



Reproducibility & Replication: Tools, Policies and Processes

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



Outline

- Containers
- ACM Replicated Computational Results
- DOE SW Productivity and Sustainability Plan



Trilinos Docker Project



Typical Trilinos Cmake Script (edison): Contain in Docker

```
cmake \
-D MPI_CXX_COMPILER="CC" \
-D MPI_C_COMPILER="cc" \
-D MPI_Fortran_COMPILER="ftn" \
-D Teuchos_ENABLE_STACKTRACE:BOOL=OFF \
-D Teuchos_ENABLE_LONG_LONG_INT:BOOL=ON \
-D Trilinos_ENABLE_Tpetra:BOOL=ON \
-D Tpetra_ENABLE_TESTS:BOOL=ON \
-D Tpetra_ENABLE_EXAMPLES:BOOL=ON \
-D Tpetra_ENABLE_EXPLICIT_INSTANTIATION:BOOL=ON \
-D Teuchos_ENABLE_EXPLICIT_INSTANTIATION:BOOL=ON \
-D TPL_ENABLE_MPI:BOOL=ON \
-D CMAKE_INSTALL_PREFIX:PATH="$HOME/opt/Trilinos/tpetraEval" \
-D BLAS_LIBRARY_DIRS="$LIBSCI_BASE_DIR/gnu/lib" \
-D BLAS_LIBRARY_NAMES="sci" \
-D LAPACK_LIBRARY_DIRS="$LIBSCI_BASE_DIR/gnu/lib" \
-D LAPACK_LIBRARY_NAMES="sci" \
-D CMAKE_CXX_FLAGS="-O3 -ffast-math -funroll-loops" \
\
..
```



- Containerization technology
- Think: Virtual machine
 - But light-weight, native.
 - Composable, layered.
- Driven by the data science community.



Trilinos usage via Docker

- WebTrilinos Tutorial
 - <https://hub.docker.com/r/sjdeal/webtrilinos>
- <http://johntfoster.github.io/posts/peridigm-without-building-via-Docker.html>
 - `docker pull johntfoster/trilinos`
 - `docker pull johntfoster/peridigm`
 - `docker run --name peridigm0 -d -v `pwd`:./output johntfoster/peridigm \`
 `Peridigm fragmenting_cylinder.peridigm`
 - Etc...



WebTrilinos

webtrilinos matrixportal

C++ Code Page ?

Insert template or

Text area with rows and columns.

Run with process(es) and thread(s).

Please type your C++ code below.



This web site is hosted by St. John's University, MN.
Credits: M. Sala, M. Phenow, J. Hu, R. Tuminaro.
Last updated on June 30, 2015 - 9:53 am CDT.



webtrilinos matrixporta

C++ Code Page ?

Insert template or

Text area with rows and columns.

Run with process(es) and thread(s).

Please type your C++ code below.

```
#include "Teuchos_ParameterList.hpp"
#include "AztecOO.h"

int main(int argc, char *argv[])
{
#ifdef HAVE_MPI
    MPI_Init(&argc,&argv);
    Epetra_MpiComm Comm( MPI_COMM_WORLD );
#else
    Epetra_SerialComm Comm;
#endif

    Teuchos::ParameterList Galerilist;

    // The problem is defined on a 2D grid, global size is nx * nx.
    int nx = 30;
    Galerilist.set("n", nx * nx);
    Galerilist.set("nx", nx);
    Galerilist.set("ny", nx);
    Epetra_Map* Map = Galerilist.CreateMap("Linear", Comm, Galerilist);
    Epetra_RowMatrix* A = Galerilist.CreateCrsMatrix("Laplace2D", Map, Galerilist);
```



This web site is hosted by St. John's University, MN.
Credits: M. Sala, M. Phenow, J. Hu, R. Tuminaro.
Last updated on June 30, 2015 - 9:53 am CDT.

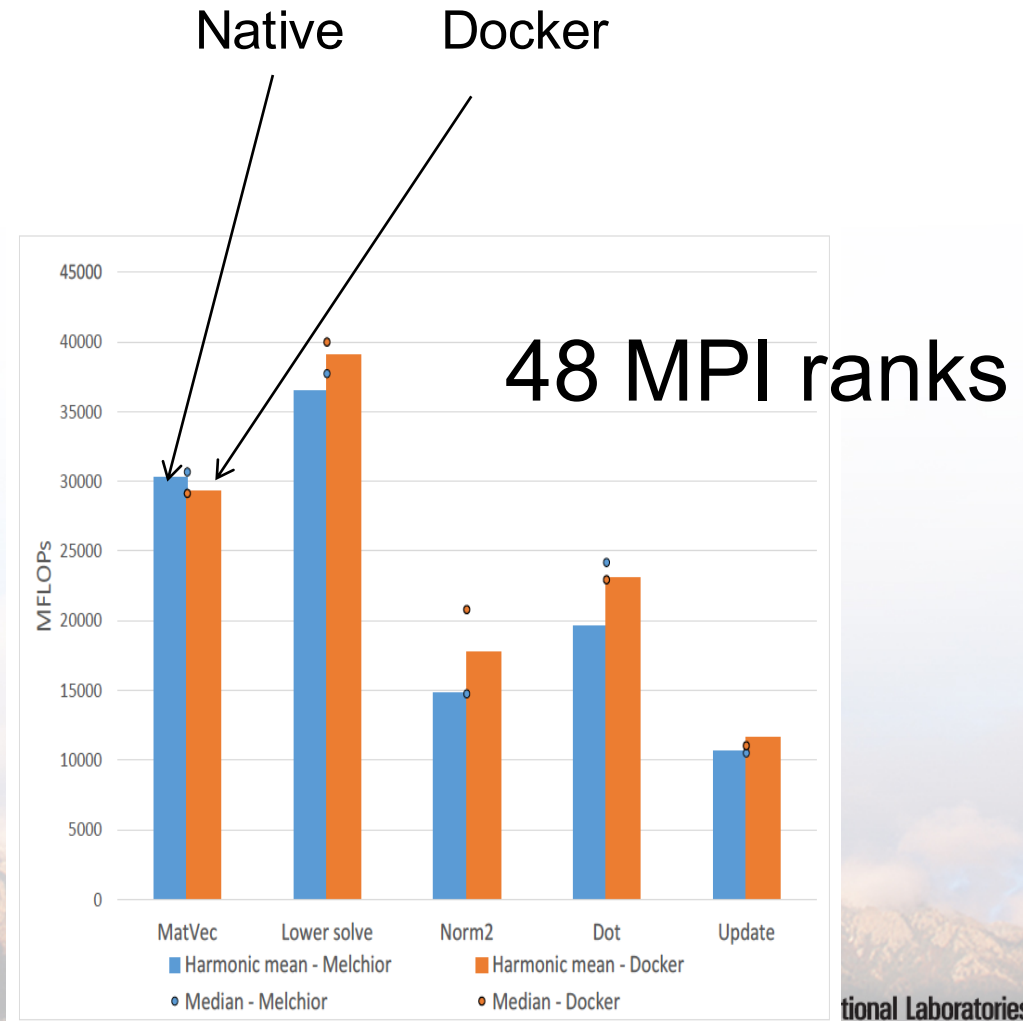


Sandia National Laboratories

First Docker MPI Results (Sean Deal)

SJU Cluster, Epetra Basic Perf Test

- MatVec
- Lower Solve
- Norm2, Dot, Update
- Harmonic mean of 5 tests
- 4M Eq per proc





Future Docker Efforts

- Standard Developer Environment:
 - Enables reproducibility (error states).
 - Contains all third-party libraries:
 - Trilinos can use dozens of TPLs (SuperLU, MUMPS, ParMetis, etc., MKL? Licensing?)
 - Contains one or a few compiler versions:
 - Enables uniform error detection/correction.
 - No ambiguity about built/test environment.
- Productivity improvement:
 - Libraries pre-built. (Full build can take hours).
- Specialized containers:
 - Turnkey capabilities.
 - Example: Scalable SVD using Anasazi.
 - Target device builds (GPU).

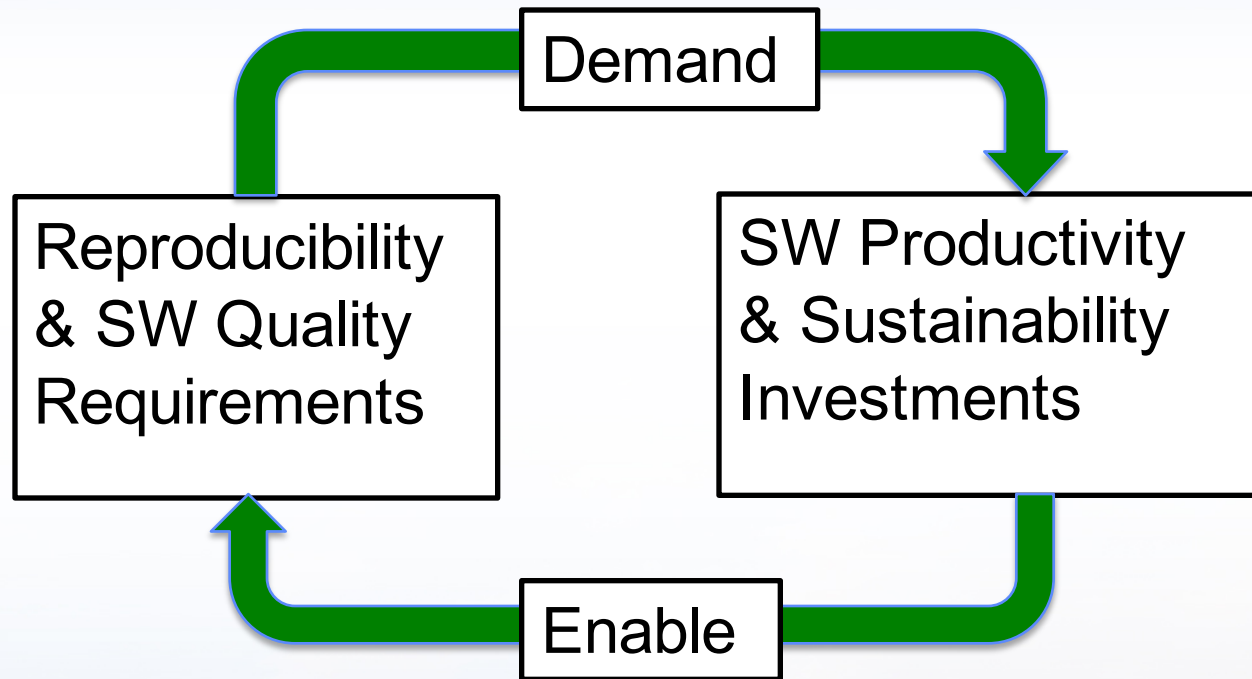




(Computational) Scientific Productivity



Reproducibility/Quality & Productivity/Sustainability



Cover letter
excerpt from RCR
candidate paper

Thank you for taking the time to consider our paper for your journal.

XXX has agreed to undergo the RCR process should the paper proceed far enough in the review process to qualify. ***To make this easier we have preserved the exact copy of the code used for the results (including additional code for generating detailed statistics that is not in the library version of the code).***

Reproducibility Terminology

V. Stodden, D. H. Bailey, J. Borwein, R. J. LeVeque, W. Rider, and W. Stein. 2013. Setting the Default to Reproducible: Reproducibility in Computational and Experimental Mathematics. (2013).
http://icerm.brown.edu/html/programs/topical/tw12_5_rcem/icerm_report.pdf

- **Reviewable Research.** The descriptions of the research methods can be independently assessed and the results judged credible. (This includes both traditional peer review and community review, and does not necessarily imply reproducibility.)
- **Replicable Research.** Tools are made available that would allow one to duplicate the results of the research, for example by running the authors' code to produce the plots shown in the publication. (Here tools might be limited in scope, e.g., only essential data or executables, and might only be made available to referees or only upon request.)
- **Confirmable Research.** The main conclusions of the research can be attained independently without the use of software provided by the author. (But using the complete description of algorithms and methodology provided in the publication and any supplementary materials.)
- **Auditable Research.** Sufficient records (including data and software) have been archived so that the research can be defended later if necessary or differences between independent confirmations resolved. The archive might be private, as with traditional laboratory notebooks.
- **Open or Reproducible Research.** Auditable research made openly available. This comprised well-documented and fully open code and data that are publicly available that would allow one to (a) fully audit the computational procedure, (b) replicate and also independently reproduce the results of the research, and (c) extend the results or apply the method to new problems.

- 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 reviewed.
- TOMS Research Category:
 - Stronger: Previous implicit “real software” requirement is explicit.
 - New: Special designation for replicated results.

ACM TOMS Replicated Computational Results (RCR)

- Submission: Optional RCR option.
- Standard reviewer assignment: Nothing changes.
- RCR reviewer assignment:
 - Concurrent with standard reviews.
 - As early as possible in review process.
 - Known to and works with authors during the RCR process.
- RCR process:
 - Multi-faceted approach, Bottom line: Trust the reviewer.
- Publication:
 - Replicated Computational Results Designation.
 - The RCR referee acknowledged.
 - Review report appears with published manuscript.





RCR Process: Two Basic Approaches

1. Independent replication (3 options):

- A. Transfer of, or pointer to, author's software.
- B. Guest account, access to author's software.
- C. Observation of authors replicating results.

Or (Untested, rare)

2. Review of computational results artifacts:

- Results may be from an unavailable system.
- Leadership class computing system.
- In this situation:
 - Careful documentation of the process.
 - Software should have its own substantial **V&V** process.

TOMS:

- First RCR paper in TOMS issue 41:3
 - Editorial introduction.
 - van Zee & van de Geijn, BLIS paper.
 - Referee report.
- Second: TOMS 42:1
 - Hogg & Scott.
- Third: TOMS 42:4.
- Fourth ID'ed.

TOMACS

- Initial article complete.
- 4 in the system.





Big Picture of TOMS RCR

- Improve science.
 - Quality of prose in publications: Good.
 - Quality of data: (Very!) poor.
- So bad now:
 - Trust comes from seeing a “cloud” of similar papers with similar results.
 - Which could still be wrong (built on a common bad piece).
 - Replicability: First step toward improvement.
- Engage a “dark portion” of the R&D community.
 - Reviewers not among typical reviewer pool.
 - Practitioners, users. Expert at use of Math SW.

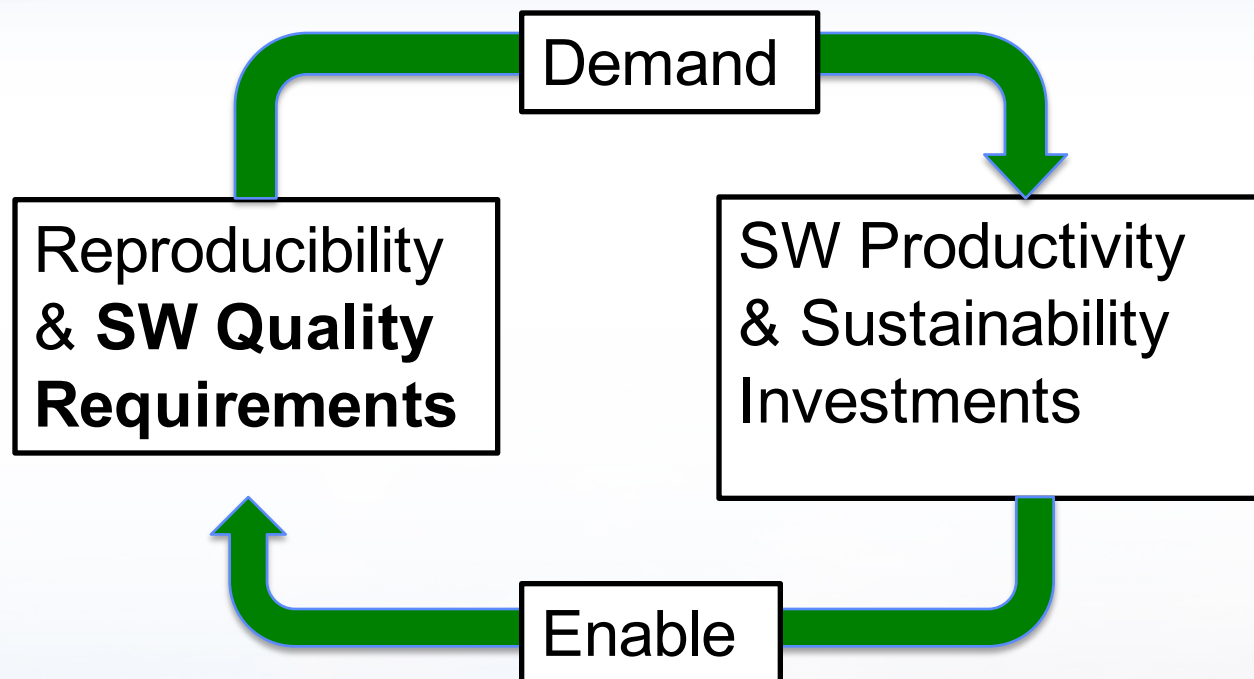


RCR Status & Futures

- TOMACS: Adopted TOMS RCR.
- ACM: Setting up electronic workflow support.
- Author participation:
 - TOMACS: Optional, indefinitely.
 - TOMS: Optional, mandatory in the future.
- Start of RCR review:
 - TOMACS: Only after paper is likely to be accepted.
 - TOMS: As soon as possible (to reduce risk at end).
- RCR Reviewer sources:
 - TOMACS: Interested scientists; if students then under supervision.
 - TOMS: Professional practitioners.



Reproducibility/Quality & Productivity/Sustainability



New funding proposal element:
SW Productivity and Sustainability Plan

Your question:

What's a SW Productivity and Sustainability Plan?



DOE SW Productivity and Sustainability Plan (SW PSP).

No viable SW plan, no money (eventually)

- **Key Entities:**

- DOE Biological and Environmental Research (BER).
- DOE Advanced Scientific Computing Research (ASCR)
- IDEAS Scientific SW Productivity Project

- **Milestone:**

- First-of-a-kind SW Productivity and Sustainability Plan.



DOE BER SW PSP Requirements (I)

- Describe overall SW development process.
 - Software lifecycle, testing, documentation and training.
- Development tools and processes:
 - source management, issue tracking, regression testing, SW distribution.
- Training and transition:
 - New and departing team members.
- Continuous process improvement:
 - Getting better at productivity and sustainability.



Creating Incentives: Carrots & Sticks

- Opportunity:
 - Publishers & Funding Agencies can provide increased incentive for better behavior.
- Challenge:
 - Cultural change.
 - Increased cost: Fewer results, more cost (but better quality).
- Goal:
 - Increase incentives to improve reproducibility.
 - Improve SW quality, making reproducibility easier.
- Example:
 - TOMS Replicated Computational Results effort.
 - DOE SW Productivity and Sustainability Plan.
- **Bottom Line:**
 - **You get published *if your results are reproducible.***
 - **You get money *if you have a solid software plan.***

Productivity++ Initiative

Ask: *Is My Work* _____ ?

Productivity++

- ✓ Traceable
- ✓ In Progress
- ✓ Sustainable
- ✓ Improved



Version 1.2

<https://github.com/trilinos/Trilinos/wiki/Productivity---Initiative>



Sandia National Laboratories