Understanding and Improving the Perceived Value of Software
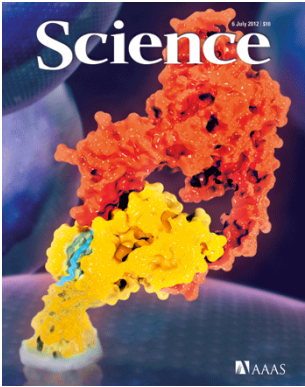
Daniel S. Katz

d.katz@ieee.org
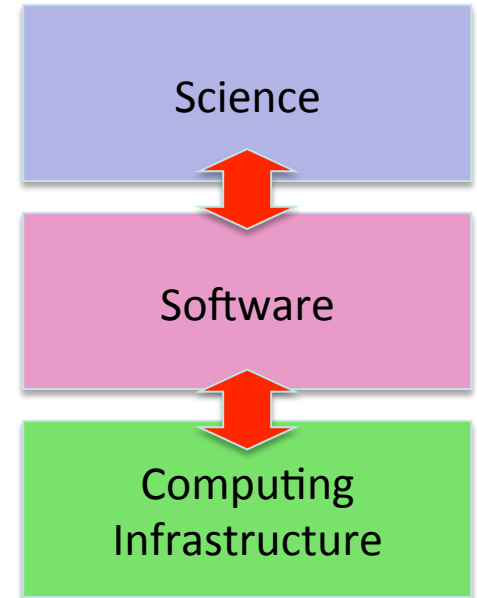
@danielskatz

University of Chicago &
Argonne National Laboratory
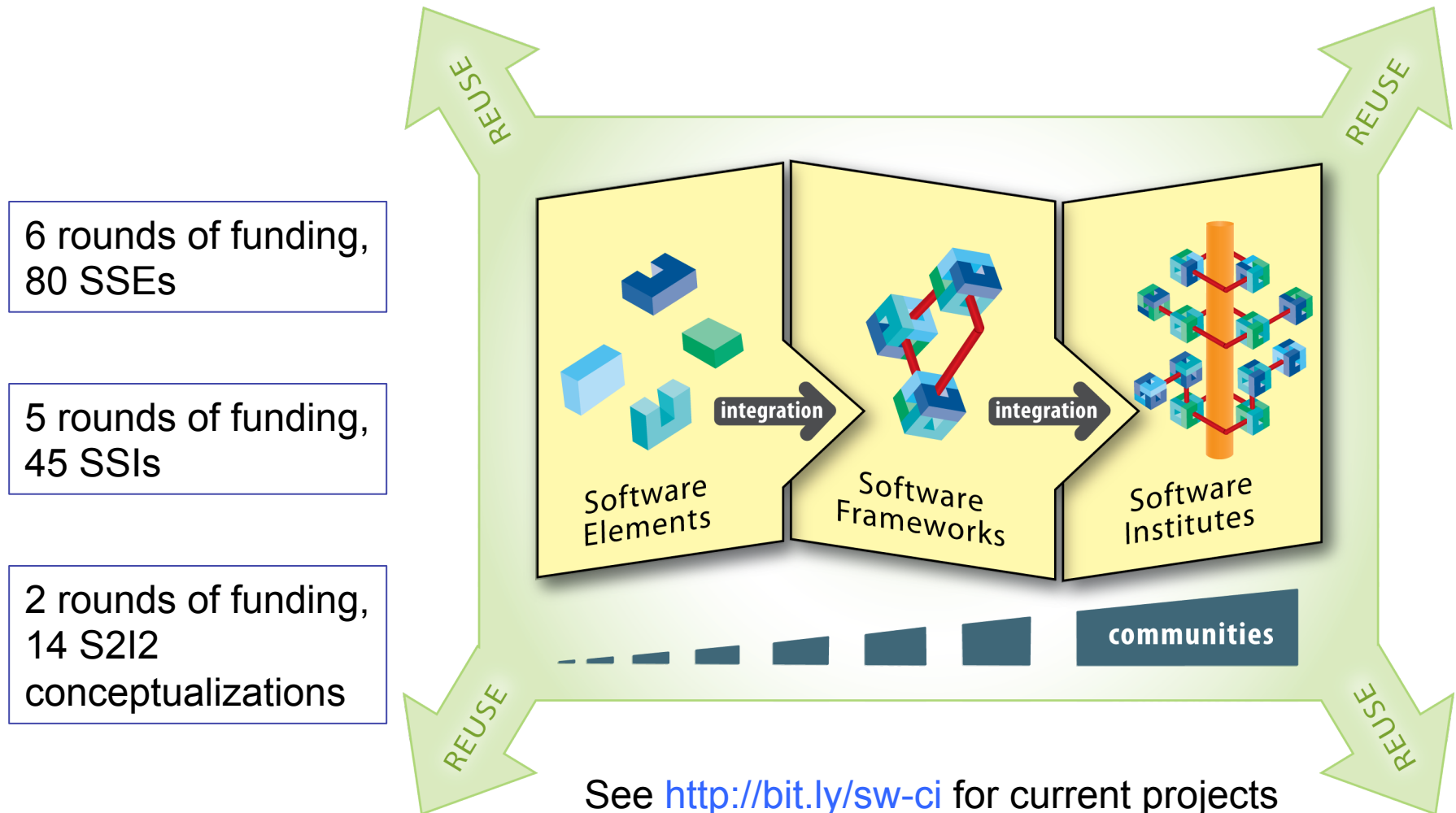
Software as Infrastructure

- Software (including services) essential for the bulk of science
  - About half the papers in recent issues of Science were software-intensive projects
  - Research becoming dependent upon advances in software
  - Significant software development being conducted across NSF: NEON, OOI, NEES, NCN, iPlant, etc
  - Wide range of software types: system, applications, modeling, gateways, analysis, algorithms, middleware, libraries
- **Software** is not a one-time effort, it **must be sustained**
  - Development, production, and **maintenance** are people intensive
  - Software life-times are long vs hardware
  - Software has under-appreciated value

Science
Software
Computing Infrastructure

For software to be sustainable, it must become infrastructure

NSF Software Infrastructure Projects

6 rounds of funding,
80 SSEs

5 rounds of funding,
45 SSIs

2 rounds of funding,
14 S2I2
conceptualizations

See http://bit.ly/sw-ci for current projects

SI^2 Solicitation and Decision Process

- Proposal reviews well -> lead PO's role becomes matchmaking
  - Find other program officers with funds, and convince them that they should spend their funds on the proposal
- Unidisciplinary project (e.g. bioinformatics app)
  - Work with single program officer, either likes the proposal or not
- Multidisciplinary project (e.g., molecular dynamics)
  - Work with multiple program officers, ...
- Omnidisciplinary project (e.g. http, math library)
  - Try to work with all program officers, often told "it's your responsibility"

To judge software, need to understand/forecast impact

Measuring Impact – Scenarios

1. Developer of open source physics simulation
   - Possible metrics
     - How many downloads? (easiest to measure, least value)
     - How many contributors?
     - How many uses?
     - How many papers cite it?
     - How many papers that cite it are cited? (hardest to measure, most value)

2. Developer of open source math library
   - Possible metrics are similar, but citations are less likely
   - What if users don't download it?
     - It's part of a distro
     - It's pre-installed (and optimized) on an HPC system
     - It's part of a cloud image
     - It's a service

- Future impacts – let proposers suggest

ACI Software Cluster Programs

- In these programs, ACI works with other NSF units to support projects that lead to software as an element of infrastructure

- Issue: amount of software that is infrastructure grows over time, and grows faster than NSF funding

Q: How can NSF ensure that software as infrastructure continues to appear, without funding all of it?
A: Incentives

- The devil is in the details

Other Software Discussions

- Working Towards Sustainable Software for Science: Practice and Experience (WSSSPE)
    - Google: WSSSPE
    - 3 previous workshops, two upcoming
    - WSSSPE2.1 Friday, SciPy, Austin
    - WSSSPE3 Sept 29-30, Boulder, CO
- Lessons:

Many of the issues in developing sustainable software are social, not technical

Software work is inadequately visible in ways that "count" within the reputation system underlying science

Where We Are

- To judge software, need to understand/forecast impact
- Q: How can NSF ensure that software as infrastructure continues to appear, without funding all of it?
- A: Incentives
- Many of the issues in developing sustainable software are social, not technical
- Software work is inadequately visible in ways that "count" within the reputation system underlying science

Hypothesis: better measurement of contributions can lead to rewards (incentives), leading to career paths, willingness to join communities, leading to more sustainable software

Moving Forward - NSF

- Recent CISE/ACI & SBE/SES Dear Colleague Letter: Supporting Scientific Discovery through Norms and Practices for Software and Data Citation and Attribution (NSF 14-059, http://www.nsf.gov/pubs/2014/nsf14059/nsf14059.jsp)
  - Need well-developed metrics to assess the impact and quality of scientific software and data
  - Explore new norms and practices for software and data citation and attribution, so that data producers, software and tool developers, and data curators are credited
- 6 projects and 3 collaborative workshops funded

Moving Forward - Dan

- Products (software, paper, data set) are registered
  - Credit map (weighted list of contributors—people, products, etc.) is an input
  - DOI is an output
  - Leads to **transitive credit**[1]
    - E.g., paper 1 provides 25% credit to software A, and software A provides 10% credit to library X -> library X gets 2.5% credit for paper 1
    - Helps developer show: "my tools are important"
  - Issues:
    - Social: Trust in person who registers a product
    - Technological: How[2], Registration system

[1]D. S. Katz, "Transitive Credit as a Means to Address Social and Technological Concerns Stemming from Citation and Attribution of Digital Products," Journal of Open Research Software, v.2(1): e20, 2014. DOI: 10.5334/jors.be
[2]D. S. Katz, A. M. Smith, "Implementing Transitive Credit with JSON-LD," 2nd Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE2), 2014. URL: http://arxiv.org/abs/1407.5117

Moving Forward - Community

- Career paths – Is there a role for non-tenure-track researchers who produce software, data, etc. in universities?
    - Assuming yes, do universities recognize and support this? If not, how to get them to?
    - Potential workshop and possible NRC study
- What is needed to support reproducibility of science, in terms of data and software?
- Project Credit: http://projectcredit.net
- Force11, Software Citation working group: https://github.com/force11/force11-scwg
- Lots of entities with similar interests in both software and data, e.g. RCUK, NIH, DOE, Wellcome, Sloan & Moore, Mozilla, Apache, etc.
- Participate in WSSSPE

Resources

- NSF Software as Infrastructure Vision:
  http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf12113
- Implementation of NSF Software Vision:
  http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504817
- Software Infrastructure for Sustained Innovation (SI2) Program
  - Scientific Software Elements (SSE) & Scientific Software Integration (SSI) solicitation: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf14520
  - 2013 PI meeting: https://sites.google.com/site/si2pimeeting/
  - 2014 PI meeting: https://sites.google.com/site/si2pimeeting2014/
  - Awards: http://bit.ly/sw-ci
- Working towards Sustainable Software for Science: Practice and Experiences (WSSSPE)
  - Home: http://wssspe.researchcomputing.org.uk (includes links to all slides & papers)
  - 1st workshop paper: http://doi.org/10.5334/jors.an
  - 2nd workshop paper: http://arxiv.org/abs/1507.01715
  - 3rd workshop site: http://wssspe.researchcomputing.org.uk/wssspe3/
- NSF 14-059: "Dear Colleague Letter - Supporting Scientific Discovery through Norms and Practices for Software and Data Citation and Attribution"
  - http://www.nsf.gov/pubs/2014/nsf14059/nsf14059.jsp
- Transitive Credit Papers
  - http://dx.doi.org/10.5334/jors.be
  - http://arxiv.org/abs/1407.5117

Credits:

- SI2 Program:
  - Current program officers: Dan Katz, Rajiv Rammath, William Chang, John Cherniavsky, Almadena Chtchelkanova, Cheryl Eavey, Evelyn Goldfield, Sol Greenspan, Daryl W. Hess, Hao Ling, Peter McCartney, Bogdan Mihaila, Dimitrios V. Papavassiliou, Andrew Pollington, Massimo Ruzzene, Nigel Sharp, Eva Zanzerkia
  - Formerly-involved program officers: Manish Parashar, Gabrielle Allen, Sumanta Acharya, Jean Cottam-Allen, Rudolf Eigenmann, Eduardo Misawa, Barbara Ransom, Thomas Russell, Thomas Siegmund, Paul Werbos

- WSSSPE:
  - Organizers: Daniel S. Katz, Gabrielle Allen, Sou-Cheng (Terrya) Choi, Neil Chue Hong, Karen Cranston, Sandra Gesing, Lorraine Hwang, Manish Parashar, David Proctor, Erin Robinson, Matthew Turk, Colin C. Venters, Nancy Wilkins-Diehr
  - WSSSPE summary paper authors: Daniel S. Katz, Sou-Cheng T. Choi, Hilmar Lapp, Ketan Maheshwari, Frank Löffler, Matthew Turk, Marcus D. Hanwell, Nancy Wilkins-Diehr, James Hetherington, James Howison, Shel Swenson, Gabrielle D. Allen, Anne C. Elster, Bruce Berriman, Colin Venters, Neil Chue Hong, Frank Seinstra, Matthew Jones, Karen Cranston, Thomas L. Clune, Miguel de Val-Borro, Richard Littauer
  - WSSSPE keynote speakers: Phil Bourne, Arfon Smith, Kaitlin Thaney, Neil Chue Hong