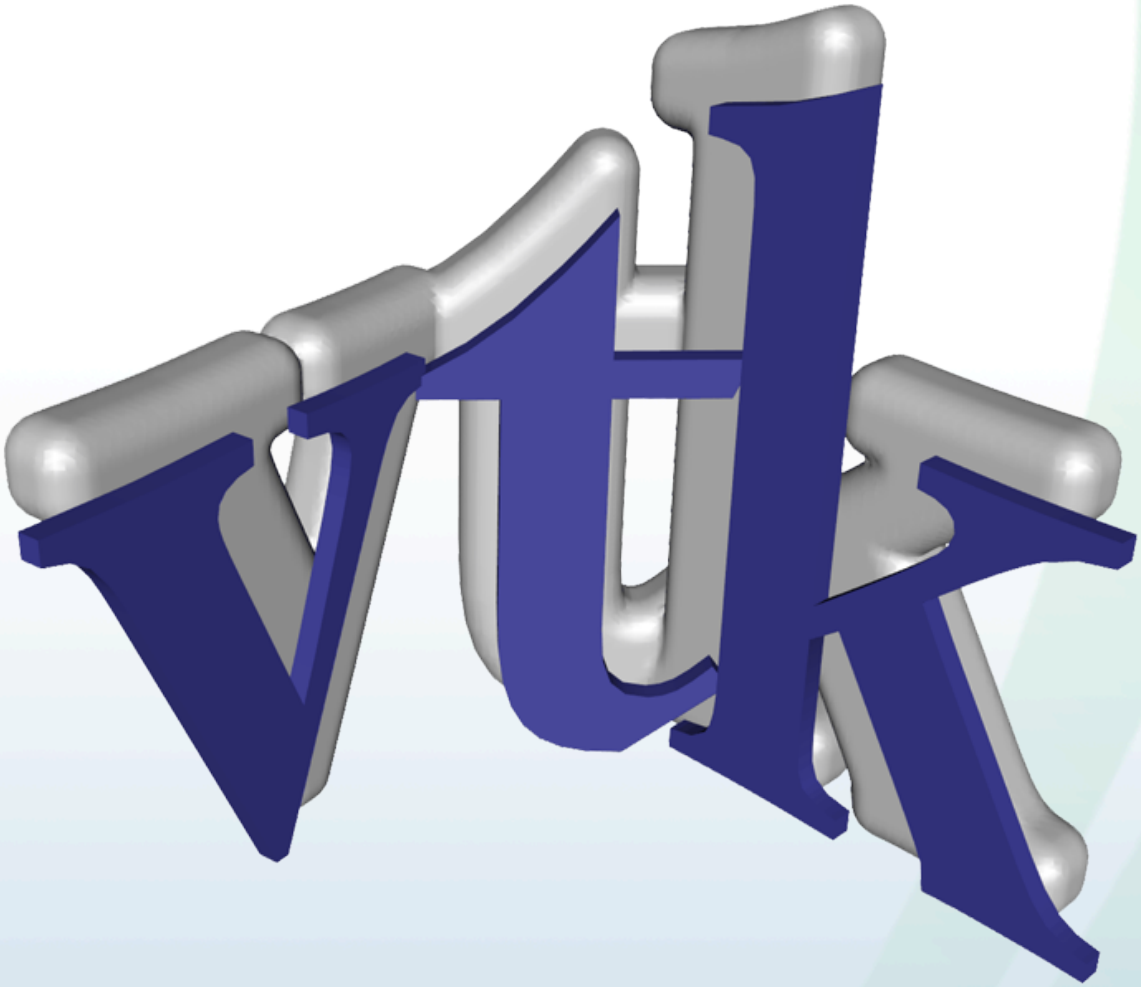

Parallel
Speedup
Superlinear
Linear
Typical
Success
Sublinear
Processors
Negative

Kitware

SCALE



Kitware

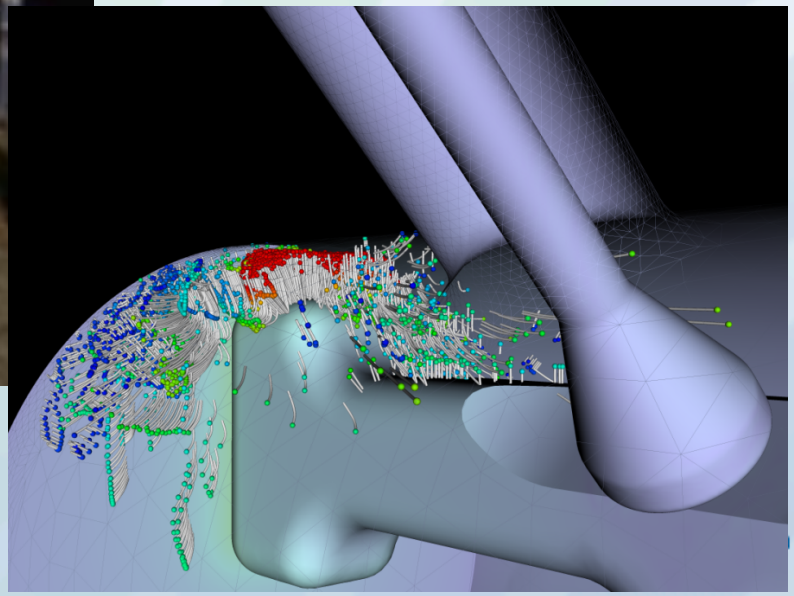


20 years old – Dec. 2013

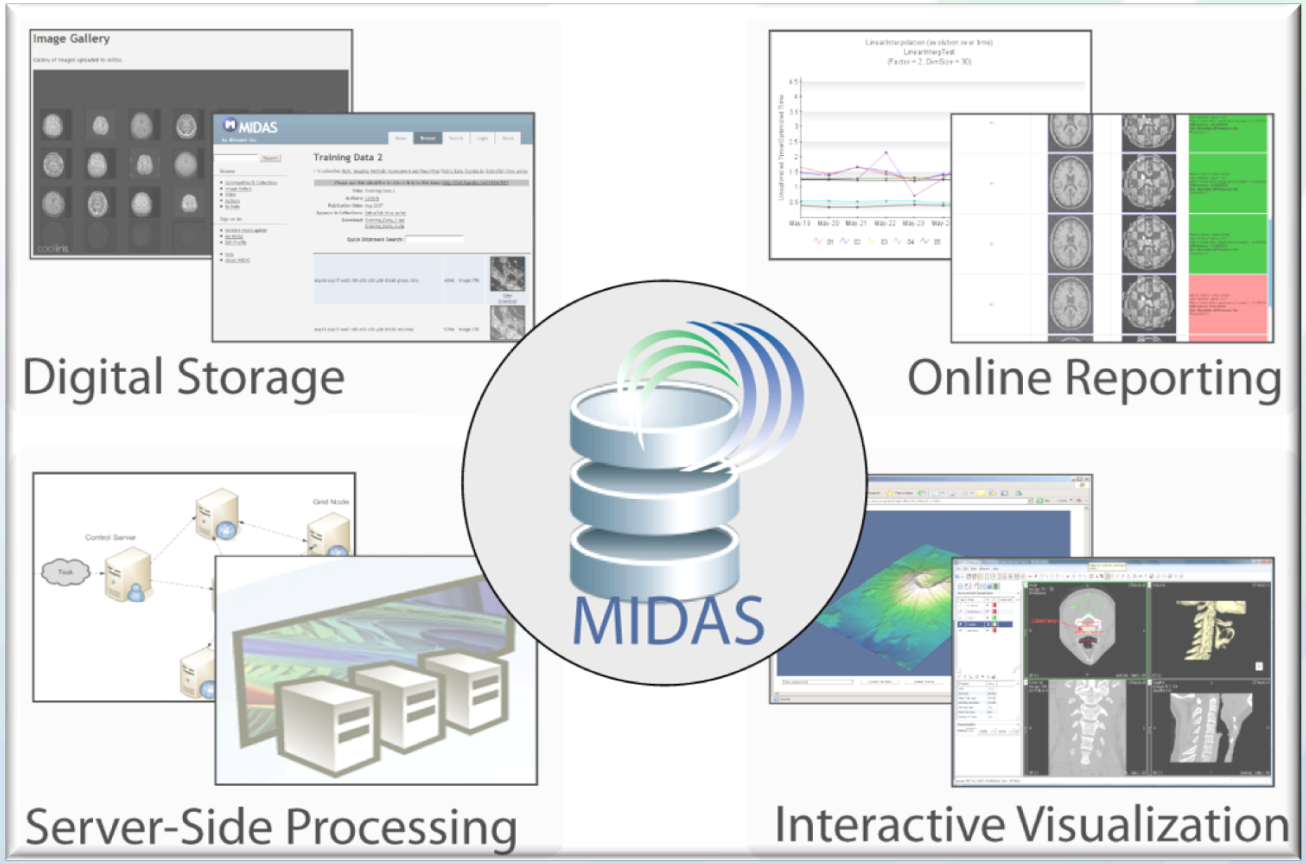
Kitware

ParaView Won HPCwire Readers' and Editors' Choice Awards (Again)

ParaView

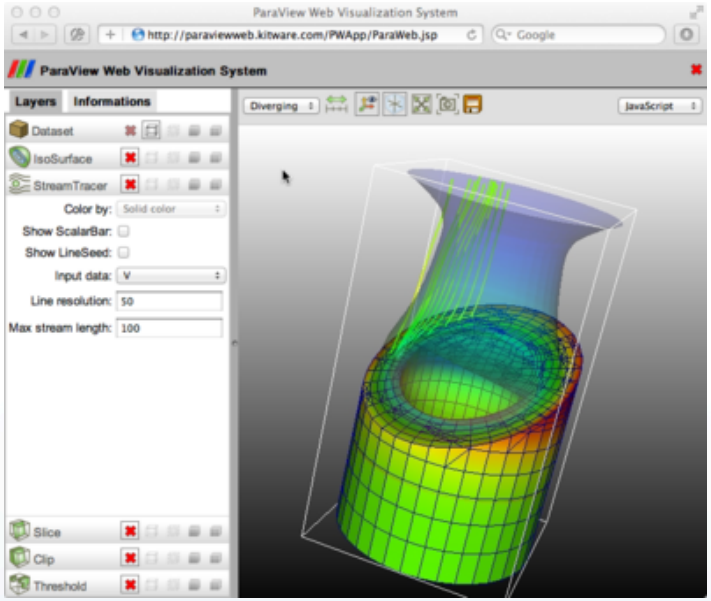


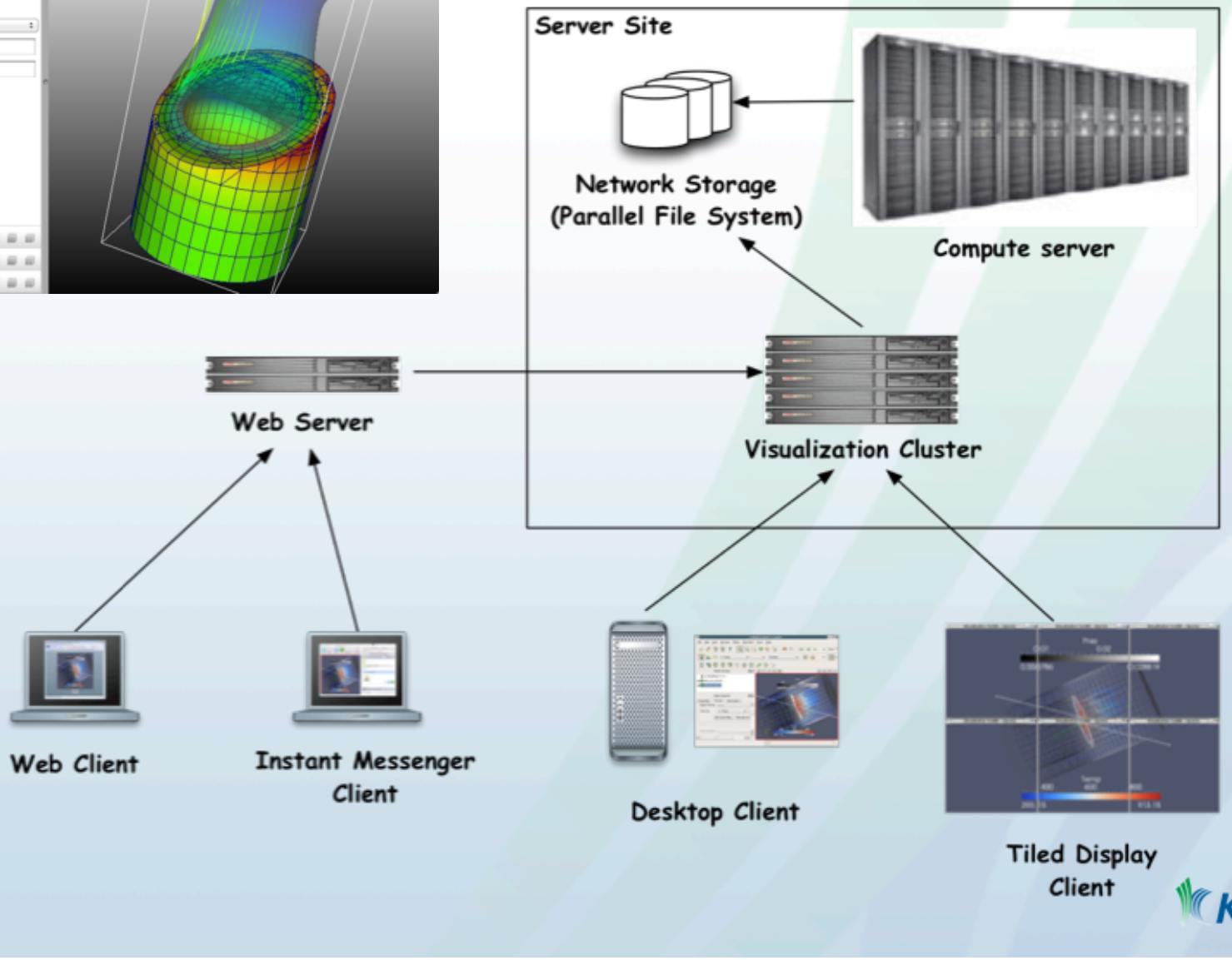
Data-Centric Computing



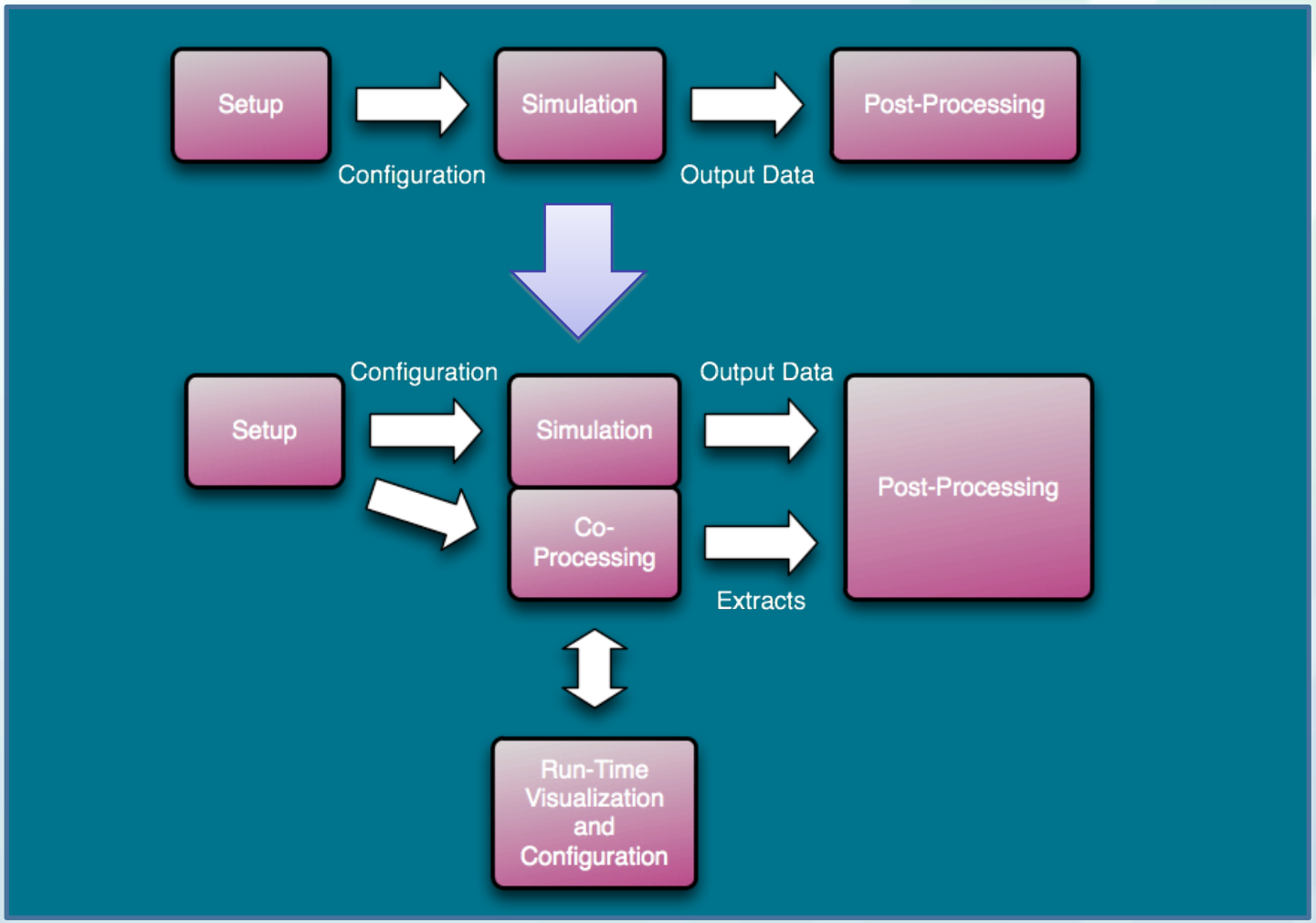
Kitware

ParaViewWeb



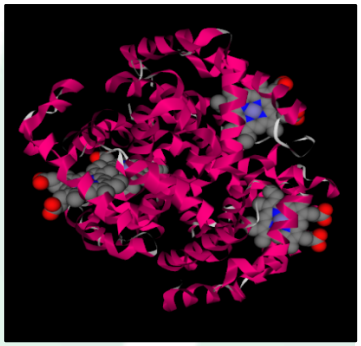


Co-Processing (Catalyst)



Kitware

Computational Chemistry



Log File

Input File

Edit/Analyze

Results

ChemData

Job Submission

ezhpc

ezhpc

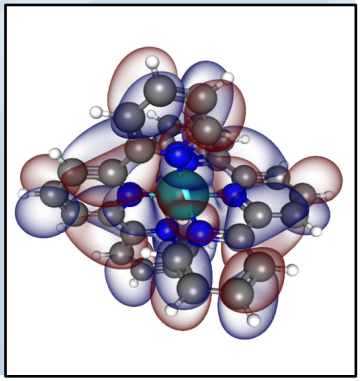
MoleQueue

Local

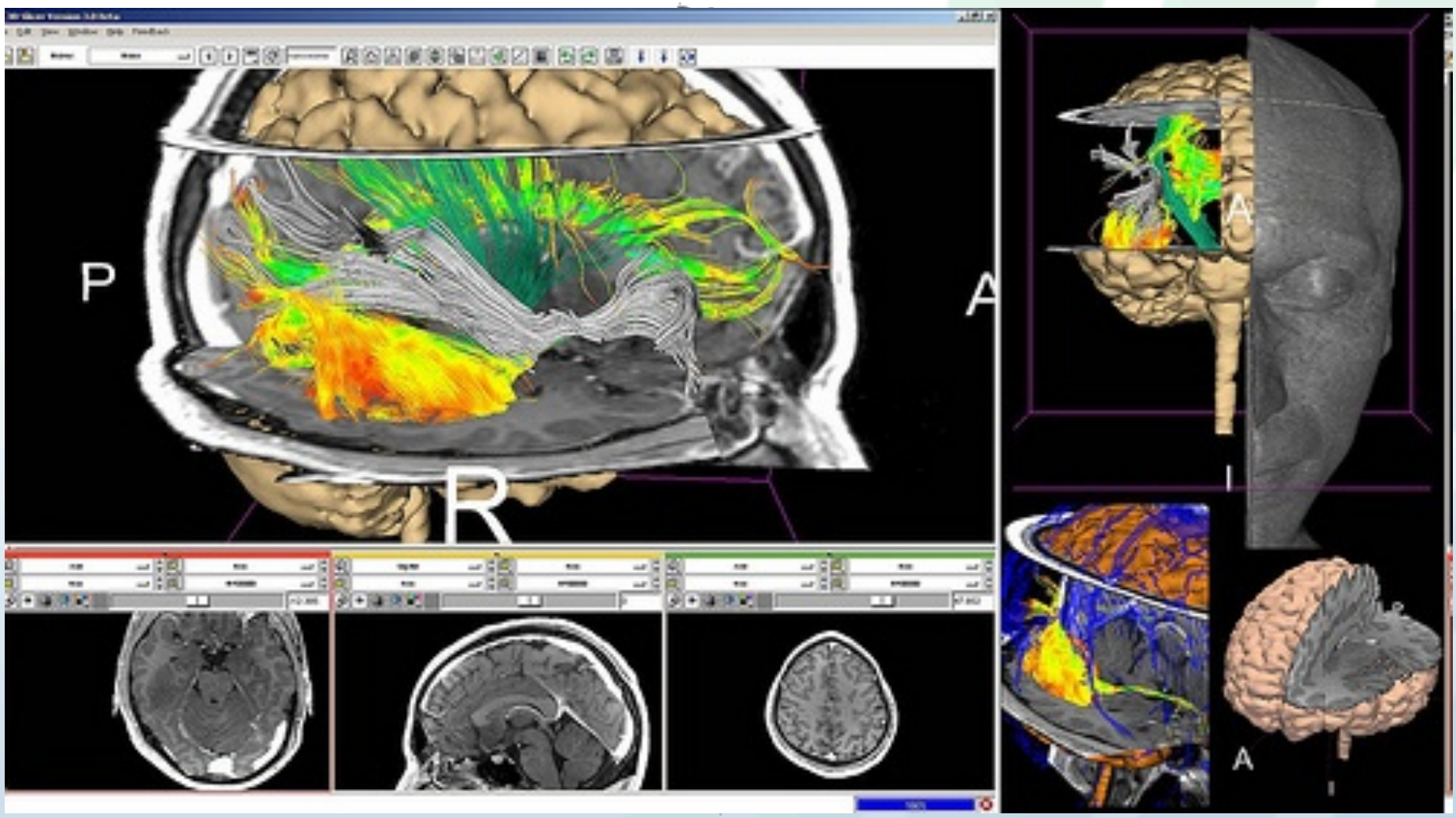
Remote

Calculation

GAMESS

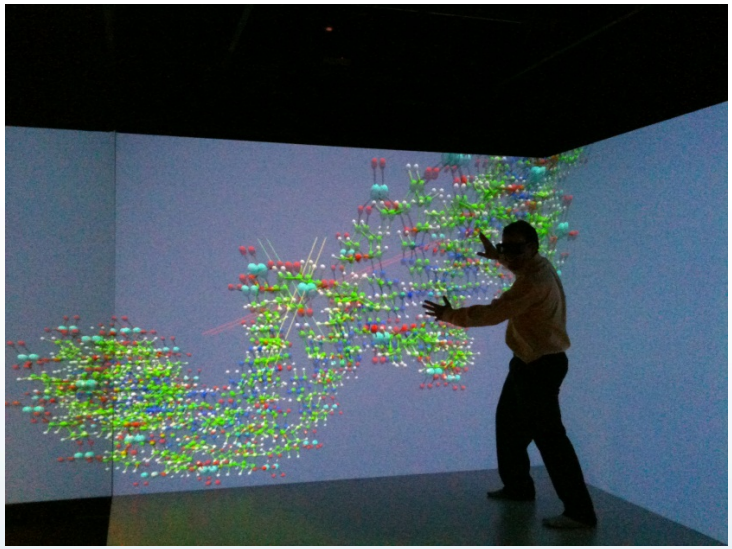


3D Slicer – Medical Image Analysis

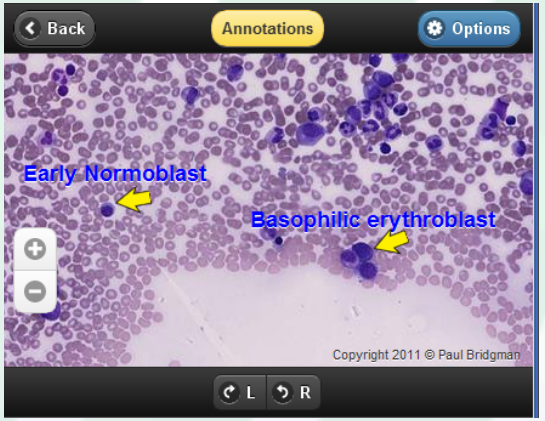


Kitware

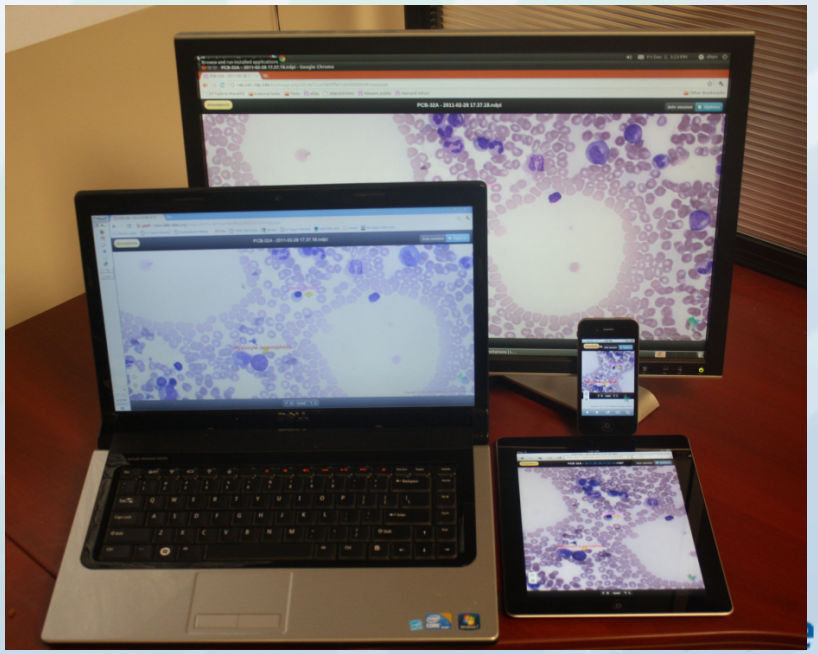
Mobile to Supercomputer



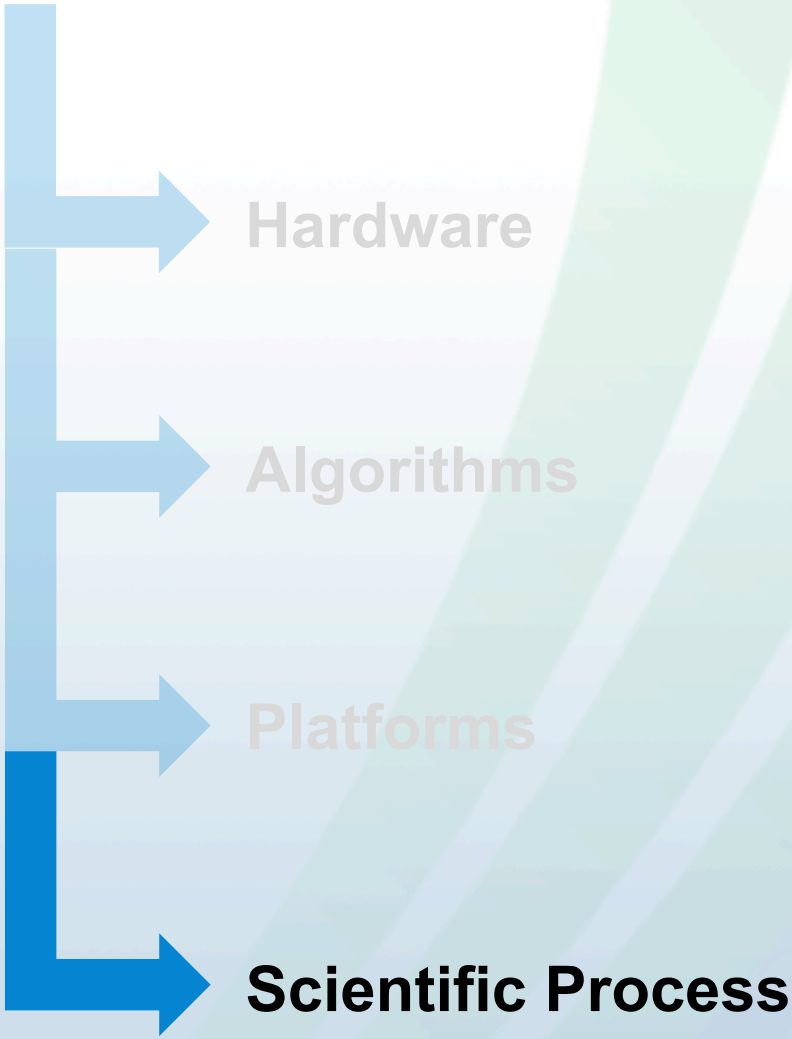
ParaView



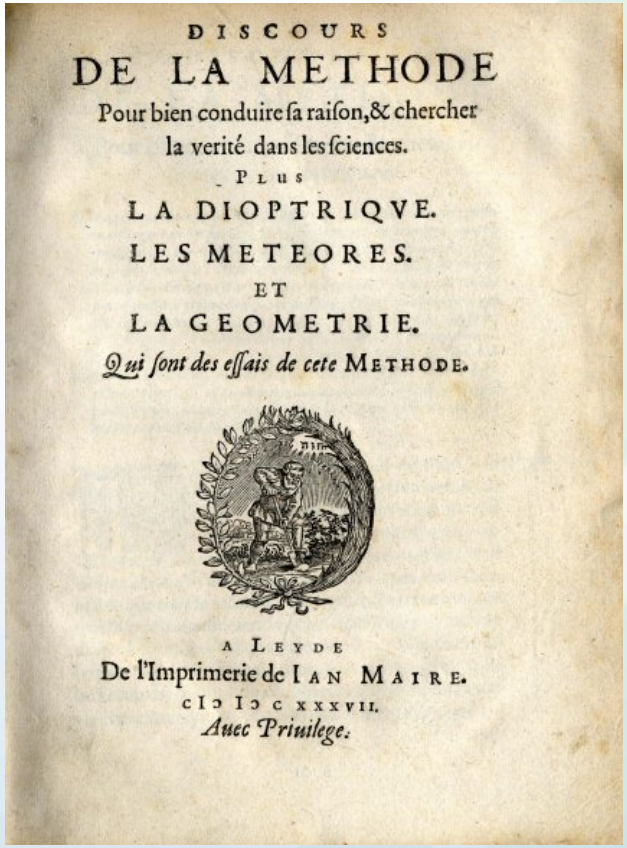
Kiwi / VES



SCALE



Kitware


DISCOURS
DE LA METHODE
Pour bien conduire ſa raiſon,& chercher
la verité dans les ſciences.
PLUS
LA DIOPTRIQVE.
LES METEORES.
ET
LA GEOMETRIE.
Qui ſont des eſſais de cete METHODE.

A LEYDE
De l'Imprimerie de IAN MAIRE.
CIↃ IↃ C XXXVII.
Auec Priuilege.

Discourse on the (Scientific) Method,
Descartes 1637

Kitware

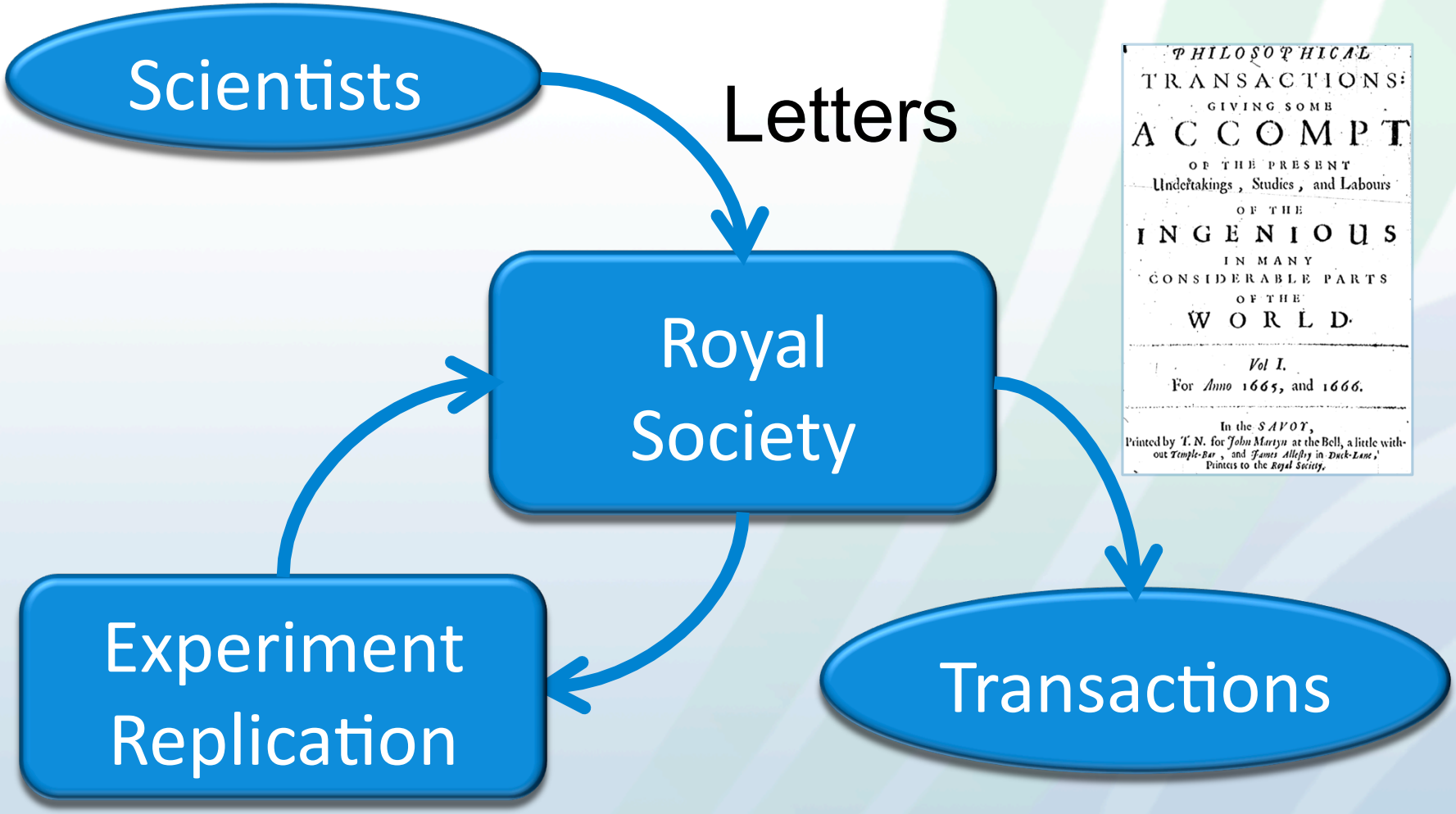
If it's not reproducible, it's not Science

Nullius in Verba

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

Kitware

Scientific Publishing Origins



PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.

Kitware

Evolution

Kitware


Happy Ending!
....or is it?

Peer-reviewers are expected to ***replicate*** *experiments*

Kitware


The thirst for
{ fame | power | control | money }

Career Pressures

"Publish or Perish" or what they taught me in Graduate School

Author

Kitware

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.

Kitware

Example Reproducibility Challenge: White Matter Tracts in Medical Imaging (DTI Imaging at *MICCAI 2011)*

- 8 international teams participated
- 3D visualization and standardized comparison of different tractography
- All used the same diffusion MRI dataset

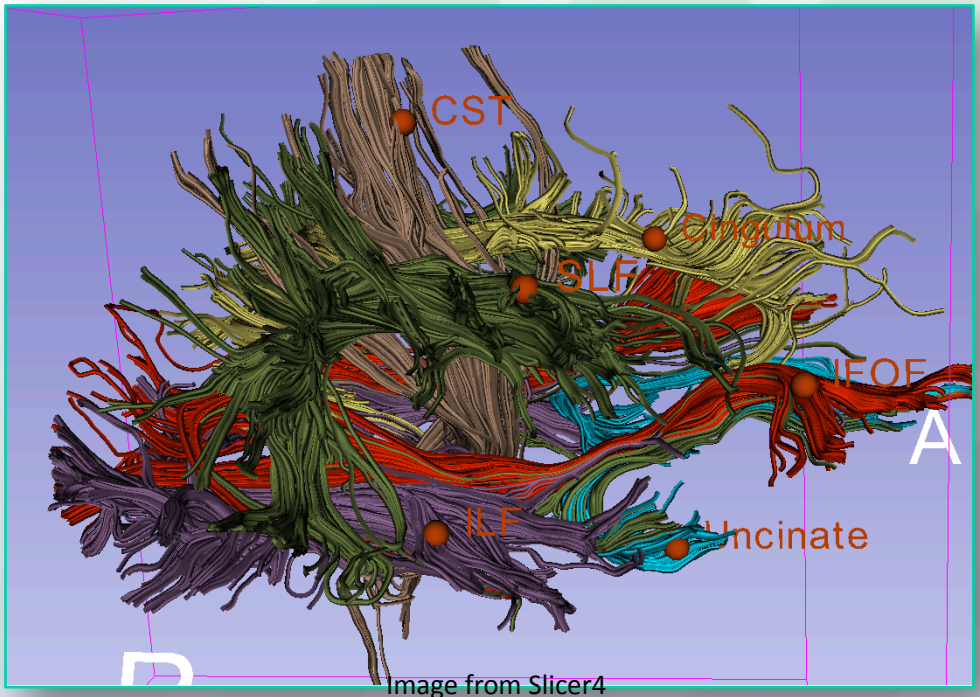
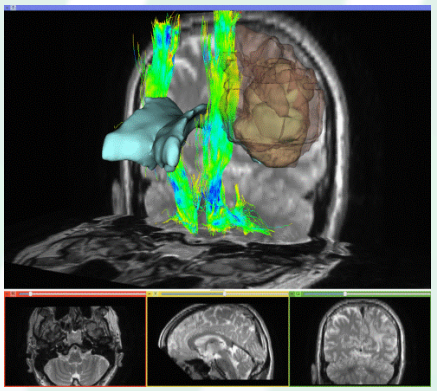


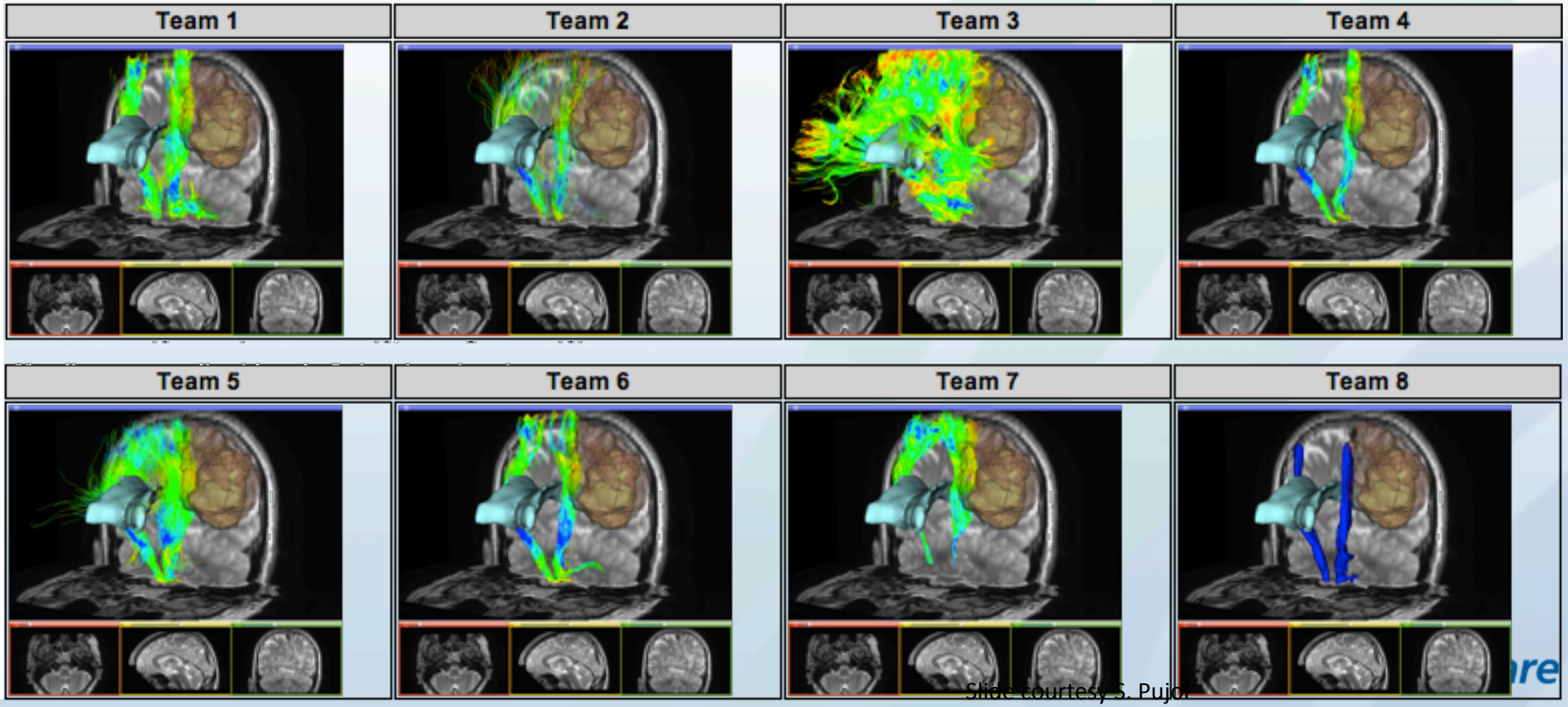
Image from Slicer4

Kitware

MICCAI Workshop Results

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





Slide courtesy S. Pujol

Publishing in the Modern Age?

- Time to post a PDF file on the Web
  - Typically 1 hour, ~0 marginal cost

vs

- Time to publish a paper in a journal
  - Typically 2 years
- Cost to publish a paper in a journal
  - About 500€ / paper
- Cost to read the same paper
  - About 30€ / paper

Kitware

Publishing:
Some Economic Repercussions

- **Subscription costs are out of control**
  - **Harvard University:** canceling “too expensive” journal subscriptions due to expense. Asking professors to publish in open access journals.
  - **UK:** Minister of Science David Willetts that all publicly funded research should be published as open access
  - **World Bank** announced that all existing and new publications, reports and documents will be open access by July 2012.
  - **Boycott of Elsevier:**
    - **E.g., In 2011: > $7K for a subscription to *Theoretical Computer Sciences***

Threatening access to scientific results

Kitware

Intellectual Property Pressures

Pros

- Estimated licensing revenue of patents to US Universities $2 billion

Cons

- After subtracting costs, net revenue $0 - $600 million remains to Universities (est.)
- Results in a corrupting influence on science
- Creates resistance to collaboration
- Reduces academic-industrial collaboration

Bayh-Dole Act (US): *enabled and encouraged academic institutions to patent the IP from government funded research. Gave rise in the US to University Patent & Licensing Operations*.

Kitware

SCIENTIFIC AMERICAN™

Permanent Address: http://www.scientificamerican.com/article.cfm?id=secret-computer-code-threatens-science

Secret Computer Code Threatens Science

Missing source code can allow bad science to slip through the cracks and means extra headaches for scientists who want to closely follow up on new studies or check for errors

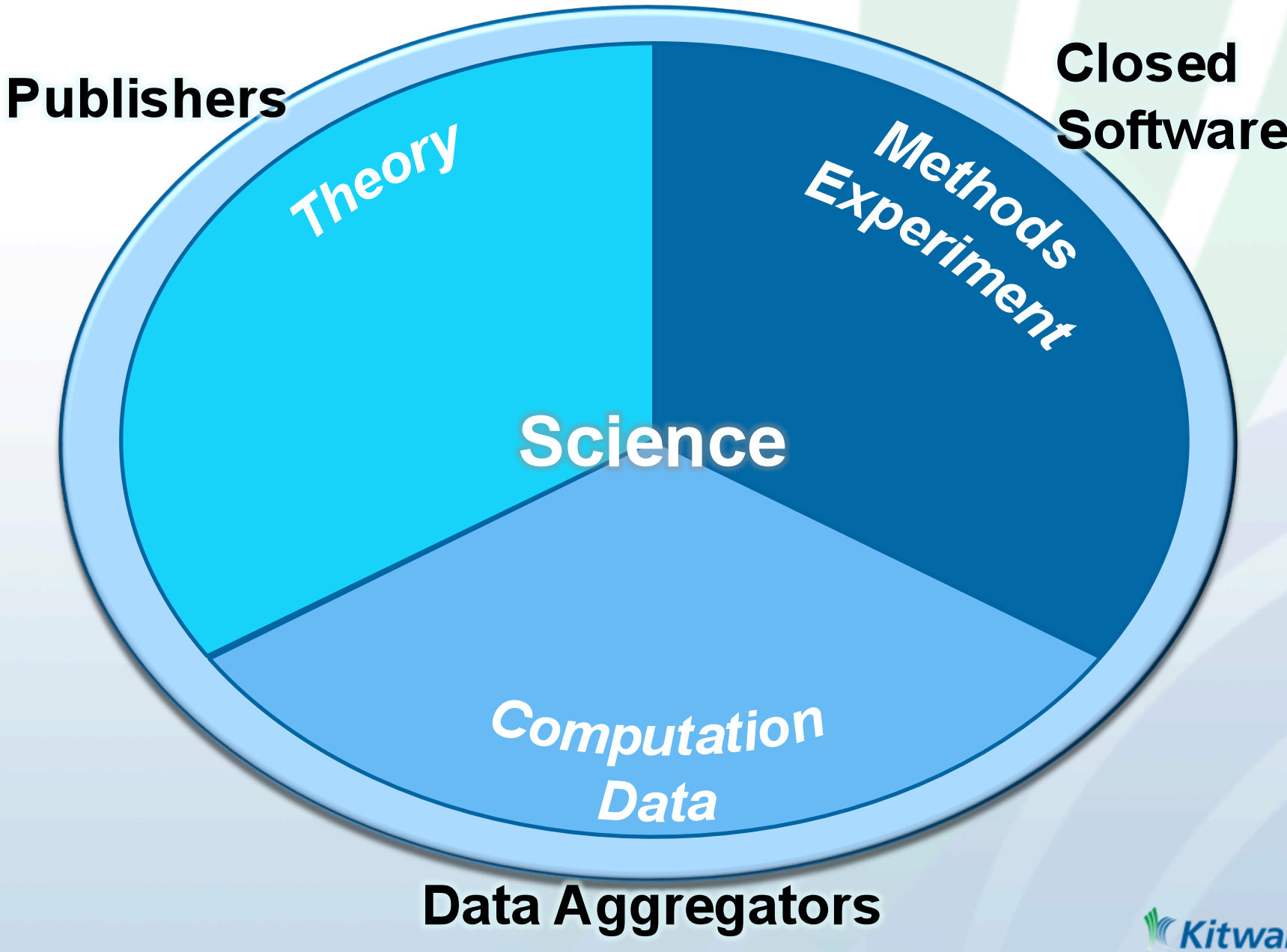
By Jeremy Hsu and InnovationNewsDaily | Friday, April 13, 2012 | 49 comments

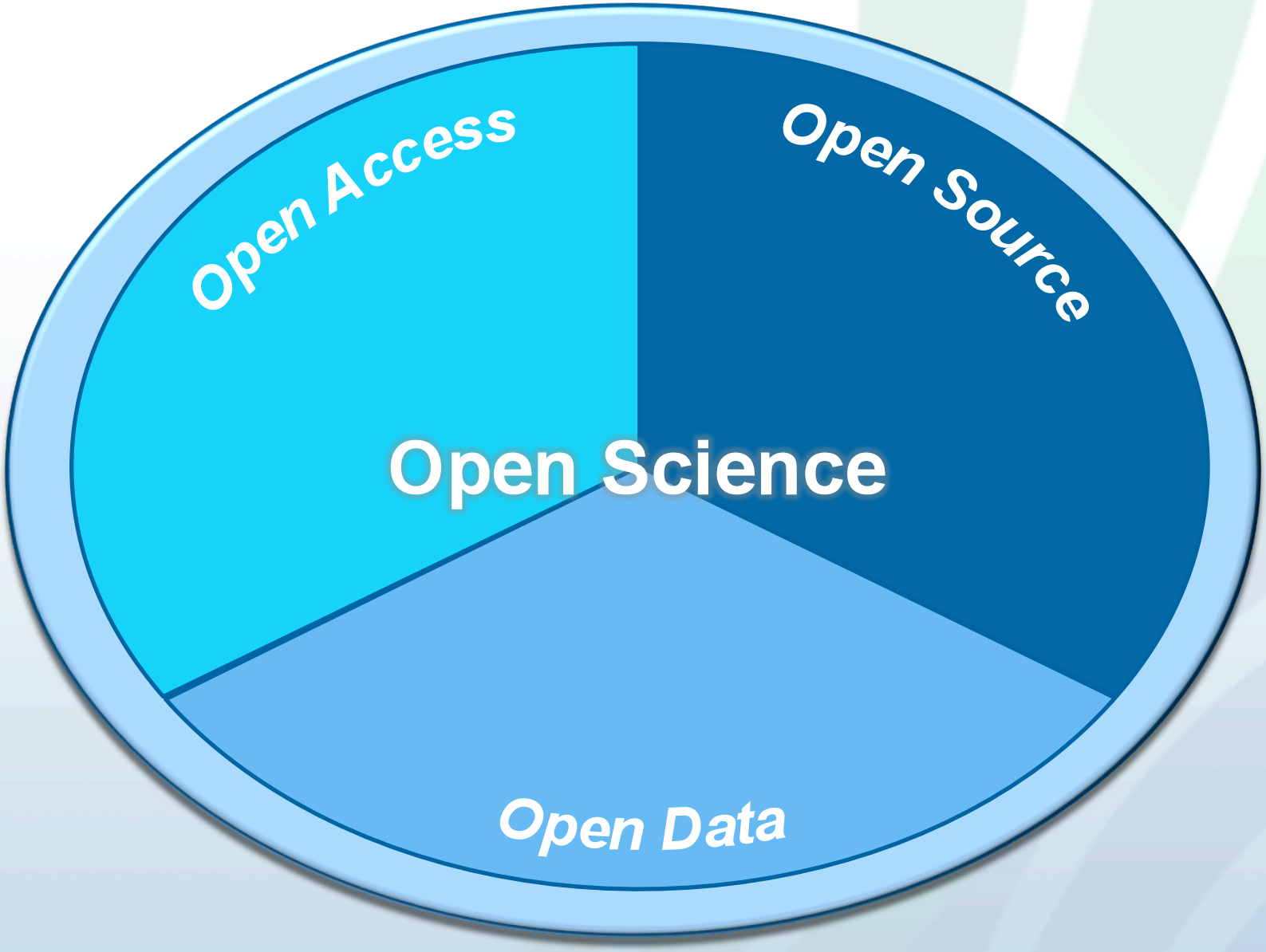
Modern science relies upon researchers sharing their work so that their peers can check and verify success or failure. But most scientists still don't share one crucial piece of information — the source codes of the computer programs driving much of today's scientific progress.

Such secrecy comes at a time when many researchers write their own source codes — human-readable instructions for how computer programs do their work — to run simulations and analyze experimental results. Now, a group of scientists is arguing for new standards that require newly published studies to make their source codes available. Otherwise, they say, the scientific method of peer review and reproducing experiments to verify results is basically broken.

Public Policy Controversies

Kitware


Publishers
Closed Software
Theory
Methods Experiment
Science
Computation Data
Data Aggregators
Kitware

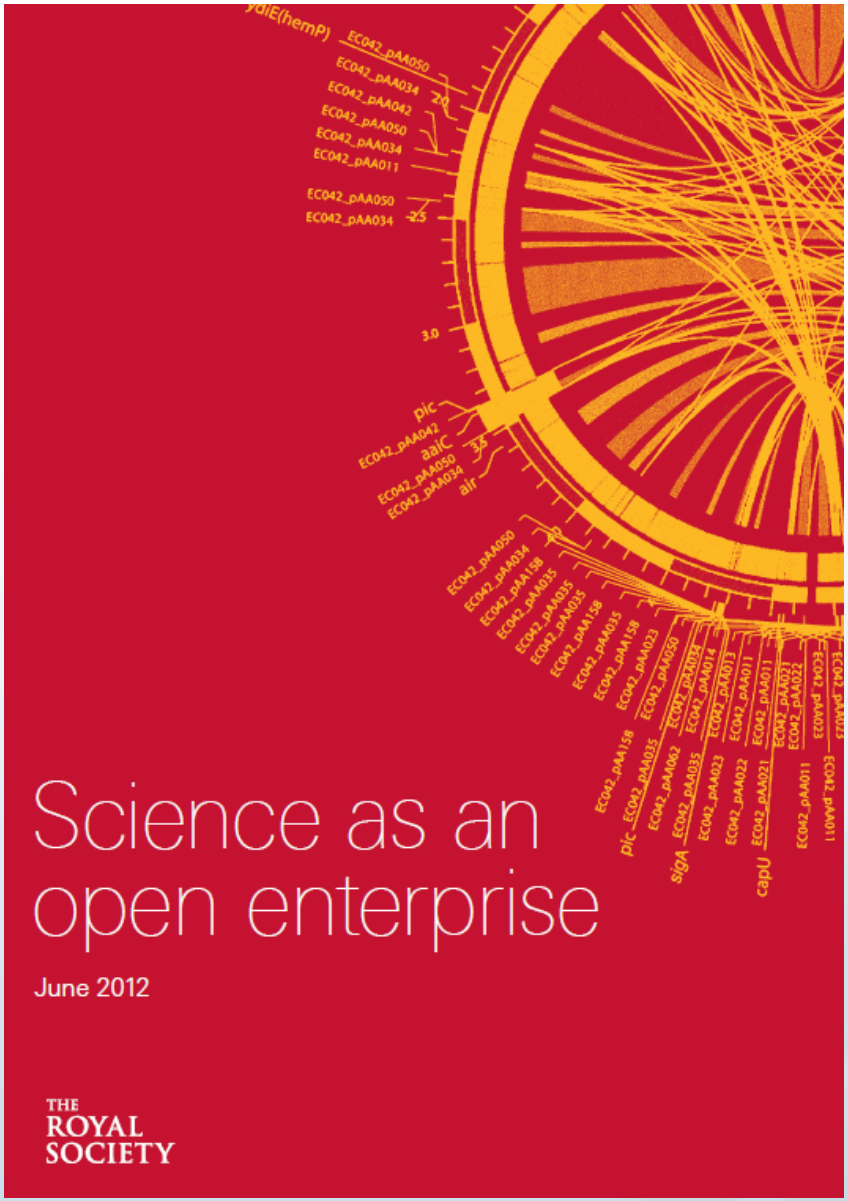


Kitware

Open Science

- Ensures reproducibility through unfettered access
  - Access to publications **(open access)**, data **(open data)**, and methods **(open source)**
- Enables easier commercialization (since implementation details are known and available)
  - Provides significant cost savings
  - Enables a service-based business model
- Addresses the demands / complexity of modern computational methods
  - Facilitates the combination of software systems, libraries and toolkits
- Fosters collaboration
  - Removes barriers to exchanging information

Kitware



"Science's capacity for self-correction comes from its openness to scrutiny and challenge."

-- Geoffrey Boulton

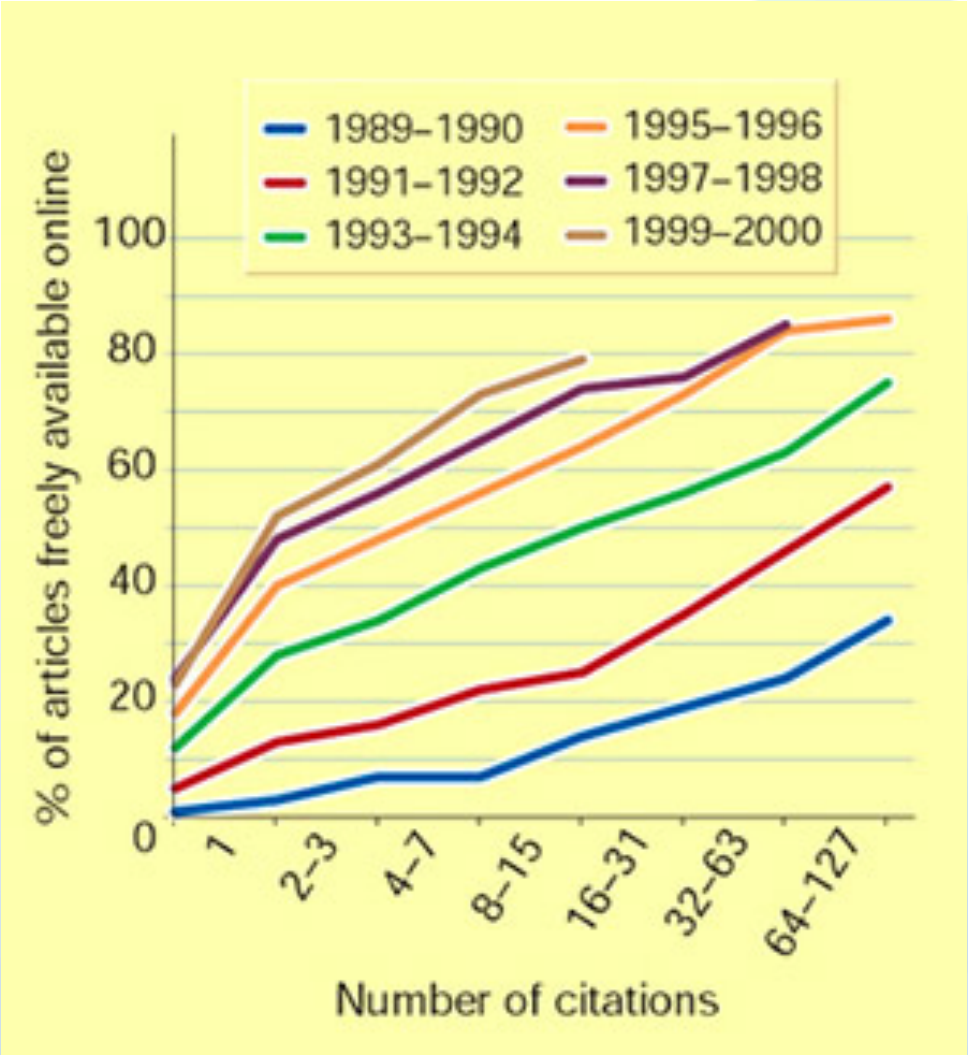
Kitware

Royal Society (recommendations): *Science as an Open Enterprise*

1. Scientists should communicate the data they collect and the models they create, via methods that allow free and open access.
2. Dataset metrics should give credit by using internationally recognized standards for data citation.
3. Learned societies and academies should promote collaboration to exploit the opportunities provided by more effective data sharing.
4. Learned societies and academies should promote the benefits of new data sharing tools.
5. The common data policies for research councils in the UK need to be updated within the next year to require data management plans.
6. As a condition of publication, scientific journals should progressively enforce requirements for traceable and usable data available through an article, when they are intrinsic to the arguments in that article.
7. Publishers should encourage and support incentives for the citation of datasets.
8. Governments should recognize the potential of open data and open science to enhance the excellence of the science base

Kitware

Be Selfish and Share

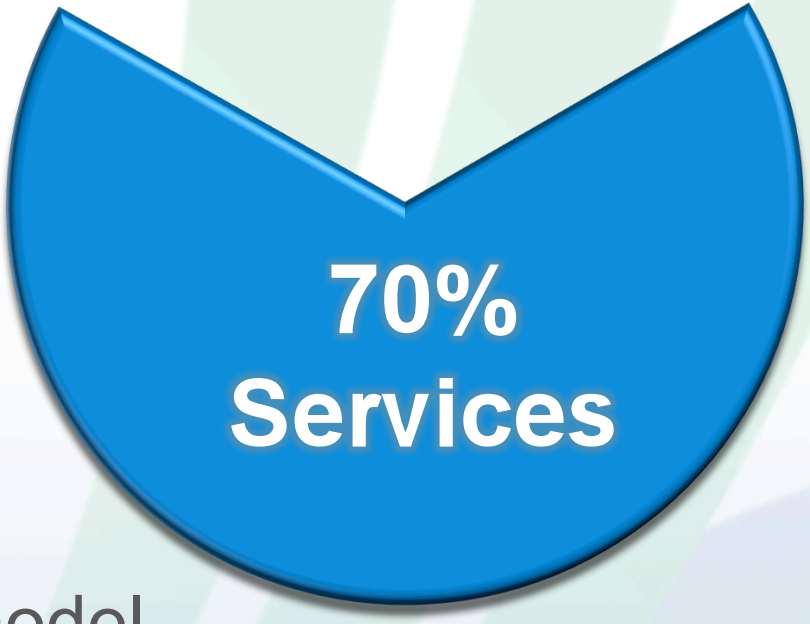


Steve Lawrence, *Nature*

Kitware

Business Trends

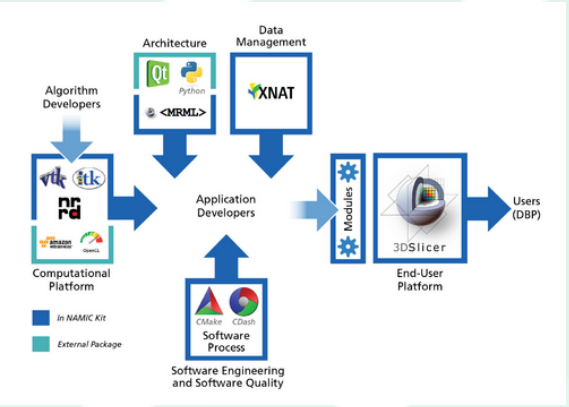
- Services are (est.) 2/3 of total software market



- Growing recognition of the "open source" business model
  - Reduced cost
  - Avoidance of vendor lock-in
  - Agility
  - Improved collaboration
  - Better quality

Kitware

Open Source Value



NAMIC Kit (from Ohloh.net)

Package	Lines of code	Person years	Price tag at 100k per person year
Slicer	1,270,816	361	$36,122,644
KWW	207,208	54	$ 5,406,516
VTK	1,853,529	538	$53,808,076
ITK	848,383	237	$23,719,173
CMake / CTest / CPack	323,454	86	$8,590,888
CDash	78,226	19	$1,897,060
XNat	200,985	51	$5,149,987
Total	4,782,601	1346	$134,694,344

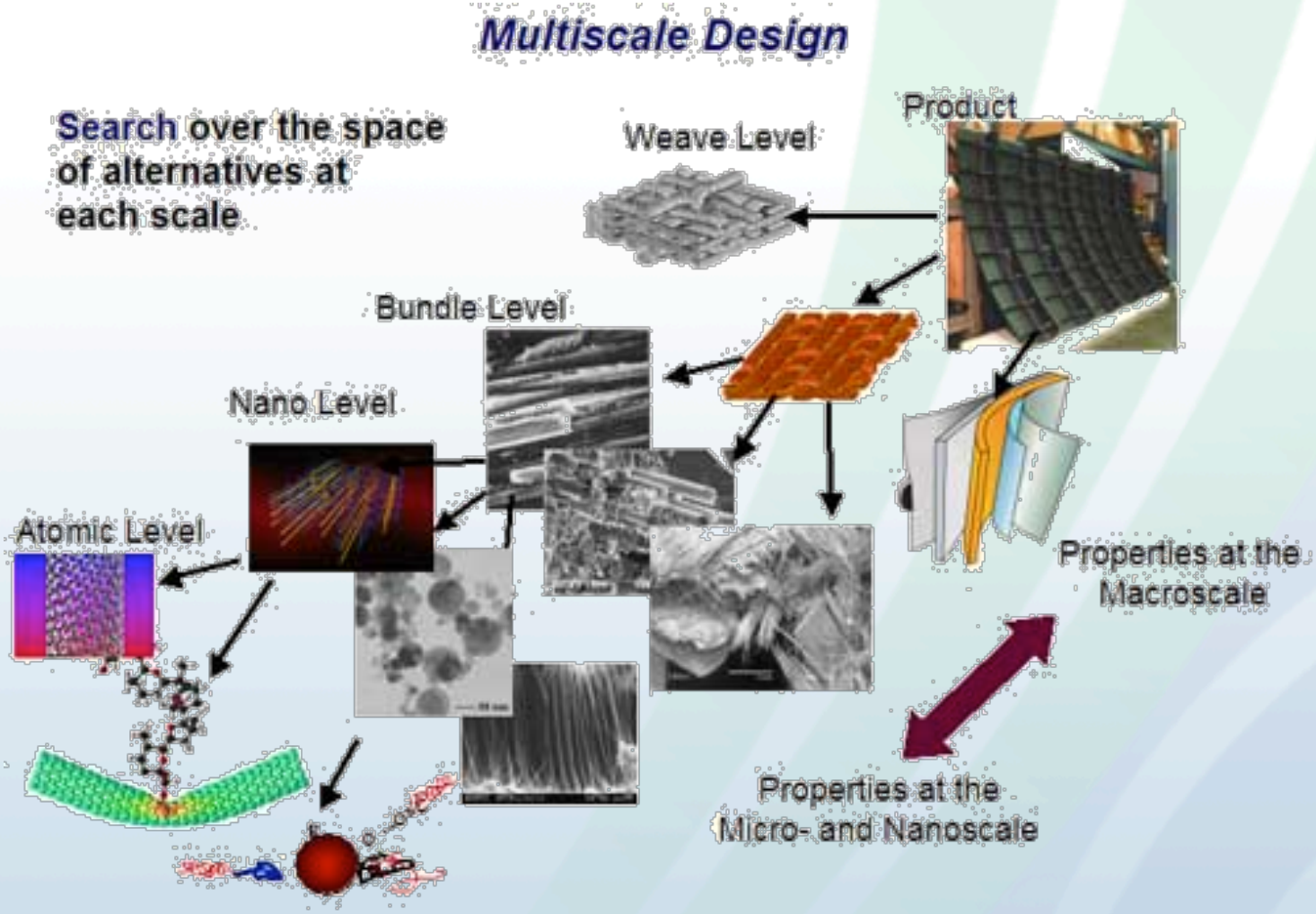
National Alliance for Medical Image Computing

A NIH National Center for Biomedical Computing

Funded under the NIH Roadmap Initiative

Kitware

Multi-Scale

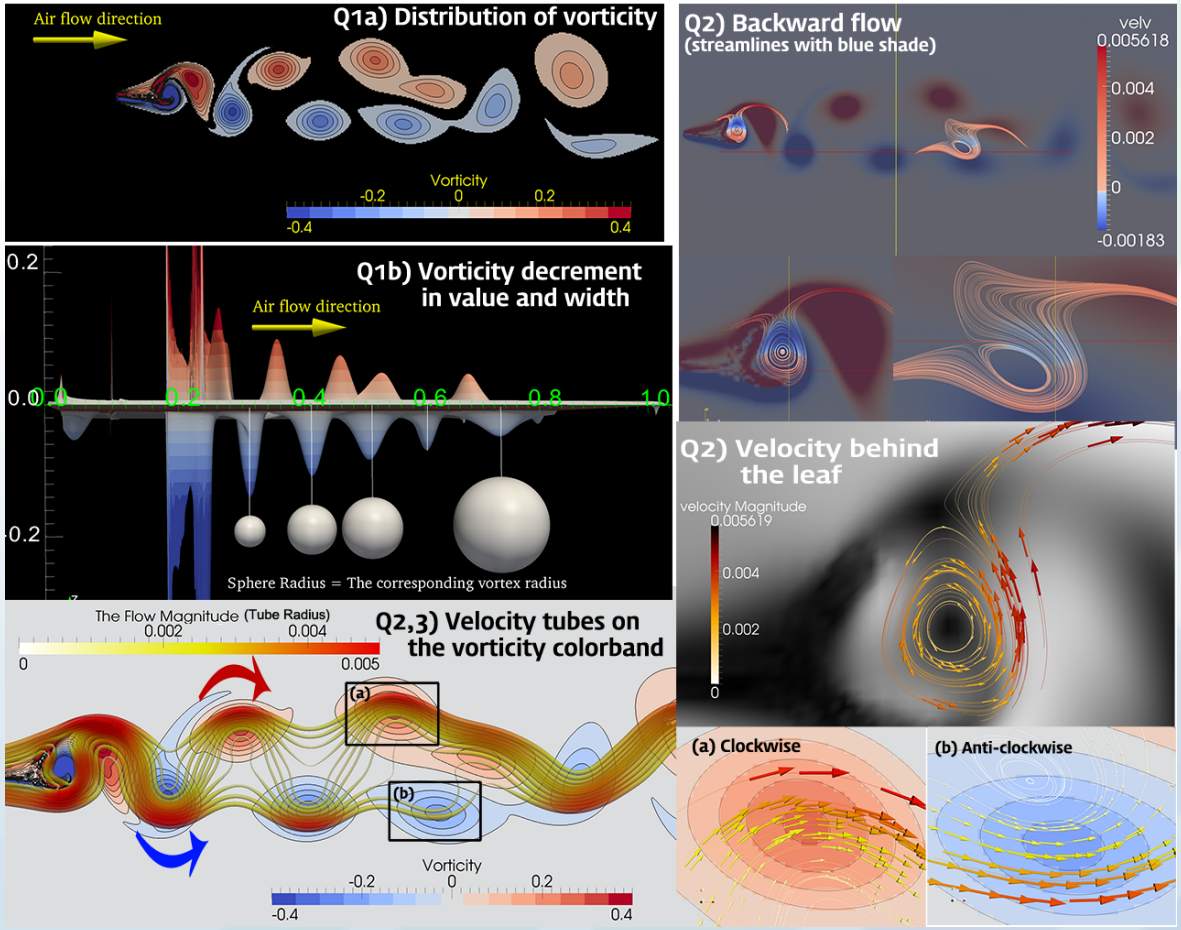


SCOREC RPI

Kitware

Ever More Complex Science

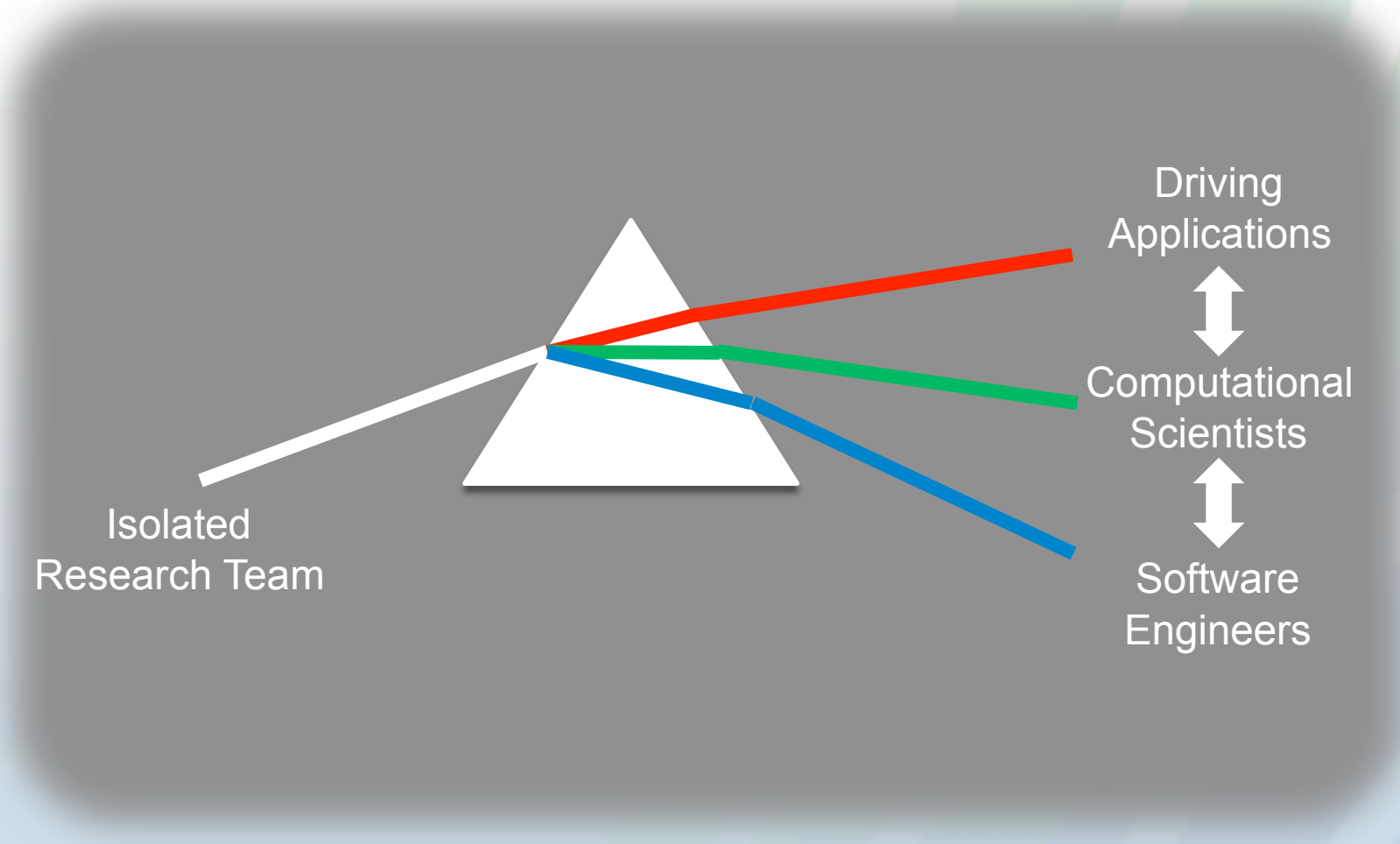
- Analysis
- Simulation
- Optimization



ParaView, Joo Hwi Lee and Namdi Brandon, UNC Visualization Class

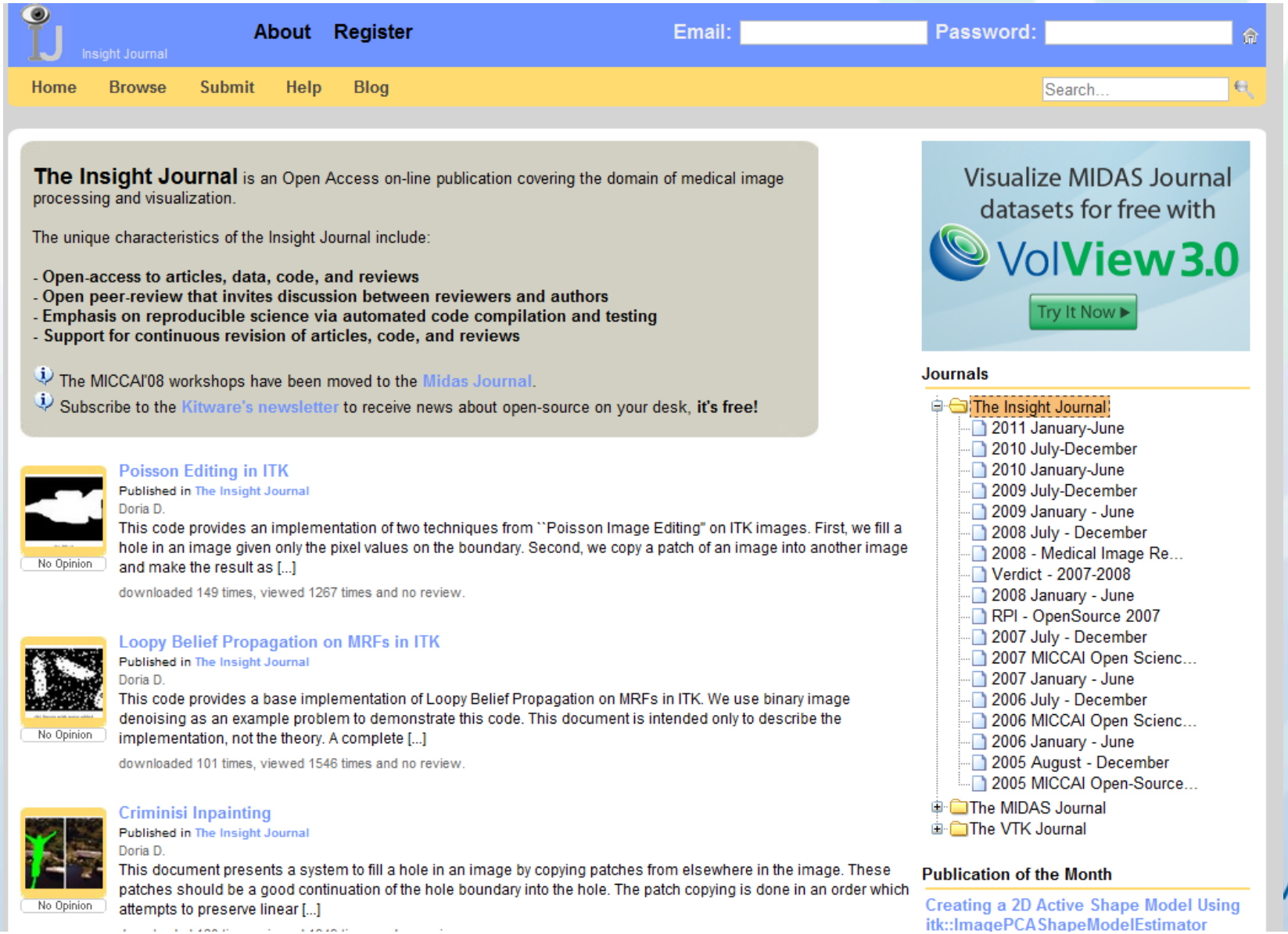
Kitware

Open Science is Scalable Science



Kitware

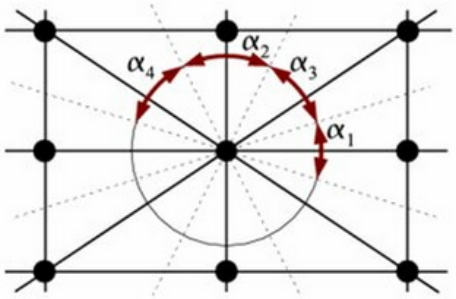
Open Access: The Insight Journal



Review Process Details

Efficient N-Dimensional surface estimation using Crofton formula and run-length encoding

Lehmann G., Legland D.
INRA



Please use this identifier to cite or link to this publication: http://hdl.handle.net/10380/3342

Published in The Insight Journal - 2012 January-June.

Submitted by Gaetan Lehmann on 02-21-2012.

Unlike the measure of the area in 2D or of the volume in 3D, the perimeter and the surface are not easily measurable in a discretized image.
In this article we describe a method based on the Crofton formula to measure those two parameters in a discritized image. The accuracy of the method is discussed and tested on several known objects. An algorithm based on the run-length encoding of binary objects is presented and compared to other approaches.
An implementation is provided and integrated in the LabelObject/LabelMap framework contributed earlier by the authors.

Code

There is no code review at this time.

Reviews

"U to the rescue, again!" by Nick Tustison on 2012-02-09 08:55:22 for revision #2

Resources
Download Package
Download Paper, View Paper
Download Source code
Author repository
Github

Statistics more
Global rating:
Review rating: [review]
Code rating:
Paper Quality:
2 comments

Information more
Categories: Data Representation, Filtering
Keywords: Crofton, RLE, Perimeter, Surface,
Toolkit: ITK
Export citation: Bibtex Export

Share
+1 Share Tweet

Associated Publications
Boolean Operations on Surfaces for VTK

View license
Send a message to the author

Another Notable System

- OSA's Interactive Science Publishing ISP
  - Active publications download data and viewer
  - *An ISP paper, top download for a period of five months for the journal of Applied Optics; ranked #1 in InfoBase (OSA's digital database)*

Oxygen advection and diffusion in a three-dimensional vascular anatomical network

Qianqian Fang,[1] Sava Sakadžić,[1] Lana Ruvinskaya,[1] Anna Devor,[1,2] Anders M. Dale,[3] and David A. Boas[1,*]

[1]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, 149 13th Street, Charlestown, Massachusetts 02129, USA
[2]Department of Neurosciences, University of California, San Diego, California, USA
[3]Department of Radiology, University of California, San Diego, California, USA
*Corresponding author: dboas@nmr.mgh.harvard.edu

Abstract: There is an increasing need for quantitative and computationally affordable models for analyzing tissue metabolism and hemodynamics in microvascular networks. In this work, we develop a hybrid model to solve for the time-varying oxygen advection-diffusion equation in the vessels and tissue. To obtain a three-dimensional temporal evolution of tissue oxygen concentration for realistic complex vessel networks, we used a graph-based advection model combined with a finite-element based diffusion model and an implicit time-advancing scheme. We validated this algorithm for both static and dynamic conditions. We also applied it to a complex vascular network obtained from a rodent somatosensory cortex. Qualitative agreement was found with *in-vivo* experiments.

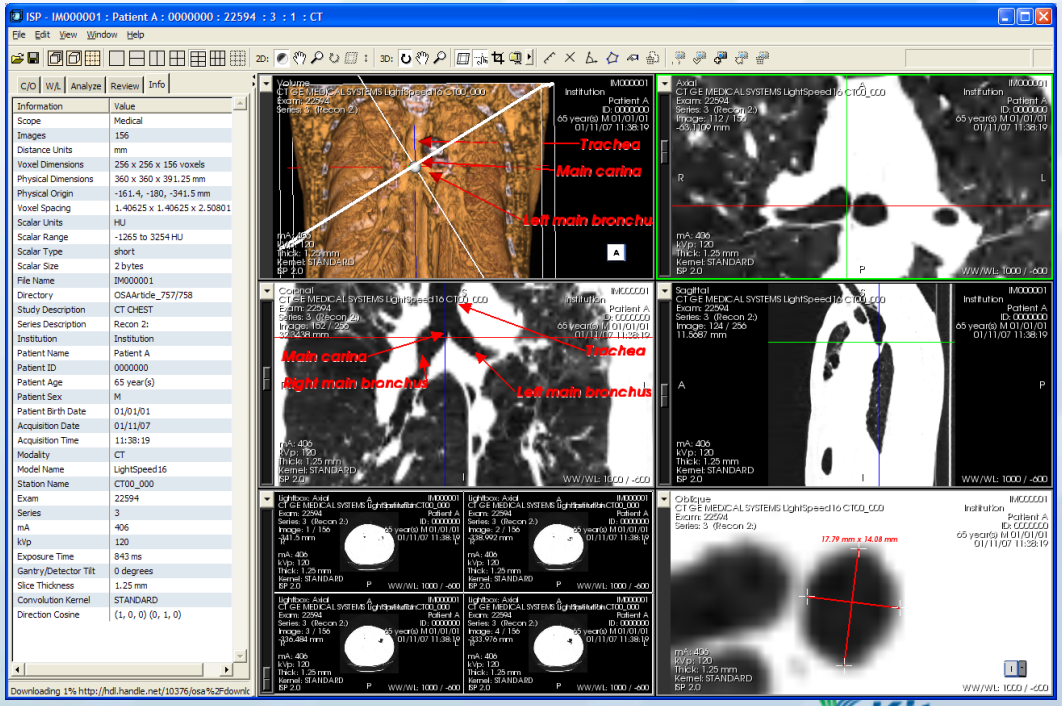
OCIS codes: (180.6900) Three-dimensional microscopy; (100.6890) Three-dimensional image processing.

[*]Datasets associated with this manuscript are available at http://hdl.handle.net/10376/1041.

References and links

1. M. Cope and D. T. Delpy, "System for long-term measurement of cerebral blood flow and tissue oxygenation on newborn infants by infra-red transillumination," Med. Biol. Eng. Comput. 26, 289-294 (1988).
2. A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," Trends Neurosci. 20, 435-442 (1997).
3. A. P. Gibson, J. C. Hebden, and S. R. Arridge, "Recent advances in diffuse optical imaging," Phys. Med. Biol. 50, R1-R43 (2005).
4. A. Grinvald, E. Lieke, R. D. Frostig, C. D. Gilbert, and T. N. Wiesel, "Functional architecture of cortex revealed by optical imaging of intrinsic signals," Nature 324, 361-364 (1986).
5. J. D. Briers, "Laser Doppler, speckle and related techniques for blood perfusion mapping and imaging," Physiol. Meas. 22, R35-R66 (2001).
6. A. K. Dunn, H. Bolay, M. A. Moskowitz, and D. A. Boas, "Dynamic imaging of cerebral blood flow using laser speckle," J. Cereb. Blood Flow Metab. 21, 195-201 (2001).
7. A. Villringer, A. Them, U. Lindauer, K. Einhaupl, and U. Dirnagl, "Capillary perfusion of the rat brain cortex," Circ. Res. 75, 55-62 (1994).
8. D. Kleinfeld, P. P. Mitra, F. Helmchen, and W. Denk, "Fluctuations and stimulus-induced changes in blood flow observed in individual capillaries in layers 2 through 4 of rat neocortex," Proc. Natl. Acad. Sci. USA 95, 15741-15746 (1998).
9. R. L. Haupt and D. D. Burns, "Quantitative, depth-resolved imaging of oxygen concentration by phosphorescence lifetime measurement," Appl. Spectrosc. 47, 1594-1599 (1993).
10. L. Sylvan, L. Qin, H. Fang, M. M. Zaman, E. Vitkin, I. C. Ghiran, S. Balachandar, M. Modell, C. Anderson, L. M. Kimmel, P. B. Cipelletti, K. H. Lim, S. D. Freedman, I. Bigio, B. P. Sachs, E. B. Hanlon, and L. T. Perelman, "Confocal light absorption and scattering spectroscopic microscopy monitors organelles in live cells with no exogenous labels," Proc. Natl. Acad. Sci. USA 104, 17255-17260 (2007).
11. A. D. Estrada, A. Ponticorvo, T. N. Ford, and A. K. Dunn, "Microvascular oxygen quantification using two-photon microscopy," Opt. Lett. 33, 1038-1040 (2008).
12. Y. Wang, B. A. Bower, J. A. Izatt, O. Tan, and D. Huang, "In vivo total retinal blood flow measurement by Fourier domain Doppler optical coherence tomography," J. Biomed. Opt. 12, 041215 (2007).

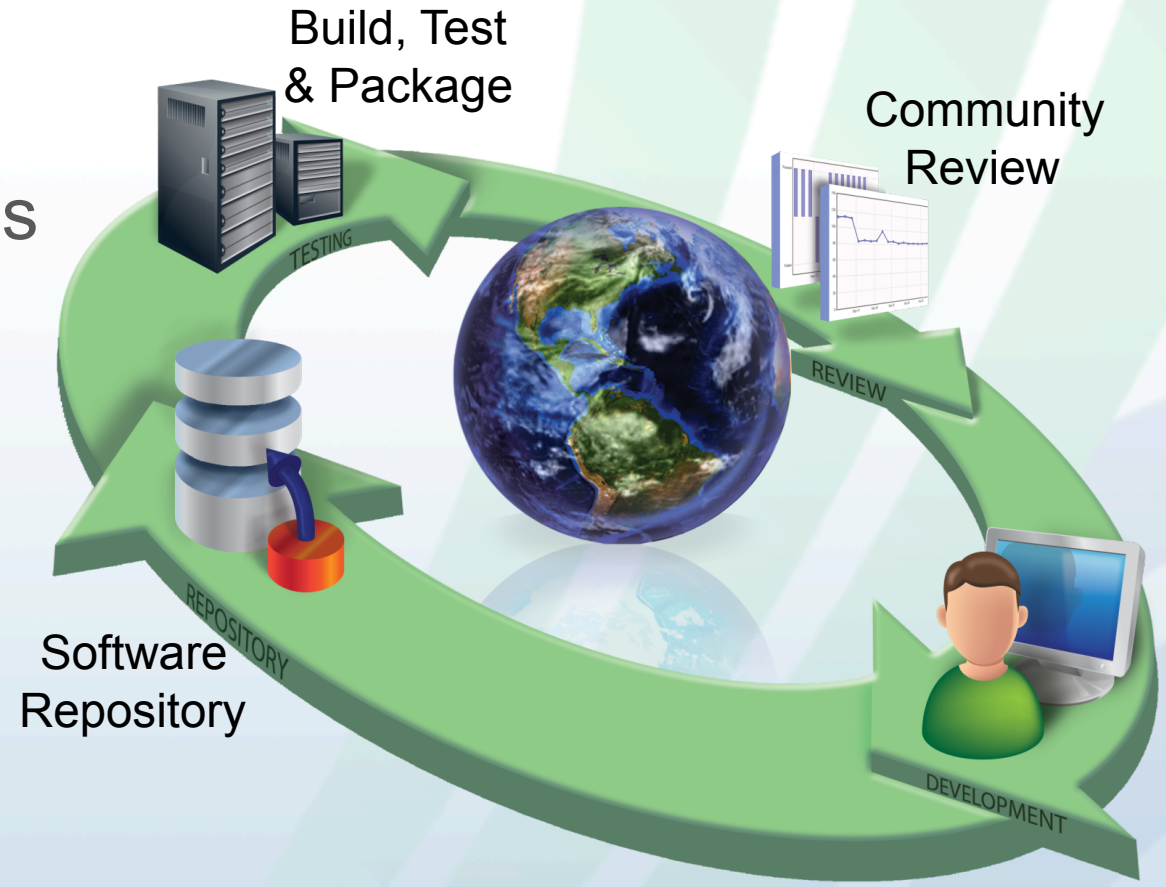
#98218 - $15.00 USD Received 1 Jul 2008; revised 3 Oct 2008; accepted 8 Oct 2008; published 15 Oct 2008



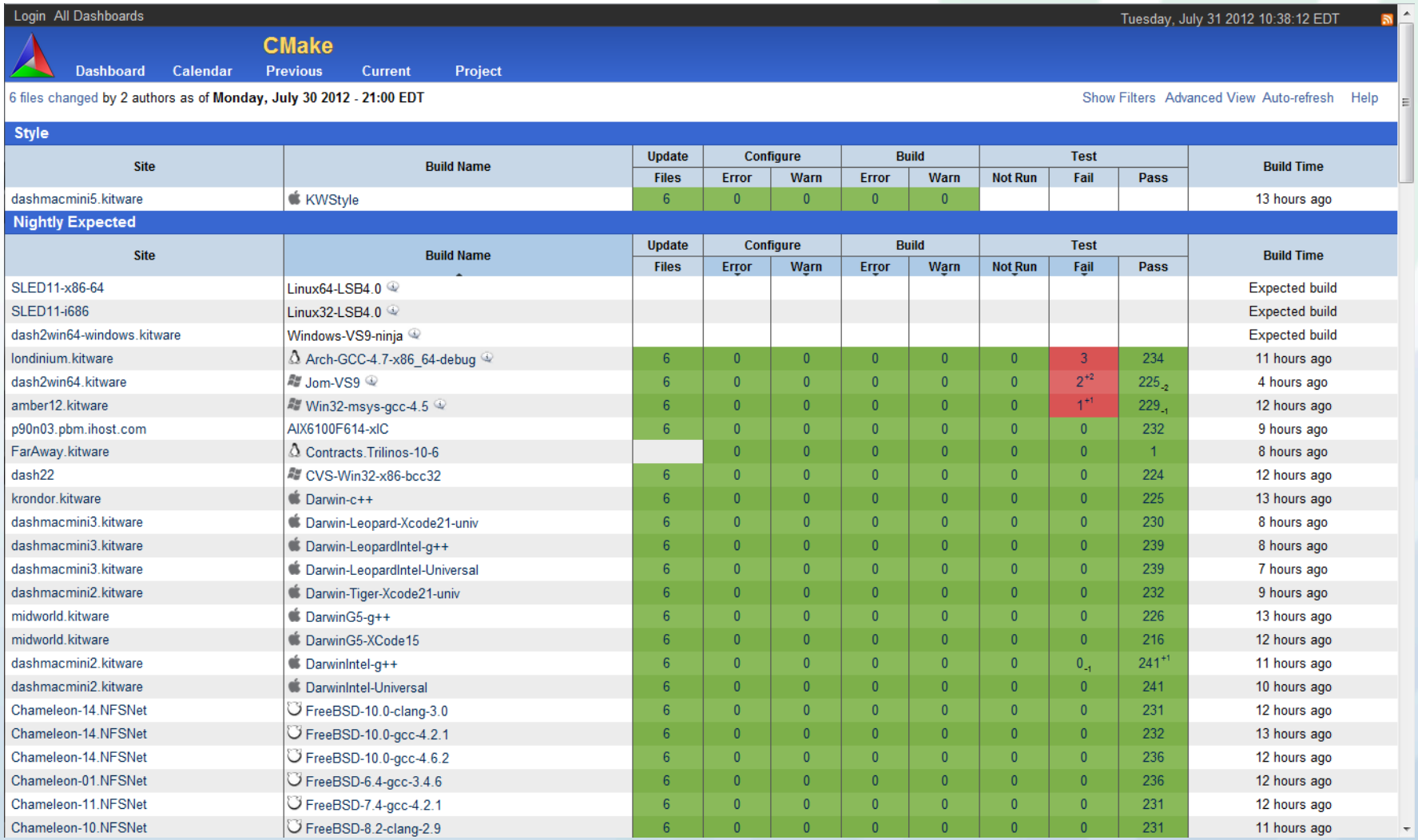
Kitware

Open Software Processes

- Require feedback loops
- Engage communities



Kitware



Login All Dashboards

Tuesday, July 31 2012 10:38:12 EDT

CMake

Dashboard Calendar Previous Current Project

6 files changed by 2 authors as of **Monday, July 30 2012 - 21:00 EDT**

Show Filters Advanced View Auto-refresh Help

Style

Site	Build Name	Update	Configure		Build		Test			Build Time
		Files	Error	Warn	Error	Warn	Not Run	Fail	Pass	
dashmacmini5.kitware	KWStyle	6	0	0	0	0				13 hours ago

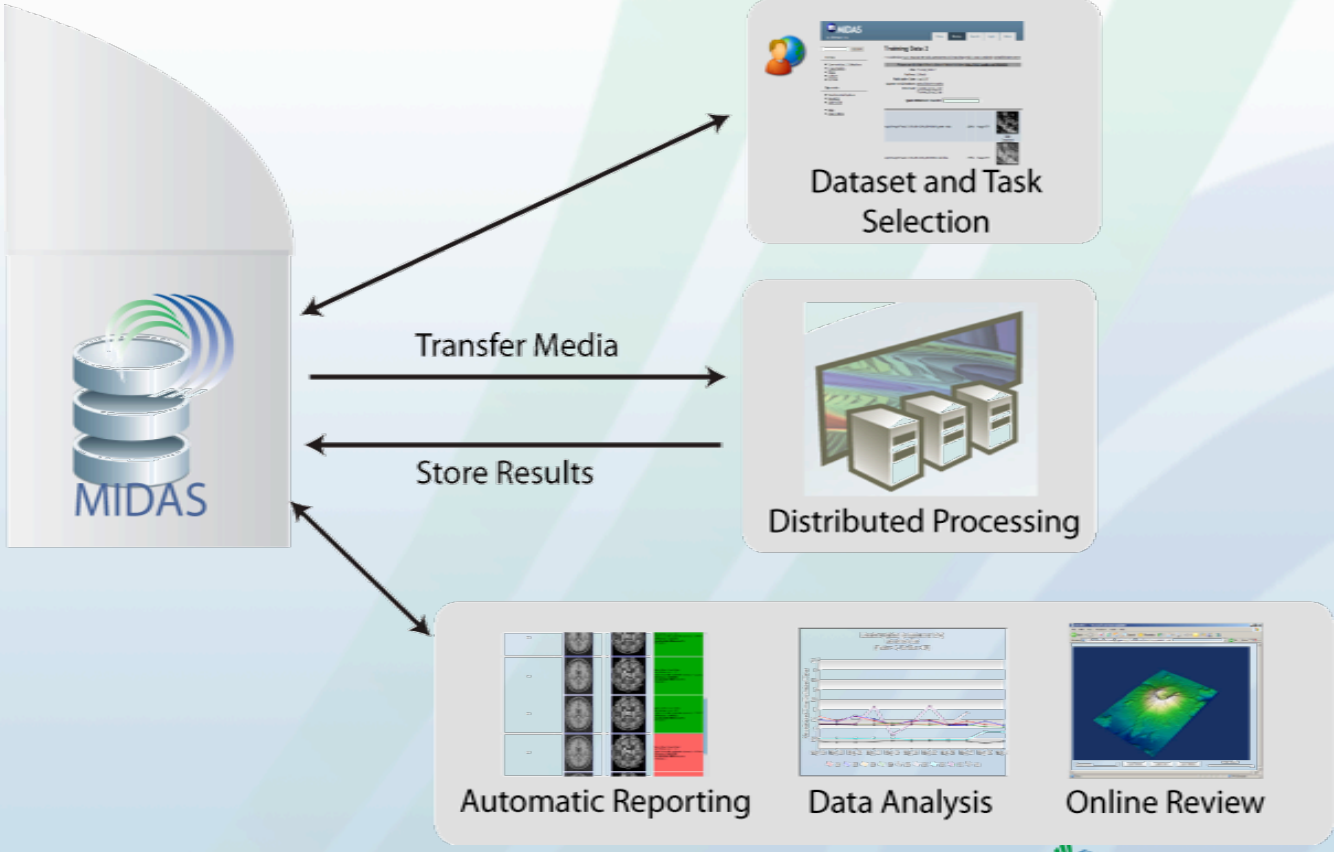
Nightly Expected

Site	Build Name	Update	Configure		Build		Test			Build Time
		Files	Error	Warn	Error	Warn	Not Run	Fail	Pass	
SLED11-x86-64	Linux64-LSB4.0									Expected build
SLED11-i686	Linux32-LSB4.0									Expected build
dash2win64-windows.kitware	Windows-VS9-ninja									Expected build
londinium.kitware	Arch-GCC-4.7-x86_64-debug	6	0	0	0	0	0	3	234	11 hours ago
dash2win64.kitware	Jom-VS9	6	0	0	0	0	0	2^{+2}	225_{-2}	4 hours ago
amber12.kitware	Win32-msys-gcc-4.5	6	0	0	0	0	0	1^{+1}	229_{-1}	12 hours ago
p90n03.pbm.ihost.com	AIX6100F614-xlC	6	0	0	0	0	0	0	232	9 hours ago
FarAway.kitware	Contracts.Trilinos-10-6		0	0	0	0	0	0	1	8 hours ago
dash22	CVS-Win32-x86-bcc32	6	0	0	0	0	0	0	224	12 hours ago
krondor.kitware	Darwin-c++	6	0	0	0	0	0	0	225	13 hours ago
dashmacmini3.kitware	Darwin-Leopard-Xcode21-univ	6	0	0	0	0	0	0	230	8 hours ago
dashmacmini3.kitware	Darwin-LeopardIntel-g++	6	0	0	0	0	0	0	239	8 hours ago
dashmacmini3.kitware	Darwin-LeopardIntel-Universal	6	0	0	0	0	0	0	239	7 hours ago
dashmacmini2.kitware	Darwin-Tiger-Xcode21-univ	6	0	0	0	0	0	0	232	9 hours ago
midworld.kitware	DarwinG5-g++	6	0	0	0	0	0	0	226	13 hours ago
midworld.kitware	DarwinG5-XCode15	6	0	0	0	0	0	0	216	12 hours ago
dashmacmini2.kitware	DarwinIntel-g++	6	0	0	0	0	0	0_{-1}	241^{+1}	11 hours ago
dashmacmini2.kitware	DarwinIntel-Universal	6	0	0	0	0	0	0	241	10 hours ago
Chameleon-14.NFSNet	FreeBSD-10.0-clang-3.0	6	0	0	0	0	0	0	231	12 hours ago
Chameleon-14.NFSNet	FreeBSD-10.0-gcc-4.2.1	6	0	0	0	0	0	0	232	13 hours ago
Chameleon-14.NFSNet	FreeBSD-10.0-gcc-4.6.2	6	0	0	0	0	0	0	236	12 hours ago
Chameleon-01.NFSNet	FreeBSD-6.4-gcc-3.4.6	6	0	0	0	0	0	0	236	12 hours ago
Chameleon-11.NFSNet	FreeBSD-7.4-gcc-4.2.1	6	0	0	0	0	0	0	231	12 hours ago
Chameleon-10.NFSNet	FreeBSD-8.2-clang-2.9	6	0	0	0	0	0	0	231	11 hours ago

Kitware

Open Data

- Large complex data must be available to support the scientific process
- Data-centric computing enables the community to access data



Kitware

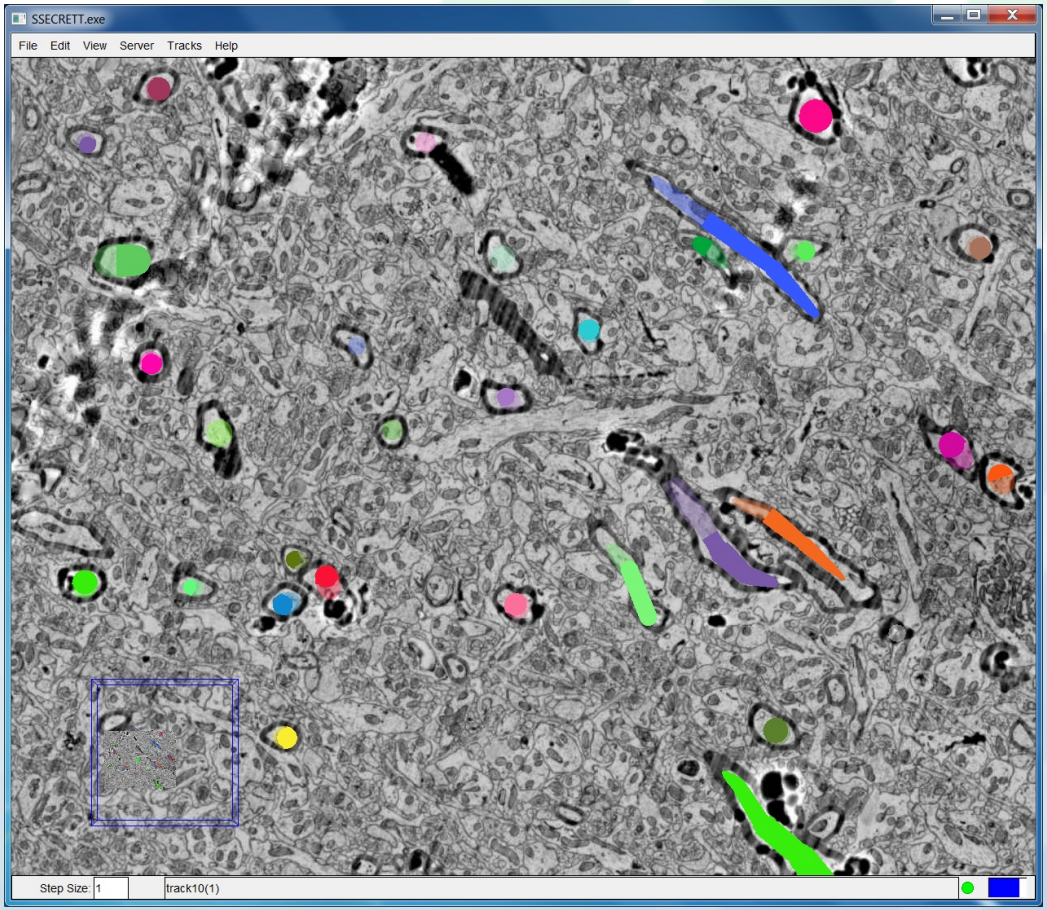
Challenges?

- Quality of newer Open Access journals?
  - Some like PLoS, BioMed Central may have even better review processes
- Status quo is threatened
- New business models required
- **Hosting large data and software, and providing access, is not cheap**

Kitware

Hosting Large Data

- Mouse connectome
- ~5 nanometer resolution (Electron microscopy)
- $100{,}000^2$ x 10,000



Kitware


Questions?
will.schroeder@kitware.com