

LSSw Meeting 6

March 17, 2022

Preview for LSSw Meeting 7: April 21, 2022

Topic: Expanding the Scope of What is Reusable: A panel discussion

Description: Application-specific, libraries and tools have also had some success, for example, the Co-Design Centers sponsored by the Exascale Computing Project, but have received less attention and can be more challenging to sustain.

Panelists to help explore expanding the kinds of functionality that can be encapsulated for reuse:

- Slaven Peles, ORNL
- Andrew Salinger, SNL
- Andrew Siegel, ANL
- Angela Herring, LANL
- One more, TBD

Prompts:

- Do you think there is value in designing, implementing, and delivering application-specific libraries, tools, and environments as reusable components?
- What has worked and not worked well with past efforts in this area?
- What are some near-term opportunities to componentize in your application area?
- How could this kind of software collection be adapted and sustained?

LSSw Meeting 6: Mar 17, 2022

- Topic: Scientific Software Ecosystems: A panel discussion
- Description: This month we have panelists representing other technical software ecosystems:
 - Anita Carleton, CMU, SEI
 - Lorraine Hwang, UC Davis, CIG
 - Elizabeth Sexton-Kennedy, Fermi Lab, HSF
 - Andy Terrel, Xometry, NumFocus
 - Theresa Windus, Iowa State, MolSSI
- Prompts:
 - What is the value proposition of your ecosystem to its developer and user communities?
 - What is the business model of your ecosystem (how do people fund their efforts)?
 - What are some of the challenges you face in providing value?
 - What are your sustainability challenges?



Scientific Software Ecosystems: A Panel Discussion

Anita Carleton

March 17, 2022

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Document Markings

Copyright 2022 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

DM22-0236

CMU SEI is a DoD R&D Federally Funded Research and Development Center



Established in 1984 at Carnegie Mellon University (CMU) in Pittsburgh, PA

Charged to advance the state of the practice of software engineering, cybersecurity, and AI Engineering

Collaborates with other departments in CMU and broadly in academia, government, and industry

Capable of conducting both fundamental research and classified work

~610 staff members

FY20 total funding ~\$140M. Offices in Pittsburgh and DC, with locations near customer facilities in MA, TX, and CA

Anita Carleton, Division Director of the Software Solutions Division (SSD) at Carnegie Mellon University's Software Engineering Institute

- More than 30 years of technical and senior leadership experience in the software engineering industry.
- Mission of SSD is to advance the state of the practice in software engineering through applied research, development, and transition of innovative technologies for building and acquiring software-intensive systems, with the specific goal of making software a strategic advantage for the Department of Defense.
- Provides leadership through research collaborations with CMU and others in academia, directly engaging with major defense acquisition projects, partnering with industry, and informing DoD policy makers.



Most recently led an SEI study engaging the software engineering community to define a national agenda for software engineering research and development for the next decade titled *Architecting the Future of Software Engineering: A National Agenda for Software Engineering Research & Development*.

What Are The Value Propositions Of Your Ecosystem To Its Developer And User Communities?

- We are a CATALYST for the DoD.
- We bring value by BRIDGING disparate communities (researcher, practitioner, acquirer AND academia, industry, government), who operate according to their own respective incentives,
- We ACCELERATE the ADOPTION of the best techniques and practices and ACCELERATE RESEARCH in key areas where the state of the art falls short.
- We do this, in part, by addressing customer-specific needs, but always in a way that also addresses the community's needs as well.

It is in this way that ***we make software a strategic advantage for the DoD.***



What Is The Business Model Of Your Ecosystem (how do people fund their efforts)?

The SEI's choice of technical work is driven by its sponsoring agreement and by its close alignment with DoD technology modernization priorities:

- Congressionally allocated funding for selected research.
- SEI performs work on direct program work plans that benefit Major Defense Acquisition Programs (MDAPs), executive agencies, service branches, combat support agencies and service labs.
- Additionally, the SEI works with federal civilian, independent federal agencies, and private sector companies and collaborates with CMU and other universities on research projects.
- Influences the software engineering community by holding leadership positions in key technical societies such as the Institute of Electrical and Electronics Engineers (IEEE), the largest technical professional organization. SEI technical staff serve as IEEE Computer Society President, on the IEEE Board of Governors, as the *IEEE Software* Editor-in-Chief, and on the *IEEE Software* Advisory Board.



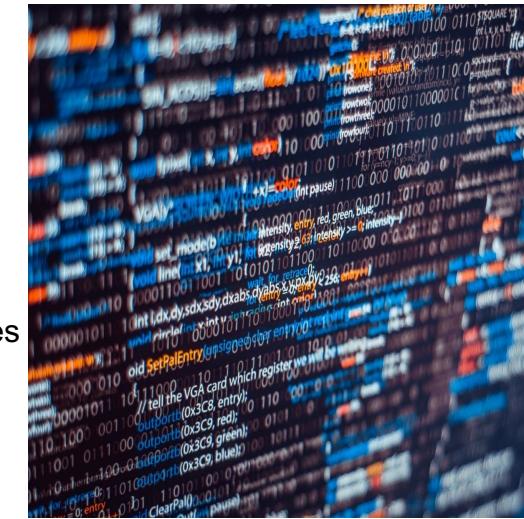
What Are Some Of The Challenges You Face In Providing Value To Your Members And Other Stakeholders?

The DoD faces enduring challenges as the need for software innovation and cybersecurity evolves and intensifies. The 2018 National Defense Strategy (NDS) and other DoD guidance make clear that the DoD needs its software-enabled systems to

- **Bring Capabilities** that make new missions possible or improve the likelihood of success of existing ones
- **Be Timely** so that the DoD is able to field new software-enabled systems and their upgrades faster than our adversaries
- **Be Trustworthy** in construction and implementation and resilient in the face of operational uncertainties including known and yet unseen adversary capabilities
- **Be Affordable** such that the cost of acquisition and operations, despite increased capability, is reduced and predictable and provides a cost advantage over our adversaries

Other challenges:

- Tension between today's challenges and future anticipated needs...How do we convince people to pay for something that is not yet causing them problems?
- Another challenge is that we are a small organization, so that we need to produce work products that have benefit beyond a specific customer.
- We also have to find other sources of leverage like standards, communities of interest, practitioner conferences,



What Are Your Sustainability Challenges?

Design for Sustainability in Computing will require:

- Fundamentally new and disruptive research across all aspects of computing including modeling, design, reuse, programming, data management, and digital and computing-based technologies
- Discovering new design and analysis principles for engineering intelligent software systems
- Contributing to the development of a discipline for AI engineering by building on its software engineering strengths in assurance, architecture, formal analysis, DevOps, and automation
- Design for reusability principles across some or all levels of the entire computing stack to avoid obsolescence and enable longevity for devices (e.g., smartphones, IoT), including modular design for updating or common product-line sharing
- Advances in computer architectures including reconfigurable architectures
- Investigating emerging technologies, such as integration of quantum components into software systems, use of low-code platforms, and designing self-modifying architectures

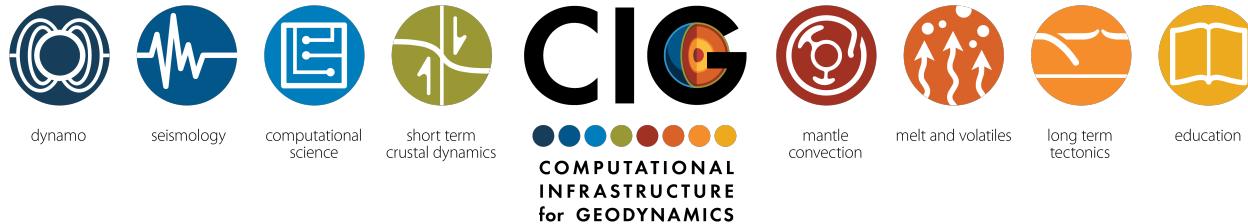


Lorraine J. Hwang

University of California Davis

CIG Director

- Background in seismology – waveform and crustal structure modeling
-

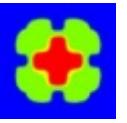


Founded in 2005 to reduce duplication of effort in software development for the geodynamics community and support the development of robust and scalable software.

What are the value propositions of your ecosystem to its developer and user communities?



Open communities and  open software and tools built on modern, robust, well tested and documented software and components.

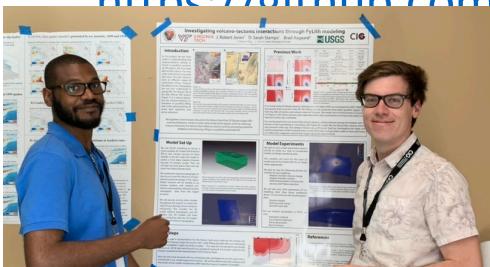


Software Best Practices

https://github.com/geodynamics/best_practices



Training



Workshops



Tutorials



Hackathons

What is the business model of your ecosystem (how do people fund their efforts)?

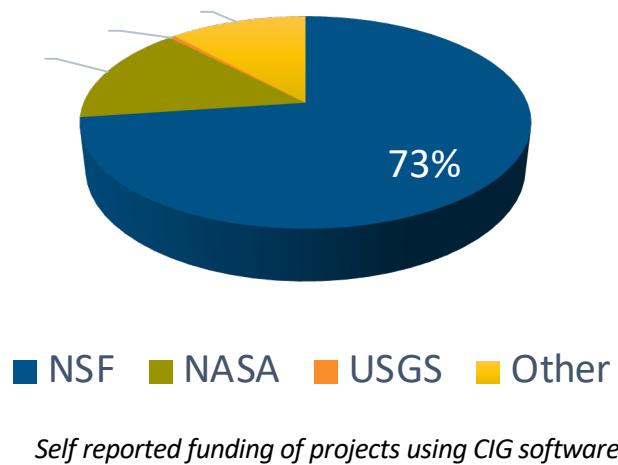
Membership (89):

93% University

7% Governmental

Grant funding:

Primarily NSF, NASA, USGS, AFRL, DOE,
international, some private foundations.



CIG funded by NSF Geoinformatics with contributions from other programs.

What are some of the challenges you face in providing value to your members and other stakeholders?

Keeping up with the pace of change.

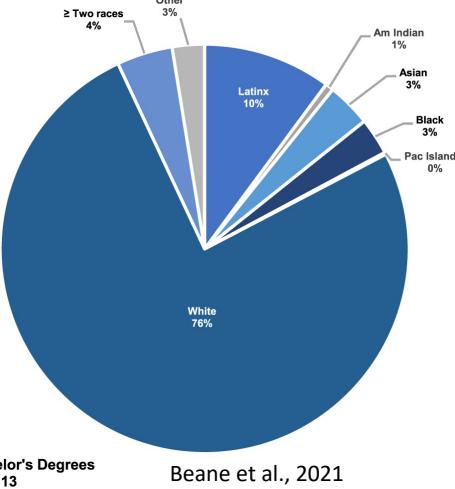
Supporting data-intensive and complex workflows

Supporting and maintaining resources (people and technology)

What are your sustainability challenges?

Diversifying and maintaining:

- Funding sources
- The pipeline
 - Diversity
 - Associate Project Scientist \$80k, industry 2x's
- Limiting technical debt – depends on ecosystem of standardized and well tested components.
- Interoperability – between codes, between components, between platforms



Elizabeth Sexton-Kennedy (Liz)

- Head of the Computing Sector, CIO at Fermilab
- 17+13 as a scientific software developer / architect / manager at Energy Frontier experiments in HEP
 - 2018 - Computing and Software project then program manager for CMS
 - 2005 - Software developer / architect for CMS
 - 2001 – Computing and Software program manager for CDF
 - 1988 – Software developer / architect for CDF

Prompts

From the perspective of the scientific software ecosystem you represent:

- The primary value to our community is the efficient and timely delivery of the science mission. In HEP almost all domain scientists are also developers, however the need for software engineers is accelerating.
- People are part of a scientific program that is funded for decades at a time, except when they are not...
- The biggest challenge has been convincing stakeholders that S&C is NOT easy. It is not just something you can assign to a graduate student or postdoc.
- Software developed outside of programs as part of a research proposal like SciDAC or ECP loose funding when the research project is done. These projects are expected to hand off the software to a program but there is often no one there to hand it off to.

Andy R. Terrel

- Positions
 - Former President of NumFOCUS (2012 – 2021),
 - Founding Team of Anaconda (2012 – 2015),
 - VP, Data and Algorithms at Xometry (current)
- Use of Tools
 - Library author of Dask, SymPy, FEniCS, and various PyData packages
 - Conference organizer, PyData, SciPy, and various <Package>Cons
 - Academic publications in HPC software automation,
 - Industry data science leader

NumFOCUS

- a 501(c)(3) public charity
- promotes open practices in research, data, and scientific computing
- serving as a fiscal sponsor for open source projects and organizing community-driven educational programs.
- Responses based on our 45 [Sponsored](#) and 51 [Affiliated](#) projects.

Value Proposition

- NumFOCUS's comprehensive stewardship includes
 - project advocacy,
 - legal and operational support,
 - fiscal administration and grant management,
 - governance support,
 - DEI assistance,
 - community outreach,
 - events, and
 - user education.
- Very similar to Universities and other Open Source Non-Profits (Apache, Mozilla, Linux Foundation, etc.)

Value Differentiation

- Focused on Scientific Software
- A funding network (or "fabric") made up of diverse sources (we take checks from everyone)
- A transparent and inclusive governance structure
- Not concentrated at a single entity (Universities and corporations have a major challenge supporting the diversity of our community)

Business Model

NumFOCUS Organization

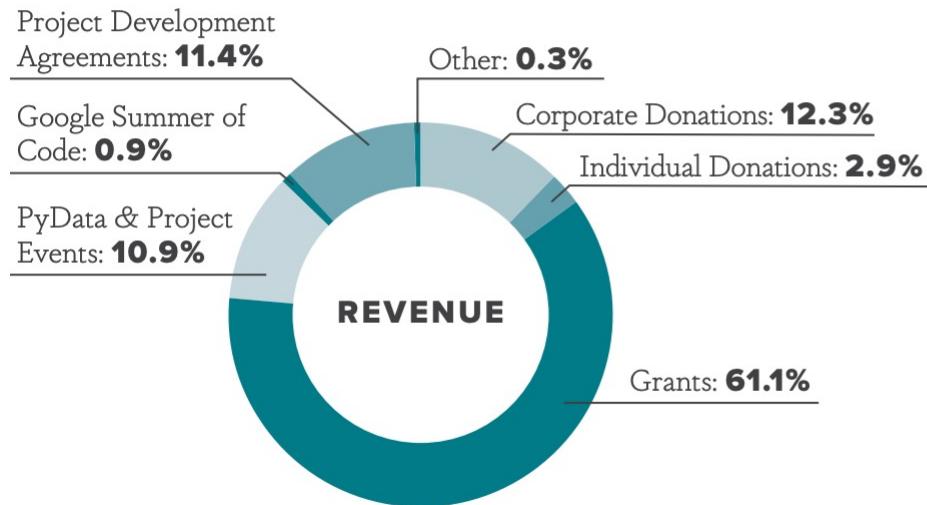
- Conferences
- Corporate Sponsorships
- Individual donations

Project accounts

- Government Grant
- Private Grants
- Corporate donations for specific projects

2020 Annual Report

REVENUE



REVENUE*	\$5,140,483.54
Corporate Donations	\$633,334.23
Individual Donations	\$140,469.00
Grants	\$3,154,332.00
PyData & Project Events	\$564,627.81
Google Summer of Code	\$45,000.00
Project Development Agreements	\$589,780.83
Merchandise Sales	\$5,809.73
Interest	\$7,129.94
	\$3,916,700.00

*Restricted for projects and programs

See <https://numfocus.org/community/mission/annual-report>

Challenges in providing our value

- Keeping management structures community led but not pushing more work onto volunteers
- Lack of funding geared toward project management, documentation, and community management
- Qualified people to facilitate community stewardship

What are your sustainability challenges?

- Attracting contributors from broader communities
- Identifying and assessing relevant skill sets, domain knowledge, and technical skills

Theresa Windus

- Distinguished Professor of Chemistry at Iowa State University and Senior Scientist at Ames Laboratory
- Over 30 years of experience in Computational Chemistry and High-Performance Computing
 - Molecular Sciences Software Institute, <https://molssi.org/>
 - NWChemEx project director



What are the value propositions of your ecosystem to its developer and user communities?

- Ecosystem is computational molecular sciences developers and users
 - Chemistry, materials, biophysics, biology, medicine, climate science, engineering, ...
- Central tools of our field and enable us to do the science
- As a computational scientist, software development is an integral part of understanding the problem/mathematics/physics
- Developed and used in academia, national labs, and industry
- Range from small to large codes
- Some are decades old, while others began in the last few years
- Moving toward more interoperability
- Components that can be used (examples: libxc and libint, basis set exchange)

What is the business model of your ecosystem (how do people fund their efforts)?

- Academics tend to be open source
 - Proposals – usually for new science, physical models
 - Some exceptions to the latter are ECP, SciDAC (to some extent), NSF MolSSI, NSF CISE, etc.
- Commercial has varying models
 - Charge all or just industry and national labs
 - Proposals – STTR and SBIR
 - Often has an academic “backend” to help drive the innovation
- MolSSI is pursuing other models

What are some of the challenges you face in providing value to your members and other stakeholders?

- Maintenance and porting of software
 - Hardware advances quickly and our ecosystem has challenges keeping up
 - Software library changes that discourage interoperability
 - Hardening, porting, and tuning of software
 - Documentation for developers and users
- Significant fragment of software is developed in academic research
 - Different goals and incentives than commercial software developers
 - Less focus on maintenance, documentation, and user interfaces
 - Turn-over is high
 - Students/postdocs who are supposed to move to full-time careers
 - Lure of other industrial positions – pay, stability, new challenges

What are your sustainability challenges/potential solutions?

- Challenges as in previous slide
- Funding for good software engineering processes within the ecosystem
 - MolSSI Support
- Bringing business experience alongside scientists to help build and try different business models (subscriptions, philanthropy, etc.)