Overview of the IDEAS Software Productivity Project

Mike Heroux (SNL), Lois Curfman McInnes (ANL), J. David Moulton (LANL), David Bernholdt (ORNL), Hans Johansen (LBNL)

And all IDEAS project members

DOE High Performance Computing Operational Review (HPCOR) on Scientific Software Architecture for Portability and Performance
Washington, DC, September 15, 2015

Confluence of trends

- **Fundamental trends:**
  - Disruptive HW changes: Requires thorough algorithm/code refactoring
  - Demands for coupling: Multiphysics, multiscale
- **Challenges:**
  - Need refactorings: Really, continuous change
  - Modest app development funding: No monolithic apps
  - Requirements are unfolding, evolving, not fully known *a priori*
- **Opportunities:**
  - Better design and SW practices & tools are available
  - Better SW architectures: Toolkits, libraries, frameworks
- **Basic strategy: Focus on <u>productivity</u>**

Interoperable Design of Extreme-scale Application Software (IDEAS)

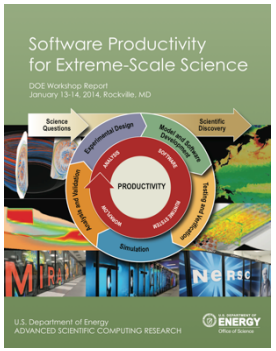
Motivation

Enable ***increased scientific productivity,*** realizing the potential of extreme-scale computing, through ***a new interdisciplinary and agile approach to the scientific software ecosystem.***

Objectives

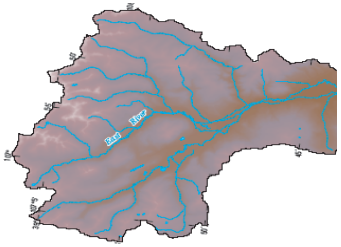
Address confluence of trends in hardware and increasing demands for predictive multiscale, multiphysics simulations.

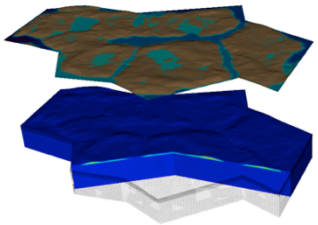
Respond to trend of continuous refactoring with efficient agile software engineering methodologies and improved software design.

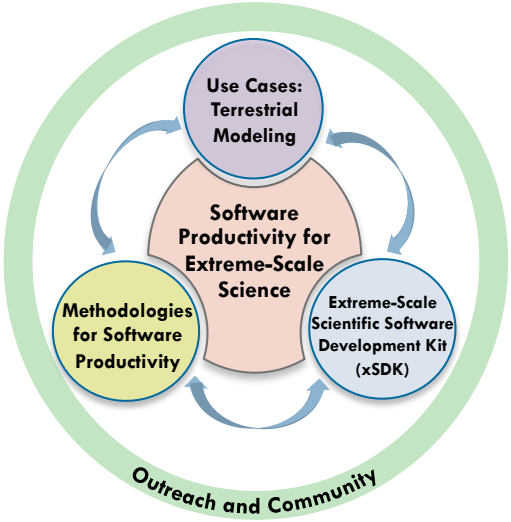


Impact on Applications & Programs

Terrestrial ecosystem ***use cases tie IDEAS to modeling and simulation goals*** in two Science Focus Area (SFA) programs and both Next Generation Ecosystem Experiment (NGEE) programs in DOE Biologic and Environmental Research (BER).







Approach

ASCR/BER partnership ensures delivery of both crosscutting methodologies and metrics with impact on real application and programs.

Interdisciplinary multi-lab team (ANL, LANL, LBNL, LLNL, ORNL, PNNL, SNL)

ASCR Co-Leads: Mike Heroux (SNL) and Lois Curfman McInnes (ANL)

BER Lead: David Moulton (LANL)

Topic Leads: David Bernholdt (ORNL) and Hans Johansen (LBNL)

Integration and synergistic advances in three communities deliver scientific productivity; outreach establishes a new holistic perspective for the broader scientific community.

www.ideas-productivity.org

U.S. DEPARTMENT OF ENERGY
Office of Science

IDEAS project structure and interactions

DOE Program Managers

ASCR: Thomas Ndousse-Fetter
BER: Paul Bayer, David Lesmes

IDEAS: Interoperable Design of Extreme-scale Application Software

ASCR Co-Leads: Mike Heroux (SNL) and Lois Curfman McInnes (ANL)
BER Lead: J. David Moulton (LANL)

Executive Advisory Board
John Cary (Tech-X)
Mike Glass (SNL)
Susan Hubbard (LBNL)
Doug Kothe (ORNL)
Sandy Landsberg (DOD)
Paul Messina (ANL)

BER Use Cases
Lead: J. David Moulton (LANL)
Carl Steefel (LBNL) *1
Scott Painter (ORNL) *2
Reed Maxwell (CSM) *3
Glenn Hammond (SNL)
Tim Scheibe (PNNL)
Laura Condon (CSM)
Ethan Coon (LANL)
Dipankar Dwivedi (LBNL)
Jeff Johnson (LBNL)
Eugene Kikinzon (LANL)
Sergi Molins (LBNL)
Steve Smith (LLNL)
Carol Woodward (LLNL)
Xiaofan Yang (PNNL)

Methodologies for Software Productivity
Lead: Mike Heroux (SNL)
Roscoe Bartlett (ORNL)
Todd Gamblin* (LLNL)
Christos Kartsaklis (ORNL)
Pat McCormick (LANL)
Sri Hari Krishna Narayanan (ANL)
Andrew Salinger* (SNL)
Jason Sarich (ANL)
Dali Wang (ORNL)
Jim Willenbring (SNL)

Extreme-Scale Scientific Software Development Kit
Lead: Lois Curfman McInnes (ANL)
Alicia Klinvex (SNL)
Jed Brown (ANL)
Irina Demeshko (SNL)
Anshu Dubey (LBNL)
Sherry Li (LBNL)
Vijay Mahadevan (ANL)
Daniel Osei-Kuffuor (LLNL)
Barry Smith (ANL)
Mathew Thomas (PNNL)
Ulrike Yang (LLNL)

Outreach and Community
Lead: David Bernholdt (ORNL)
Katie Antypas* (NERSC)
Lisa Childers* (ALCF)
Judy Hill* (OLCF)

IDEAS productivity

* Liaison
*1 *2 *3: Leads: Use Cases 1, 2, 3

Crosscutting Lead: Hans Johansen (LBNL)

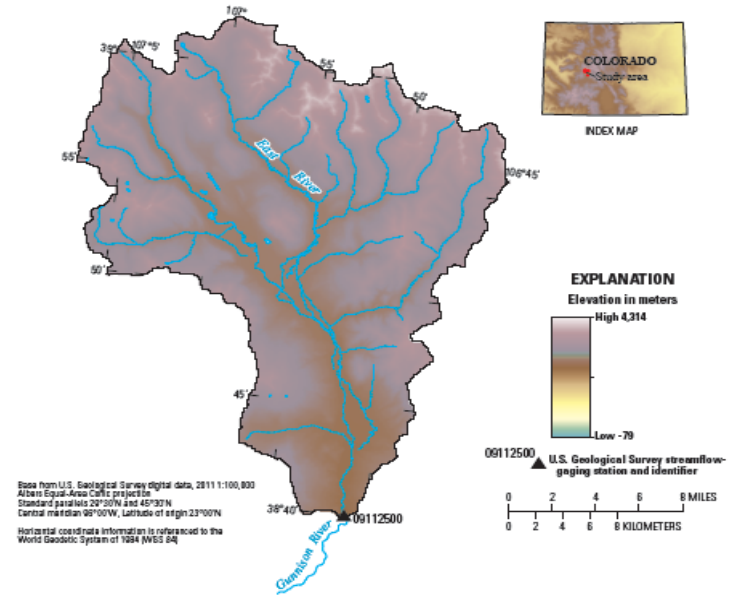
Sept 2015

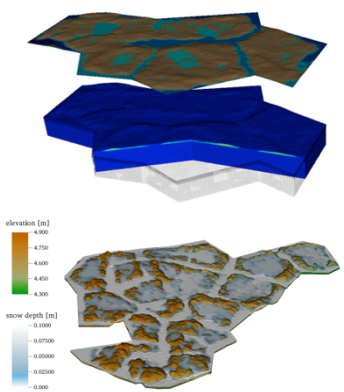
SFAs	NGEE	Exascale Co-Design	ASCR Math & CS	ALCF
CLM	ACME	Exascale Roadmap	SciDAC / NERSC	OLCF
BER Terrestrial Programs		DOE Extreme-scale Programs		DOE Computing Facilities

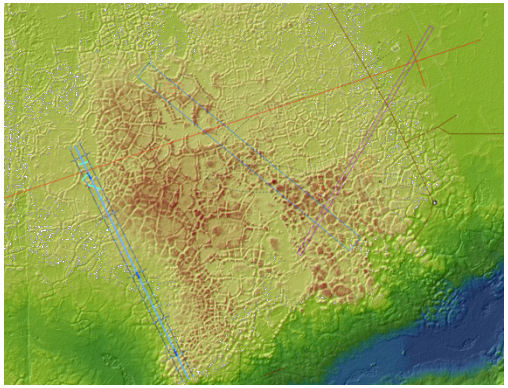
U.S. DEPARTMENT OF ENERGY
Office of Science

Use cases: Multiscale, multiphysics representation of watershed dynamics

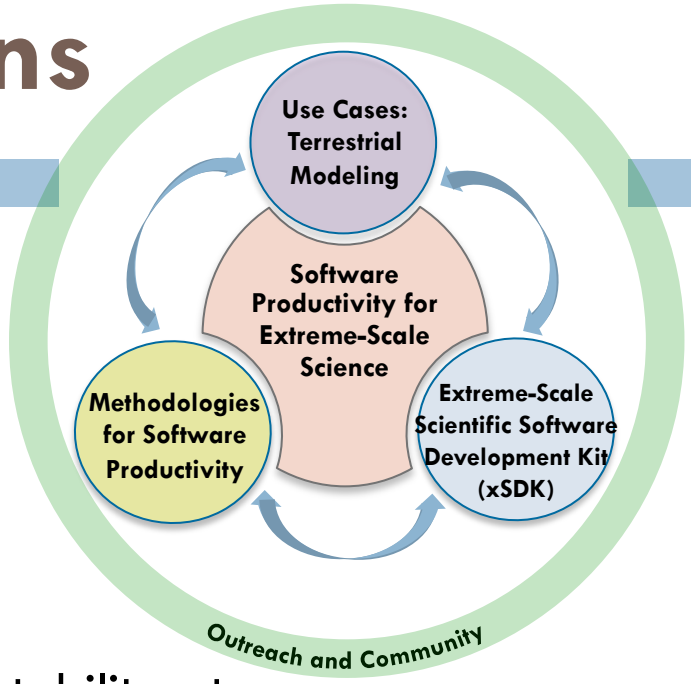
- **Use Case 1:** Hydrological and biogeochemical cycling in the Colorado River System
- **Use Case 2:** Thermal hydrology and carbon cycling in tundra at the Barrow Environmental Observatory
- **Use Case 3:** Hydrologic, land surface, and atmospheric process coupling over the continental United States
- **Leverage and complement existing SBR and TES programs:**
  - LBNL and PNNL SFAs
  - NGEE Arctic and Tropics
- **General approach:**
  - Leverage existing open source application codes
  - Improve software development practices
  - Targeted refactoring of interfaces, data structures, key components to facilitate interoperability
  - Modernize management of multiphysics integration and multiscale coupling





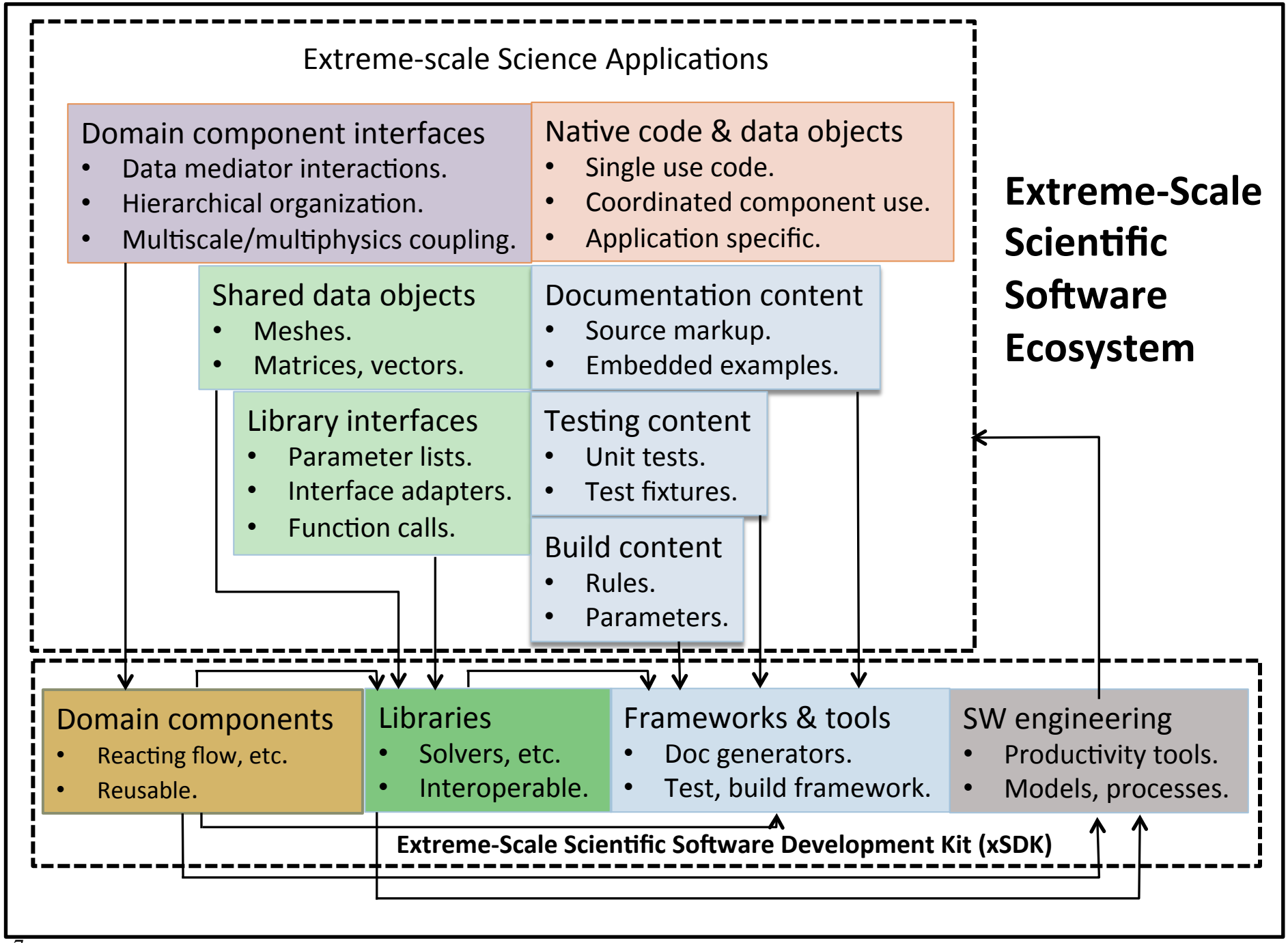


IDEAS interconnections



- **Use cases:** Drive efforts. Traceability from all efforts
  - But generalized for future efforts
- **Methodologies ("HowTo") for SWP:**
  - Infrastructure, testing, porting, refactoring, portability, etc.
  - Workflows, lifecycles: Document and formalize. Identify best practices
- **xSDK:** frameworks + components + libraries
  - Build apps by aggregation and composition
- **Outreach:** Foster communication, adoption, interaction

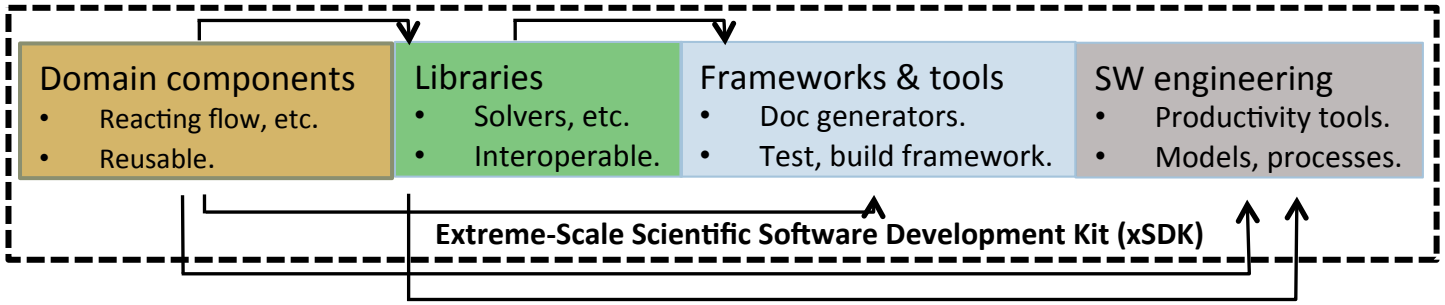
- **First of a kind: Focus on <u>software productivity</u>**


Extreme-scale Science Applications
Domain component interfaces
• Data mediator interactions.
• Hierarchical organization.
• Multiscale/multiphysics coupling.
Native code & data objects
• Single use code.
• Coordinated component use.
• Application specific.
Extreme-Scale Scientific Software Ecosystem
Shared data objects
• Meshes.
• Matrices, vectors.
Documentation content
• Source markup.
• Embedded examples.
Library interfaces
• Parameter lists.
• Interface adapters.
• Function calls.
Testing content
• Unit tests.
• Test fixtures.
Build content
• Rules.
• Parameters.
Domain components
• Reacting flow, etc.
• Reusable.
Libraries
• Solvers, etc.
• Interoperable.
Frameworks & tools
• Doc generators.
• Test, build framework.
SW engineering
• Productivity tools.
• Models, processes.
Extreme-Scale Scientific Software Development Kit (xSDK)

IDEAS xSDK

xSDK focus

- Common configure and link capabilities
  - xSDK users need full and consistent access to all xSDK capabilities
  - Namespace and version conflicts make simultaneous build/link of xSDK difficult
  - Determining an approach that can be adopted by any library or components development team for standardized configure/link processes
- Library interoperability
- Designing for performance portability



Standard xSDK package installation interface

Motivation: Obtaining, configuring, and installing multiple independent software packages is tedious and error prone.

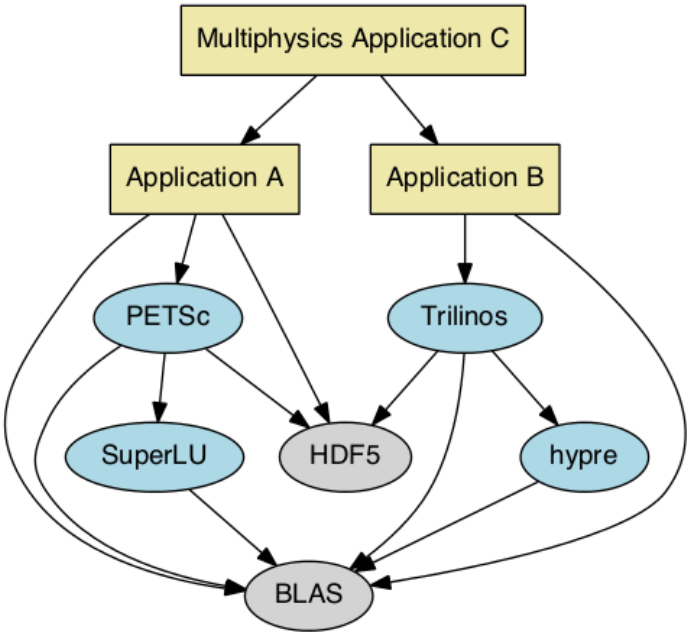
- Need consistency of compiler (+version, options), 3rd-party packages, etc.

Approach: Define a *standard xSDK package installation interface* to which all xSDK packages will subscribe and be tested

Accomplishments:

- Work on implementations of the standard by the hypre, PETSc, SuperLU, and Trilinos developers
- PETSc can now use the "scriptable" feature of the installers to simultaneously install hypre, PETSc, SuperLU, Trilinos with consistent compilers and 'helper' libraries.

xSDK Build Example



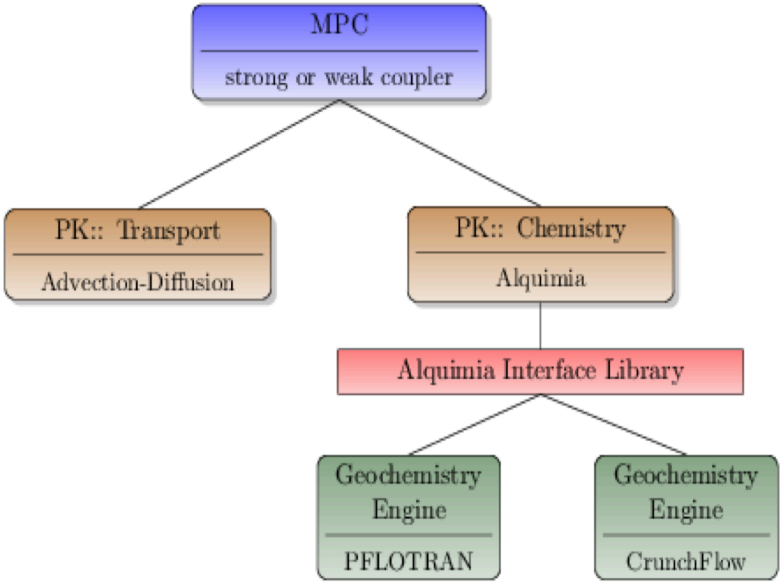
Impact: Foundational step toward seamless combined use of xSDK libraries, as needed by BER use cases and other multiphysics apps

IDEAS productivity

Enabling Interoperable Biogeochemistry with Alquimia

Several geochemistry libraries are well established in the community making geochemistry ideal to explore componentization and interface design. ***Alquimia* is an interface library, and does not perform any reaction calculations.**

- Alquimia currently assumes reactive transport uses operator-splitting.
- Fully-implicit reactive transport support is being developed in collaboration with IDEAS.
- Assists in enforcing geochemical conditions (speciation) for transport boundary conditions
- Alquimia can facilitate benchmarking of geochemical capabilities in existing codes.
- Geochemistry libraries, such as PFLOTRAN and CrunchFlow, have implemented interfaces to Alquimia.



Schematic of the Alquimia interface library providing uniform access to PFLOTRAN and CrunchFlow geochemistry in Amanzi

Alquimia is open source, https://bitbucket.org/berkeleylab/alquimia

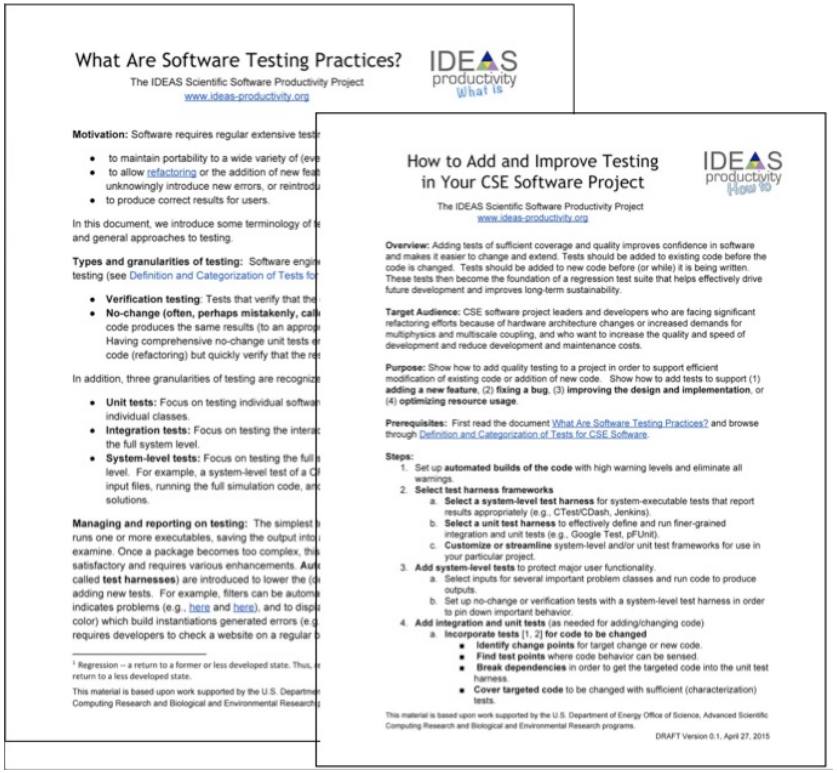
IDEAS productivity

Methodologies or “HowTo”

IDEAS 'What is' and 'How to' docs

- **Motivation:** Scientific software teams have a wide range of levels of maturity in software engineering practices
  - Baseline survey of *xSDK* and *BER Use Case* teams
- **Approach:**
  - **'What Is' docs**: 2-page characterizations of important software project topics
  - **'How To' docs**: brief sketch of best practices
    - Emphasis on ``bite-sized'' topics enables CSE software teams to consider improvements at a small but impactful scale.
- **Initial emphasis:**
  - *What is CSE Software Productivity?*
  - *What are Software Testing Practices?*
  - *How to Add and Improve Testing in Your CSE Software Project*
- **Topics in progress:**
  - Refactoring tools and approaches
  - Best practices for using interoperable libraries
  - Designing for performance portability
  - Etc.

https://ideas-productivity.org/resources/howtos



What Are Software Testing Practices?

The IDEAS Scientific Software Productivity Project

www.ideas-productivity.org

How to Add and Improve Testing in Your CSE Software Project

The IDEAS Scientific Software Productivity Project

www.ideas-productivity.org

Impact: Provide baseline nomenclature and foundation for next steps in SW productivity and SW engineering for CSE teams

Managing issues: Fundamental software process

- Issue: Bug report, feature request
- Approaches:
  - Short-term memory, office notepad
  - ToDo.txt on computer desktop (1 person)
  - Issues.txt in repository root (small co-located team)
  - …
  - Web-based tool + Kanban (distributed, larger team)
  - Web-based tool + Scrum (full-time dev team)
- IDEAS project:
  - Jira Agile + Confluence: Turnkey web platform (ACME too)
  - Kanban: Simplest of widely known Agile SW dev processes

Kanban principles

- Limit number of "In Progress" tasks
- Productivity improvement:
  - Optimize "flexibility vs swap overhead" balance. No overcommitting.
  - Productivity weakness exposed as bottleneck. Team must identify and fix the bottleneck.
  - Effective in R&D setting. Avoids a deadline-based approach. Deadlines are dealt with in a different way.
- Provides a board for viewing and managing issues

Task: Have Eureka moment by Tuesday.

IDEAS Confluence, Jira Agile, Kanban

Pages / IDEAS Common Home

Managing IDEAS Project Activities Using JIRA Agile and Kanban

Created by Michael Heroux, last modified by Jim Willenbring about an hour ago

Contents:

- Overview
- Standard JIRA Agile Issue Type Definitions
- IDEAS Issue Management Structure
  - IDEAS Projects
  - Issue Status
  - Viewing Issues
- Entering an IDEAS Project Activity
- Managing Existing IDEAS Issues

Developer Guide, on Confluence site

Kanban Board, on Jira site.

Four columns:

- To Do
- Selected
- In Progress
- Done

Kanban board

Board

QUICK FILTERS: No subtasks | No ToDo | Epics | No epics | HowTo | Outreach | UseCase | xSDK | Only My Issues | FY15 | ... Show more

24 of 64 To Do

- HOW-10 HowTo-2.1: Assess and measure SE maturity within IDEAS teams
- HOW-14 HowTo-2.1: Measure SW productivity improvement and benefits
- HOW-12 HowTo-2.1: Establish scientific software ecosystem lifecycle model
- HOW-11 HowTo-2.1: Develop What Is and How To Content for Improving CSE SW Practices
- HOW-40 HowTo-2.1: Adapt and provide tools & processes that will enhance IDEAS &
- HOW-41 HowTo-2.1: Provide training opportunities for IDEAS Team members
- HOW-4

5 of 13 Selected

- HOW-22 Develop plan an agile software ecosystem lifecycle model (ASELM) for — HowTo-2.1: Establish sci...
- HOW-28 Create a Performance Portability What-Is Document — HowTo-2.1: Develop Wha...
- HOW-37 Coordinate all interested IDEAS team members to discus Jenkins setup/use — HowTo-2.1: Provide traini...
- HOW-38 Develop strategy for addressing contributor agreements for open-source
- HOW-42 Develop training material on how to use Jira Agile and Kanban — HowTo-2.1: Provide traini...

6 of 24 In Progress

- HOW-7 Based on assessment, recommend use case testing practices and infrastructure — HowTo-2.2: Best practice...
- HOW-36 Create Confluence page for guiding IDEAS team member on how to use Jira — HowTo-2.1: Adapt and pr...
- HOW-16 Single repository git workflow best practices. — HowTo-2.1: Establish sci...
- HOW-24 Assess status of use case testing — HowTo-2.1: Assess and ...
- HOW-26 Develop a list of future How to topics to post on ideas-productivity.org — HowTo-2.1: Develop Wha...

6 of 32 Done — Release...

- ~~HOW-1~~ Hans adds IDEAS leads to JIRA and Confluence
- ~~HOW-2~~ Write up summary notes from Methodology telecon
- ~~HOW-9~~ Hans yells at Procurement about Pcard
- ~~HOW-18~~ Conduct a GQM style survey refactoring opportunities and needs of UseCases — HowTo-2.1: Establish sci...
- ~~HOW-19~~ Evaluate the readiness of UseCases in Identifying change points, Finding test — HowTo-2.1: Establish sci...
- ~~HOW-35~~ Document How the IDEAS Project will Use Jira for Issue

Outreach

Outreach and community

- **Objective:** *Begin* changing the way computational and domain science communities think about software development through increased awareness, education, and active engagement
- Building connections to computing facilities
- Presentations and posters at BER PI meetings, SE-HPCCSE14 workshop, organized SIAM CSE 2015 mini-symposium
- Leadership roles in NITRD CSESSP (sustainability and productivity) workshop (Oct 2015)
- BOF proposal for SC15 with international colleagues

Software productivity discussion at OLCF Annual User Meeting, June 2015

CSESSP workshop homepage, October 2015

IDEAS productivity

cse-software.org screen shot

cse-software.org

Computational Science & Engineering Software Forum

HOME ANNOUNCEMENTS WORKING GROUPS JOIN

Computational Science & Engineering (CSE) Software Forum

Who are we?

The Computational Science & Engineering (CSE) Software Forum is an international group of researchers and other stakeholders from national laboratories, academic institutions, and industry who are interested in (and concerned about) **development processes for software for computational science and engineering** and related technical computing areas, with a particular interest in CSE **on high-performance (parallel) computers**.

What are we doing?

We're creating a clearinghouse to **gather, discuss, and disseminate** experiences, techniques, tools, and other resources for the entire lifecycle of CSE software, from initial concept through the growth phase to maturity, re-investment, and eventually to retirement.

We want to **raise awareness** of the importance of good software practices to scientific productivity and to the quality and reliability of computationally-based scientific results, as well as raising awareness of the increasing challenges facing CSE software developers as high-end computing heads to extreme scales.

Search

RECENT POSTS

Computational Science & Engineering Software Sustainability and Productivity Challenges (CSESSP Challenges)
Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE2.1@SciPy)
Third International Workshop on Software Engineering for High Performance Computing in Computational Science and Engineering
3rd Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE2) CFP1

IDEAS productivity

Wrapup

IDEAS Value Proposition

- IDEAS is DOE's first Productivity-specific effort.
- Targeted Problems:
  - BER Use Cases: Real impact plus drivers, validators.
  - Enables focused effort on topics with broad applicability.
- SW Ecosystem:
  - xSDK: Interoperable libraries, components.
  - Reduces installation, co-play and interoperability costs.
- Methodologies:
  - Exploration, adaption, adoption for Extreme-scale CSE.
- Outreach:
  - Engaging the broader community.

Why Interact With Us?

Here today:
- Mike Heroux
- Lois McInnes
- David Bernholdt
- David Moulton
- Hans Johansen
- Ulrike Yang
- Anshu Dubey
- Thomas Ndousse-Fetter (ASCR)

- Targeting Gnarly Problems:
  - BER problems representative of multi-scale/physics trends.
  - Goal: Generalize experience to impact broader community.
- Team represents 100+ years cumulative experience:
  - Designing, developing community-oriented reusable SW.
  - Working dozens of large-scale, real apps.
  - API design and implementation.
- Opportunity to consider explicit productivity focus:
  - Conduct studies in SW methodologies, tools.
  - Engage community in collection & dissemination of ideas.

Engaging with IDEAS

- Talk to us: ideas-leads@lists.anl.mcs.gov
- See our website: http://ideas-productivity.org
- Join the community: go to http://cse-software.org and join the cse-forum mailing list (or cse-announce)

Argonne NATIONAL LABORATORY
BERKELEY LAB
Lawrence Livermore National Laboratory
Los Alamos NATIONAL LABORATORY EST. 1943
OAK RIDGE National Laboratory
Pacific Northwest NATIONAL LABORATORY
Sandia National Laboratories
SCHOOL OF MINES 1874 COLORADO
U.S. DEPARTMENT OF ENERGY | Office of Science

IDEAS productivity