Conceptualizing a US Research Software Sustainability Institute (URSSI)

Daniel S. Katz, Jeffrey C. Carver, Sandra Gesing, Karthik Ram, Nicholas Weber

http://urssi.us

Assistant Director for Scientific Software & Applications, NCSA
Research Associate Professor, CS, ECE, iSchool
d.katz@ieee.org, @danielskatz
Research software

Software developed and used for the purpose of research: to generate, process, analyze results within the scholarly process
Why do we care about research software?

• NSF
  • 1995-2016: 18,592 awards totalling $9.6 billion with project abstracts that topically include “software”
  • ~20% of the overall NSF research budget

• DOE
  • Of 3 Exascale Computing Project areas, most of 2 (application development, software) are research software
  • “ECP is a 7-year project with a cost range of $3.5B–$5.7B”
    -- Paul Messina, 2017

Collected from http://www.dia2.org in 2017
Why do we care about research software?

- 40 papers in Nature (Jan-Mar 2016)
  - 32 explicitly mentioned software
  - Average 6.5 software tools/paper
  - Most research software

- Top 100-cited papers:
  - 6 of top 13 are software papers
  - Vast majority describe experimental methods or software essential in their fields”

Nangia & Katz, 10.1109/eScience.2017.78
Nature, 10.1038/514550a
Why do we care about research software?

  - Use research software: 92% / 95% (UK/US)
  - Research not possible without software: 67% / 63%
  - Research possible but harder: 21% / 31%
  - Develop my own software: 56% / 28%

Hettrick, [https://www.software.ac.uk/blog/2016-09-12-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers](https://www.software.ac.uk/blog/2016-09-12-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers)
Science gateways are part of research software

- Science gateway frameworks and tools
  - Science gateway framework tool developers and maintainers are research software developers and maintainers
- Science gateways built on these frameworks and tools
  - Science gateway builders are research software developers and maintainers
- User-contributed elements (e.g. tools on HUBzero)
  - Science gateway teams support policies and practices that make others more likely to contribute elements to science gateways
Software sustainability

- The capacity of the software to endure

- The software will continue to be available in the future, on new platforms, meeting new needs

- The practices, both technical and non-technical, that allow software to continue to operate as expected in the future

Achieving software sustainability

- Software will collapse if not maintained
- Software bugs are found, new features are needed, new platforms arise
- Software development and maintenance is human-intensive
- Much software developed specifically for research, by researchers
- Researchers know their disciplines, but often not software best practices
- Researchers are not rewarded for software development and maintenance in academia
- Developers don’t match the diversity of overall society or of user communities
A rare success in the field

- Barriers prevent recognition and career success for research software
- An exception: Dr. Fernando Perez
- Traditional untenured research scientist (in neuroscience & computational research)
- On the side, co-developed IPython and Project Jupyter with a team
- Software had far more impact than traditional scientific outputs
- Evolved into Jupyter ecosystem
- High profile awards followed
- Now has fast-tracked tenured position at Berkeley
URSSI goals

• Conceptualize (plan) a US Research Software Sustainability Institute
• Cut across existing activities funded by NSF and beyond
• Directly and indirectly positively impact all software development and maintenance projects
• Focus on the entire research software ecosystem, including
  • People who create, maintain, and use research software
  • Research software
  • Research in research software
• Outputs:
  • Eager supportive & inclusive community
  • Concrete institute plan configured to offer valued services
  • Published survey and data that demonstrates community need
URSSI team (SGCI members in bold)

PIs
• Karthik Ram (PI, UC Berkeley)
• Jeffrey Carver (Alabama)
• **Sandra Gesing (Notre Dame)**
• Daniel S. Katz (Illinois)
• Nicholas Weber (Washington)

Advisory Committee
• Richard Arthur (GE Global)
• Michelle Barker (ARDC)
• Phil Bourne (Virginia)
• Daniel Crawford (MolSSI & Va Tech)
• Neil Chue Hong (SSI & Edinburgh)
• James Howison (Texas)
• Kurt Schwehr (Google)
• Jeff Spies (SHARE)
• **Nancy Wilkins-Diehr (retired)**

Senior Personnel
• Wolfgang Bangerth (Colorado State)
• Anshu Dubey (Argonne)
• Melissa Haendel (Oregon State)
• Mike Heroux (Sandia)
• Katy Huff (Illinois)
• **Suresh Marru (Indiana)**
• Kate Mueller (Notre Dame)
• Jarek Nabrzycki (Notre Dame)
• Kyle Niemeyer (Oregon State)
• **Marlon Pierce (Indiana)**
• Ariel Rokem (Washington)
• Arfon Smith (STScI)
• Tracy Teal (Carpentries)
• Matt Turk (Illinois)
• Rick Wagner (Argonne)
• **Mike Zentner (UCSD)**
NSF institutes context

Technologies
- MolSSI
- IRIS-HEP
- GSI (Conceptualization)
- CFDSI (Conceptualization)

Disciplines
- SGCI
- URSSI
URSSI activities

- Learn from community
  - Survey, workshops, ethnographic studies
- Iteratively build understanding of challenges
- Talk about what we’ve learned
  - Website (urssi.us), 33 blog posts (urssi.us/blog), articles, papers, talks
- Overall, build community
General Workshop & Unconference – Berkeley

• Key themes
  • **Software**: development, usability, and credit
  • **Training & workforce development**: Building upon existing models, designing new methods of instruction
  • **People**: Sustainability of RSEs as the human infrastructure of scientific software
  • **Projects**: Sustainability of software projects through organizational and institutional best practices

<table>
<thead>
<tr>
<th>Support Area</th>
<th>Supporting Software</th>
<th>Supporting People</th>
<th>Supporting the Community</th>
<th>Science &amp; research impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development Support (consulting &amp; short term small project support)</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Incubator (technology advice, business planning, usability advice, etc.)</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Training (courses &amp; guides)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Policy (research &amp; campaigns)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Community (fellowships, workshops, blogs, website)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
**General Workshop & Unconference – Chicago**

- Key themes
  - Community building
  - Career path and institutional support for RSEs
  - Training and workforce development
    - URSSI summer school, for grad students, return as TAs/mentors
    - Sustainability as related to reproducibility, usability, discoverability
  - Discussed URSSI mission and vision
- Strawhorse iteration
  - Added area: management, sustainability & governance
  - Development still unclear, scaling is problematic
  - Good ideas for other elements

Credit & Metrics Workshop – Santa Barbara

• Short talks about relevant projects
• Used Rumelt’s Strategy Kernel concept to plan potential URSSI work in this area
  1. Diagnosis (problems & challenges)
  2. Guiding principles
  3. Coherent actions
• Actions mostly in policy strawhorse area; some work done by fellows (community area)
• Many actions proposed, $3m/year to do everything, must carefully select
• Actions meant to support credit; must also work with other activities

R. Rumelt, Good Strategy Bad Strategy: The Difference and Why It Matters
Software Incubation Workshop – College Park

• How to support new and existing software projects that need to
  • Attract contributors & maintainers
  • Set up a governance model
  • Find and managing funding sources
  • Determine development practices

• Compared existing models of support and guidance
• Determined model of project phases; mapped existing efforts
• Found need to coordinate incubators and fill gaps


Participants
Apache Incubator
Code for Science & Society
eScience Institute Winter Incubators
ESIP Incubator
Mozilla Open Leaders
NumFOCUS
Software ecosystem researchers
Project stages

Schematic stages of open community for research software

Stage 0. Some code and a user of it. No sustained team.

Stage 1. Software development team, internal use.

Stage 2. Multiple software teams (different institutions) on same code (team is community), for internal use.

Stage 3. Self-governing developer community deliberately supporting broad user community.

Stage 4. Self-sustaining organization dedicated to supporting user and dev community (e.g. through commercial support, events, software foundation, etc.).

Proprietary commercialization path...

Stage 0. Some code.

Stage 1. Software development team, internal use.

Stage X. Project goes commercial, without developing an open community.

Draft survey findings, 1200 responses

- **Time**: Can’t allocate for software work as desired
- **Development process**: People-related aspects are harder than they should be, not technical aspects
- **Peer code review**: Less used than desired
- **Software development practices**: Not well-supported by tools
- **Version control**: Many still copying files to another place or zip file backups
- **Training**: Many have no training, mostly of insufficient time
- **Funding & support**: Nationally & locally, insufficient for software work
- **Performance reviews**: Don’t significantly value software contributions
- **Diversity plan**: Missing for most projects
Ethnographic studies

• Observations & semi-structured interviews over 12 months
• 3 projects of different types, in astronomy, bioengineering, hydrology
• Will produce case studies of how research software projects overcome challenges in
  • Recruiting contributors
  • Building a governance model
  • Seeking funding
  • Sharing credit
Final meeting - Chicago

- PIs, Senior Personnel, Advisory committee members, selected experts
- Design workshop
- Discussed previous work
  - Lessons learned from past workshops
  - Initial ethnographic work
  - Initial survey results
- Outlined URSSI implementation plan
- Discussed partnerships and strategies
- Proposed potential URSSI structure and activities
- Assigned initial workforce and budget to structure
URSSI plan
Mission & Vision

- PIs began to develop mission and vision statements in late 2018
- Mission: the purpose for URSSIs’s existence
- Vision: what URSSI strives to achieve
- Process
  - At first workshops asked about potential goals and vision
  - PIs found institutions with similar goals, compared mission & vision statements
  - Group facilitated by Neil Chue Hong developed drafts
  - Presented drafts at final URSSI meeting
  - Group & PIs refined final versions
Mission & Vision

• Mission
  • Our mission is to improve the recognition, development, and use of software for a more sustainable research enterprise
  • We achieve this mission through collaboratively developing education, outreach, and software services that emphasize open, transparent reproducible, and cooperative practices
  • URSSI is an institute for software expertise as well as a social infrastructure that promotes an inclusive and diverse community of research software engineers, maintainers, contributors, and users

• Vision
  • Empowered people, building better software, enabling exceptional research
Planning assumptions

- **Time**: 5 years + potential 5-year extension
- **Budget**: $3m-$5m per year
  - $3m is baseline & minimum needed
  - Build higher cost and higher return activities on that baseline
- **Activities**: set of reinforcing (but separable) activities, potentially on different timelines
  - Due to demonstrated interest in supporting URSSI goals from private foundations and potential interest from other federal agencies
- **Partners**: Improved software sustainability needed in all fields; URSSI can’t do everything
  - Many potential partners w/ overlapping goals
- **Indirection**: Can’t directly work with the US software development community
  - URSSI is small, community is large
  - URSSI need to work indirectly and leverage other groups
High-level methods

• Must choose initial targeted communities and efforts, focus on them for some period, then move on to next set

• Must work closely with partners to amplify its efforts

• Must be clever about using resources, attempt to achieve multiple outputs for activities whenever possible
Desired impact

• URSSI's ultimate desired impact
  • as stated in its vision
• is on scholarly research (science, engineering, etc.)
• It aims to achieve this impact by:
  • Improving research software
  • Empowering people
  • Making contributions to the overall research software ecosystem
Desired impact: software

Improve the sustainability of research software by

- Developing and sharing good-enough and best practices, including for testing, governance, codes of conduct, continuous integration
- Documenting various sustainability models and providing guidance on how to choose between them
- Promoting communities and appropriate governance models for open source research software
- Building models to make best use of transient developers (e.g., students)
Desired impact: people

Improve careers of research software developer & maintainers by

- Promoting new career paths
- Developing and promoting use of research software usage and impact metrics for hiring and promotion
- Promoting publication of research software
- Developing and providing training
- Encouraging a diverse set of participants to enter the field; decrease structural and systemic barriers to productive careers for members of underrepresented groups
Desired impact: research software ecosystem

Improve understanding and functioning of the ecosystem by
• Documenting use of research software in research
• Growing and participating in research software communities
• Promoting increased understanding of importance of research software in research
• Promoting software credit and citation mechanisms
• Studying methods to catalog and promote research software
• Supporting the use of software in reproducibility
### In-progress URSSI Structure

**5 years, 5 themes**

<table>
<thead>
<tr>
<th><strong>Incubator</strong></th>
<th><strong>Training</strong></th>
<th><strong>Community</strong></th>
<th><strong>Policy</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Providing technology, business planning, governance/licensing, and usability advice so that project ideas germinate into small projects, and small projects grow into viable sustainable projects.</td>
<td>Developing quick start guides, training modules, summer school</td>
<td>Supporting diverse fellowships &amp; champions, and putting on workshops and an annual conference</td>
<td>Research &amp; analysis, gathering data and understand policy needs and potential changes; Advocacy, working to make changes occur</td>
</tr>
</tbody>
</table>

**Management, Governance, Sustainability**

Driving URSSI, monitoring and acting on metrics and risks, bringing in community views and ideas, planning for future support.
Example URSSI flow

5 years, 5 themes

- Incubator
  - Project planning

- Training
  - Trained staff & contributors

- Community
  - Collaborators & colleagues

- Policy
  - Supportive institutional and funding culture

Management, Governance, Sustainability

New Software Project
Example URSSI flow

5 years, 5 themes

- Incubator
  - New topics
- Training
  - Course and modules
- Community
  - Fellowship & workshop topics
- Policy
  - New studies, advocacy plans

Management, Governance, Sustainability

Community needs
Example URSSI flow

5 years, 5 themes

- **Incubator**: Develop material
- **Training**: Develop materials, customize for communities
- **Community**: Outreach & engagement
- **Policy**: Perform analysis, Support advocacy

**Fellows**

**Management, Governance, Sustainability**
Sustainability

- Funding from other agencies & private foundations
  - Sloan expressed strong interest in URSSI, potentially could fund some activities
  - Other private foundations work in this space, have expressed interest in this type of activity, e.g., CZI, Schmidt
    - Sloan is willing to bring these foundations together and discuss how URSSI might be supported
  - DOE's related project (IDEAS-Productivity) will be ending around 2021
    - An opportunity for DOE & NSF to work together?
- Institutional support (from our institutions) & membership fees from partner institutions (including industry)
- Partners who will take on part of our mission over time (e.g., Carpentries in training)
- Services with cost recovery, perhaps that spin-off
In 5 years … More science & engineering

Better research software skills → better research software

Better research software culture → better career paths for developers & maintainers

   Better career paths → willingness to share software & join software communities

       Better research software → researchers focus on their research
Next steps

- Complete written draft plan by Dec 2019
- Submit to NSF & make public
- Run pilot summer school in Dec 2019 (URSSI “winter school”)
- Gather feedback from NSF and public on draft plan
- Gather lessons from winter school
- Complete final plan by April 2020
- Hope for NSF to offer chance to propose, and work with other potential funders
How to get involved in URSSI

• Watch the web page: http://urssi.us
• Join the mailing list – via form on the URSSI web page
• Suggest and discuss topics: http://discuss.urssi.us
• Follow on twitter: https://twitter.com/si2urssi
• Write a blog, or suggest one we should cross-post
• Use GitHub: http://github.com/si2-urssi/
  • Repos for the web site (PRs to add) and workshops
• Talk to others about URSSI (including NSF)
• If you have questions, want to suggest something, want to volunteer, email us: contact@urssi.us