The Virtues of Data Transparency

Michael A. Heroux
Scalable Algorithms Department
Sandia National Laboratories

Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin company, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia
National
Laboratories

Elements of the scientific method

1. Characterizations
2. Hypothesis development
3. Predictions from the hypothesis
4. Experiments
5. Evaluation and improvement
6. **(Independent) Confirmation**

http://en.wikipedia.org/wiki/Scientific_method

Reproducibility and Independent Verification

- Reproducibility and Independent Verification:
  - Part of rigorous implementation of the scientific method.
  - Specifically: Confirming or refuting results.
- Human tendency:
  - Rush through results collection (ahead of a paper deadline).
  - Gravitate toward results that support hypothesis.
- Evidence of reproducibility and independent verification is largely missing in many major journals.

SIAM SISC, 2010
Volume 32, Issue 3, Articles 1-10

- *Higher Order Multidimensional Upwind Convection Schemes for Flow in Porous Media on Structured and Unstructured Quadrilateral Grids,* Sadok Lamine and Michael G. Edwards
- Discrete Differential Forms for $(1+1)$-Dimensional Cosmological Space-Times, Ronny Richter and Jörg Frauendiener
- *A Phase-Field Model and Its Numerical Approximation for Two-Phase Incompressible Flows with Different Densities and Viscosities,* Jie Shen and Xiaofeng Yang
- *Analysis of Block Parareal Preconditioners for Parabolic Optimal Control Problems,* Tarek P. Mathew, Marcus Sarkis, and Christian E. Schaerer
- *Weighted Matrix Ordering and Parallel Banded Preconditioners for Iterative Linear System Solvers,* Murat Manguoglu, Mehmet Koyutürk, Ahmed H. Sameh, and Ananth Grama
- *Blendenpik: Supercharging LAPACK's Least-Squares Solver,* Haim Avron, Petar Maymounkov, and Sivan Toledo
- *Optimal Control of Parameter-Dependent Convection-Diffusion Problems around Rigid Bodies,* Timo Tonn, Karsten Urban, and Stefan Volkwein
- *An Entropy Adjoint Approach to Mesh Refinement,* Krzysztof J. Fidkowski and Philip L. Roe
- *High Order Numerical Methods to Three Dimensional Delta Function Integrals in Level Set Methods,* Xin Wen
- *Moving Least Squares via Orthogonal Polynomials,* Michael Carley

Reproducible Results?

- All 10 presented computational results.
- Only 3 mentioned the software environment used.
- Of those, 2 mentioned named software packages.
- Of those, 0 packages were available for independent use.
- 1 article briefly mentioned the hardware environment.

Note: Not picking on SIAM SISC. Many (most?) computational journals are like this.

Comparisons to Experimental Research

Study Questions Treatment Used in Heart Disease

By GARDINER HARRIS
Published: May 26, 2011

WASHINGTON — Lowering bad cholesterol levels reduces heart attack risks, and researchers have long hoped that raising good cholesterol would help, too. Surprising results from a large government study announced on Thursday suggest that this hope may be misplaced.

RECOMMEN
TWITTER
COMMENTS
E-MAIL
PRINT

- Computational Science often scorned, compared to experimental disciplines:
  - Assumptions: Experiments are independently verified.
  - Reality: They seem to increasingly suffer from the same problem.
- Examples:
  - Query: Is coffee [bad|good] for you?
  - NY Times: HDL therapy *increases* heart-attack risk.
- Tim Trucano (Sandia V&V expert):
  - As computing use increases in experimental disciplines, ability and tendency to verify results goes down.
- Is parameter space too large to expect reproducibility and independent verification?

Sandia National Laboratories

Reproducibility & Independent Verification Requirement

- In order to publish a paper: *Someone other than the authors must be able to reproduce the computational results.*
- Latitude in “reproduce”:
  - Exactly the same numerical results?
  - Exactly the same runtime?
  - Close, in the opinion of an expert reviewer?
- What about:
  - Access to the same computing environment?
  - High end systems?
- Lots of challenges.
- But just the *expectation [threat]* can drive efforts…

Fruits of the Threat

- **Visibility of data:** External review of results implies some level of data transparency.
- **Source/data management tools:** In order to guarantee that results can be reproduced, software & data must be preserved so that the exact version used to produce results is available at a later date.
- **Use of other standard tools and platforms:** In order to reduce the complexity of an environment, standard software libraries and computing environments will be helpful.
- **Documentation:** Independent verification requires that someone else understand how to use your software & data.
- **Source & data standards:** Improves the ability of others to read your source & data.
- **Testing:** Investment in greater testing makes sense because the software & data will be used by others.
- **High-quality software engineering environment:** If a research team is serious about producing high-quality, reproducible and verifiable results, it will want to invest in a high-quality SE environment to improve team efficiency.

Challenges

- Computing Environment Complexity:
  - Combinatorial explosion of parameters.
  - How to manage?
- Cost:
  - Reduced frequency of publication.
  - Increased burden on reviewers.
- Cultural:
  - Biggest?
  - Long transition.
    - New tools, processes.
  - Sense of vulnerability.

Sandia National Laboratories

ACM Transactions on
Mathematical Software

ACM TOMS

- TOMS RCR Initiative: Referee Data.
- Why TOMS? Tradition of real software that others use.
- Two categories: Algorithms, Research.
- TOMS Algorithms Category:
  - Software Submitted with manuscript.
  - Both are thoroughly reviewe.
- TOMS Research Category:
  - Stronger: Previous implicit “real software” requirement is explicit.
  - New: Special designation for replicated results.

Sandia National Laboratories

ACM TOMS Reproducible Computational Results (RCR) Process

- Submission: Optional (for now) RCR option.
- Standard reviewer assignment: Nothing changes.
- RCR reviewer assignment:
  - Concurrent with the first round of standard reviews
  - Known to and works with the authors during the RCR process.
- RCR process:
  - Multi-faceted approach.
- Publication:
  - Replicated Computational Results Designation.
  - The RCR referee acknowledged.
  - Review report appears with published manuscript.

RCR Process

- Independent replication:
  - Transfer of or pointer to software given to RCR reviewer.
  - Guest account, access to software on author's system.
  - Detailed observation of the authors replicating the results.
- Review of computational results artifacts:
  - Results may be from a system that is no longer available
  - Leadership class computing system
  - In this situation:
    - Careful documentation of the process.
    - Software should have its own substantial verification process.

Sandia National Laboratories

Conclusions

- Requiring/fostering data transparency, reproducibility & independent verification is virtuous.
  - Healthy dynamics: More rigor in generating, presenting, preserving.
  - Can lead to higher quality practices, tools, concepts and more.
- Better practices not *imposed* but deemed necessary by the research team, a means to an end.
- Cost is substantial.
- Rest of research community: Backsliding?
- Need concerted focus on data transparency:
  - Tools to make it easier (VMs, common SW environments).
  - Funding agencies to expect it.
  - Journal policies to ask/demand it.
- Best impact: Just the *expectation* of the requirement leads to better scientific practices.
- Data ownership? Transparency is better.

Sandia National Laboratories

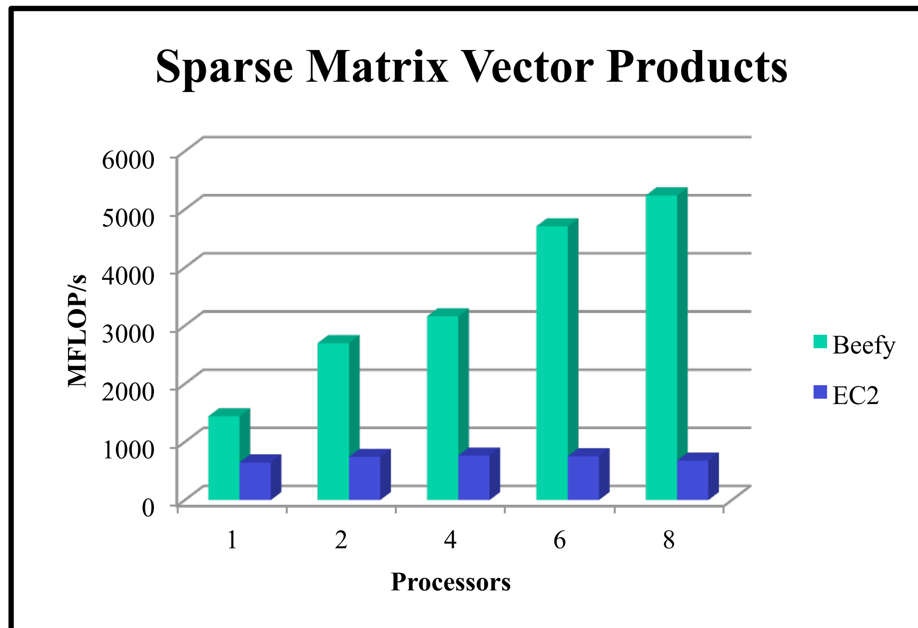
Possible approaches for making progress

Engineering of Research Software

- Research software history:
  - Informal: Very few formal approaches.
  - Basics: Repository, documentation standards, etc.
    - Missing or *ad hoc.*
  - More advanced: lifecycle model, release process.
    - Unaware of concepts.
- How can we improve?
  - Impose processes? No.
  - Trust we will all just "do a good job?" No.
  - Require independent verification of results? Maybe.

Possibilities: Portable Environments

- Pre-built SW environment
- Example:
  - Linux+GCC+MPI+valgrind+…
  - Trilinos
  - All Trilinos 3rd-party libs:
    - SuperLU, UMFPack, ParMetis, Boost…
- Issue: Performance
  - Trilinos EC2 appliance.
  - Sparse kernel benchmark.
  - Native HW (beefy) vs.
  - EC2 VM (8 cores).
- But: VirtualBox gives 80% of (single core) performance. (Levesque)



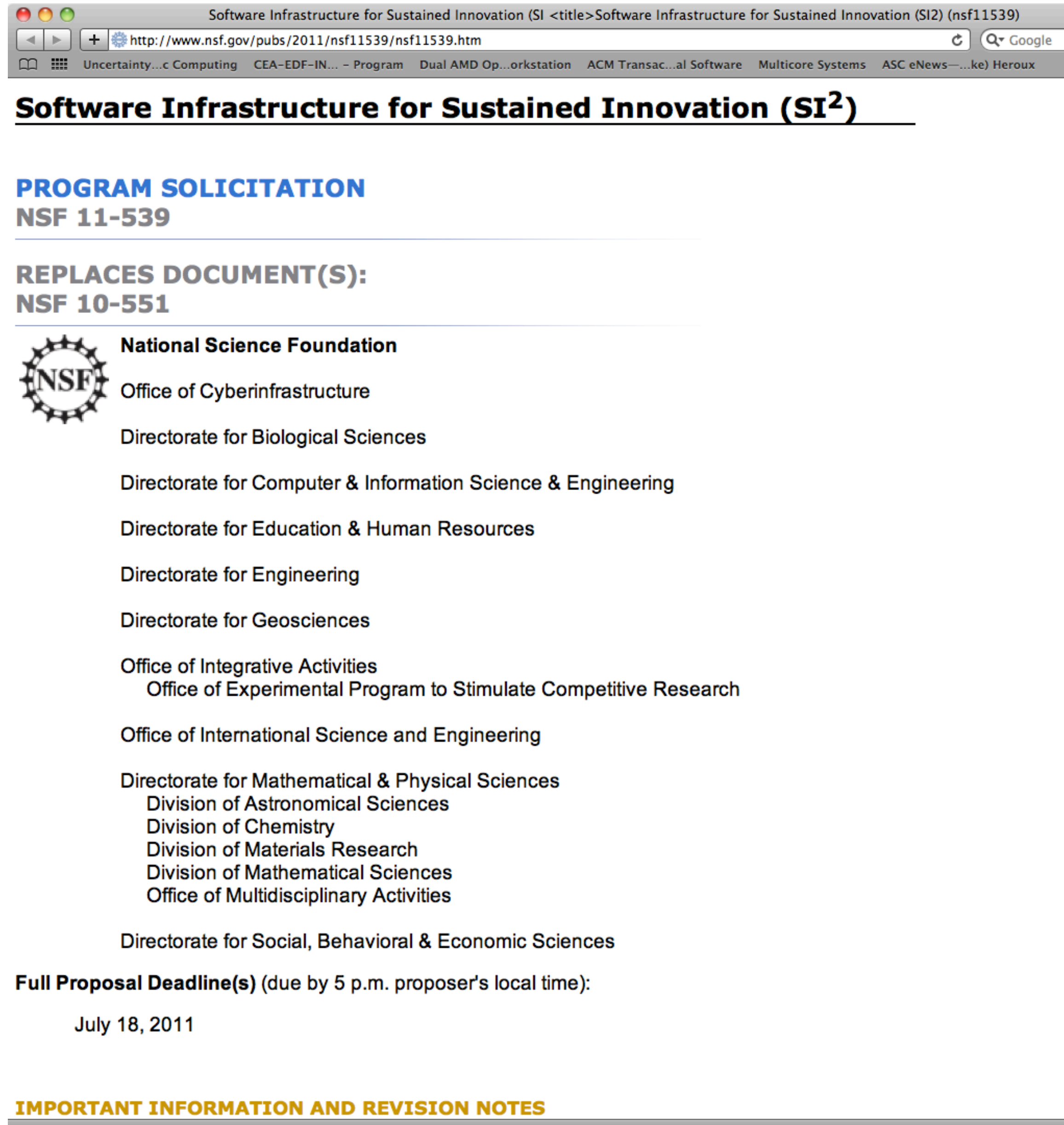
Use Common Programming Environments

PETSc

- Libraries & PEs provide:
  - Common functionality base.
  - Portability.
- Reduced cost:
  - Code size.
  - Software infrastructure.
  - Reviewer orientation.

Funding Agencies

- US DOE:
  - Long tradition producing software.
  - Challenge maintaining.
  - Maintenance *ad hoc.*
- US NSF:
  - Getting started.



Software Infrastructure for Sustained Innovation (SI^2)

PROGRAM SOLICITATION
NSF 11-539

REPLACES DOCUMENT(S):
NSF 10-551

National Science Foundation

Office of Cyberinfrastructure

Directorate for Biological Sciences

Directorate for Computer & Information Science & Engineering

Directorate for Education & Human Resources

Directorate for Engineering

Directorate for Geosciences

Office of Integrative Activities
Office of Experimental Program to Stimulate Competitive Research

Office of International Science and Engineering

Directorate for Mathematical & Physical Sciences
Division of Astronomical Sciences
Division of Chemistry
Division of Materials Research
Division of Mathematical Sciences
Office of Multidisciplinary Activities

Directorate for Social, Behavioral & Economic Sciences

Full Proposal Deadline(s) (due by 5 p.m. proposer's local time):

July 18, 2011

IMPORTANT INFORMATION AND REVISION NOTES