

# Toward a Framework for Evaluating Software Success: A Proposed First Step

Stan Ahalt<sup>1</sup>, Bruce Berriman<sup>2</sup>, Maxine Brown<sup>3</sup>, Jeffrey Carver<sup>4</sup>, Neil Chue Hong<sup>5</sup>, Allison Fish<sup>6</sup>, Ray Idaszak<sup>1</sup>, Greg Newman<sup>7</sup>, Dhabaleswar Panda<sup>8</sup>, Abani Patra<sup>9</sup>, Elbridge Gerry Puckett<sup>6</sup>, Chris Roland<sup>10</sup>, Douglas Thain<sup>11</sup>, Selcuk Uluagac<sup>12</sup>, Bo Zhang<sup>13</sup>



Funding, in part:  
1521388

## Abstract

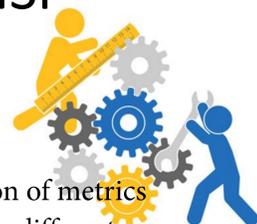
Software is a particularly critical technology in many computational science and engineering (CSE) sectors. Consequently, software is increasingly becoming an important component in the evaluation of competitive grants and the execution of research projects. As a result, software can be viewed as a scholarly contribution and has been proposed as a new factor to consider in tenure and promotion processes. However, existing metrics for evaluating the capability, use, reusability, or success of software are sorely lacking. This lack of software metrics permits the development of software based on poor development practices, which in turn allows poorly written software to “fly under the radar” in the scientific community and persist undetected. The absence of evaluation by knowledgeable peers often leads to the establishment and adoption of tools based on aggressive promotion by developers, ease-of-use, and other peripheral factors, hindering the sustainability, usefulness, and uptake of software and even leading to unreliable scientific findings. All of these factors mean that addressing the current lack of software evaluation metrics and methods is not just a question of increasing scientific productivity, but also a matter of preventing poor science.

As a first step toward creating a methodology and framework for developing and evolving software success metrics for the CSE community, we propose the creation of a software “peer-review group.” This group, comprised of grant recipients funded to develop sustainable software, would meet periodically to evaluate their own and each others’ software, developing and refining success metrics along the way. We envision the group as a pilot test for a potential larger-scale effort to establish a more formal framework for software success metrics and evaluation.

White paper corresponding to this poster: <http://dx.doi.org/10.6084/m9.figshare.1561451>

## Background

Feb 17-18, 2015 NSF  
SI2 PI workshop  
breakout team:

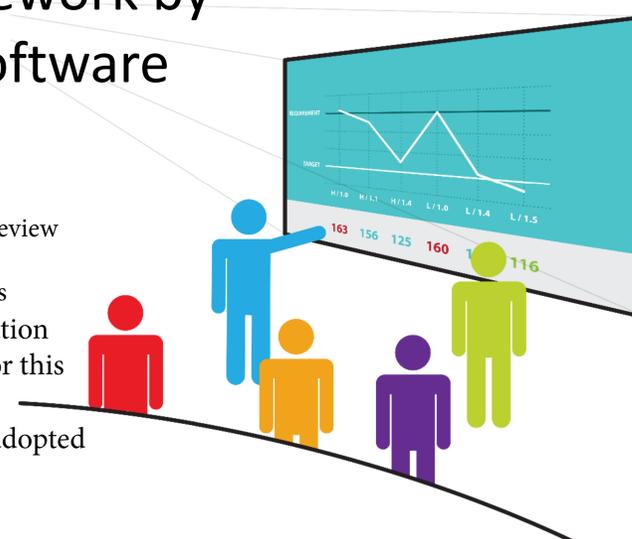


- Discussed framework for the creation of metrics
  - Define software success factors across different research domains
- Improve both the software and the metrics (i.e. “eat their own dogfood”)
  - Stakeholders put up their own software for evaluation within the network
- Build Community
  - Improve the cross-talk among different research domains of what makes software successful

## Developing Metrics while Developing Community

Build an open framework by  
which to develop software  
success metrics

- Software “peer review group”
  - Representative stakeholders who will self-review software works created by their respective communities; concurrently develop metrics
- Community-governed; no single institution oversees the activity or infrastructure for this
  - No central “group of experts”!
- Evolve into community generated and adopted standards



## Framework

Tie generation of  
metrics to actual  
evaluation of  
software



- Incent through improving their own software; benchmarking; publishing opportunities
  - Volunteer, opt-in
- Infrastructure support
  - Code, code review; in contributors infrastructure
  - Social: forums for generation of software success metrics tied to code review
  - Can metrics be fit into a common template

## Further Questions

- How will peer-review of code and corresponding peer-review of software success metrics be performed and tied together across different research domains?
  - What type of infrastructure will best facilitate this?
- How will ever-increasing open participation by the community be nurtured and measured?
- What governance structure will be put in place to ensure shared community governance of the software and metric peer-review process?
- How will factors of software success be evaluated: e.g. criticality, value, sustainability, usability, reusability, performance, functionality, capability, availability, scientific impact, usefulness, reliability?



## Next Steps

- Use CSESSP, WSSSPE, other future events to build this community and framework & author more white papers
  - Continually take advantage of these venues
  - Collaborate with other metric & framework projects
- Consider proposal around this idea to give enough runway to build community & metrics
  - Encouraging the community to take up this idea and run with it

<sup>1</sup>RENCI, University of North Carolina at Chapel Hill; <sup>2</sup>California Institute of Technology; <sup>3</sup>University of Illinois at Chicago; <sup>4</sup>University of Alabama; <sup>5</sup>Software Sustainability Institute, University of Edinburgh; <sup>6</sup>University of California, Davis; <sup>7</sup>Colorado State University; <sup>8</sup>The Ohio State University; <sup>9</sup>The State University of New York, University at Buffalo; <sup>10</sup>North Carolina State University; <sup>11</sup>University of Notre Dame; <sup>12</sup>Florida International University; <sup>13</sup>Indiana University