Research Software Engineers: Establishing Careers

Ian A. Cosden, Ph.D.
Director, Research Software Engineering for Computational & Data Science
Princeton University

Chair, Steering Committee
US-RSE Association

MAGIC Meeting
July 7, 2021
Outline

• Local example
  • Establishing Princeton’s Research Software Engineering Group

• Building a national community
  • US-RSE: The US Research Software Engineer Association

• Building a Pipeline
  • INTERSECT Project: Research Software Engineering Training
Research Software Use

• “Research software”
  • Anything used to generate, process, or analyze results you intend to appear in a publication
  • Anything from a few lines of code written by you to a professionally developed software package

Researchers using research software

<table>
<thead>
<tr>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>92%</td>
<td>95%</td>
</tr>
</tbody>
</table>

Fundamental to their research

<table>
<thead>
<tr>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>69%</td>
<td>66%</td>
</tr>
</tbody>
</table>

Some of the problems

• Writing good code requires experience/training
• Writing good code is hard and time consuming
• Domain scientists/researchers typically aren’t rewarded for good code
  • Publications result from new insight or discovery
• Graduate students and post-docs are frequently primary developers
  • Often tasked with making software design decisions as a novice programmer
  • “They aren’t even aware of the damage they have done.”
• Grad students and post-docs leave

This article has been retracted: please see Elsevier Policy on Article Withdrawal (http://www.elsevier.com/locate/withdrawalpolicy).

In our Correspondence, we reported evidence leading us to conclude that pigeons are on par with humans when tested with a behavioral task that demands simultaneous processing resources; in particular, we claimed that pigeons show faster responses than humans when sub-tasks are separated with a short STOP-CHANGE delay of 300 ms—the “SCD 300” condition (time advantage of 200 ms). We have subsequently discovered, however, that the MATLAB script that was used for the analysis of reaction times in the pigeon paradigm was wrongly indexed.
Reproducibility

• Ad-hoc software can make reproducibility hard

You can download our code from the URL supplied. Good luck downloading the only postdoc who can get it to run, though #overlyhonestmethods
One solution: RSE

1. Establish a position dedicated to research software
   • A (new) career path in the research community
   • *Research Software Engineer*

   • Role isn’t really new
     • Title, formality, and awareness *are*

   • Princeton RSE group
     • Formed in late 2016
     • Centrally located group of software experts with permanent positions
     • 2019 expanded to include other fully-funded departmental RSEs
• **Goal:** Help researchers create the most efficient, scalable, and sustainable research code possible in order to enable new scientific advances.

• Complement traditional academic research groups with embedded, long-term:
  • Software development
  • Domain specific knowledge
  • Algorithm development and selection
  • Performance tuning & optimization
  • Coding standards and techniques
What is a Princeton RSE?

1. Software Engineer/Developer
   - Design, develop, refactor
   - Build tests, documentation, etc.

2. Computational Researcher/Scientist
   - Domain expertise
   - Implement algorithms in code
   - Extract science from software

3. Traditional Research IT Support
   - Get novice researchers started
   - Answer help tickets
   - Solve error messages, installation, etc.

Diagram:
- Professional Software Engineer/Developer
- Computational Researcher/Scientist
- Traditional Research IT Support
Current RSE Partnerships

- Princeton Neuroscience Institute
- High-Energy Physics (IRIS-HEP)
- Center for Statistics and Machine Learning
- Applied & Computational Math
- Genomics
- Molecular Biology
- Civil & Environmental Engineering
- Operations Research & Financial Engineering
- Data Driven Social Science

Vacant
RSEs at Princeton

RSE Advantages:

- Combining research/domain knowledge with software engineering best practices
- Primarily focused (and evaluated) on software contributions not research output

Performance   Functionality   Reproducibility   Usability   Reusability

“Just wanted to drop a line and say that I followed the instructions for installing the CPU only version on one of my machines, and it did everything smoothly… for the first time in my life…!”

-- Researcher to Troy Comi, RSE
RSEs at Princeton

RSE Advantages:

• Combining research/domain knowledge with software engineering best practices
• Primarily focused (and evaluated) on software contributions not research output
• Institutional knowledge
• Mentor and leaders to novice developers

“Dave has transformed the way we do research in the building.”

RSEs at Princeton

RSE Advantages:
- Combining research/domain knowledge with software engineering best practices
- Primarily focused (and evaluated) on software contributions not research output
- Institutional knowledge
- Mentor and leaders to novice developers
- Educate (training, workshops, etc.)

RSE Challenges:
- Funding
- Expansion
- Career paths & classification
Outline

- Local example
  - Establishing Princeton’s Research Software Engineering Group

- Building a national community
  - US-RSE: The US Research Software Engineer Association

- Building a Pipeline
  - INTERSECT Project: Research Software Engineering Training
A Brief History of the Research Software Engineer

- Pre-2012

194 different job titles

Credit: Simon Hettrick
A Brief History of the Research Software Engineer

- Movement and Term: Born in the UK
  - 2012 SSI’s Collaborations Workshop - “Research Software Engineer”
  - Late 2013 UKRSE Association forms with ~50 members
RSE Associations Around the World

Photo by Antonia Cozcu, Jan Philipp Dietrich, de-RSE e.V. (CC BY 4.0).
Introducing the US-RSE

A community-driven organization

People:

- Writing and contributing research software at
  - Universities, laboratories, knowledge institutes, companies, & more
- Interested in research software engineer careers
  - Students, researchers, software engineers
- Identifying as RSE “allies”
  - Manage, sponsor, support

https://us-rse.org
Mission of the US-RSE Association

1. **Community**
   ○ Create a professional community to share knowledge, connections, and resources

2. **Advocacy**
   ○ Promote RSEs impact on research, highlighting the critical and valuable role RSEs serve

3. **Resources**
   ○ Access to information and material to support individuals and RSE groups

https://us-rse.org
Growth of the US-RSE

- Founded in early 2018
- Mid-2019 – now: significant growth

https://us-rse.org
Where are our members located?

https://us-rse.org/usrse-map
Where are our members located - US?

https://us-rse.org/usrse-map
US-RSE Resources

- Website
- Newsletters
- Twitter
- Syndicated blog
- RSE stories Podcast

US-RSE Newsletter
March 2021 Newsletter
Posted on March 22, 2021

In this bi-monthly newsletter, we share recent, current, and planned activities of the US-RSE Association, and related news that we think is of interest to US-RSE members. Newsletters are also available on our website alongside the growing resources and information on the US-RSE Association. A sign-up option for our newsletter is available here.

https://us-rse.org
US-RSE Community

- Most communication happens in slack
  - Get invite after joining
- Working Groups
- Events
  - Conference Panels/Workshops/Meetups
  - US-RSE Virtual Workshops
  - Monthly Community Calls

https://us-rse.org
RSE Job Board

- RSE job board and #jobs channel
  - 99 unique jobs posted all time
  - 9 as of yesterday

https://us-rse.org/jobs

https://us-rse.org
RSE Job Board

- RSE job board and #jobs channel
  - 99 unique jobs posted all time
  - 9 as of yesterday
- Second most viewed page on the site
- Success!

Valeria
5:33 PM
Tuesday, October 6th

Hi all!

I just want to write something as appreciation to this group. I feel like I am still new to the RSE world but I just want to thank everybody involved in this project. I found my very first job as an RSE exactly through this Slack space and this week it's my first week 🎉!
US-RSE Future Plans

- Raise awareness of US-RSE & RSEs
- Community engagement and development
- International RSE Survey
  - Help us better understand RSEs in the world and US
  - We need your help! Complete the survey & share with colleagues
- Annual RSE Conference and/or workshop(s)
- Regional and local events/chapters
- Advocacy program to promote RSEs
- Career development program

https://us-rse.org

#US-RSECon2023
Join us!

Free signup for newsletter and slack.

https://us-rse.org/join

Get involved & Lurkers welcome!

https://us-rse.org

contact@us-rse.org

@us_rse
Outline

- Local example
  - Establishing Princeton’s Research Software Engineering Group
- Building a national community
  - US-RSE: The US Research Software Engineer Association
- Building a Pipeline
  - INTERSECT Project: Research Software Engineering Training
How do I Become a Research Software Engineer?

Villanova University - MS in Software Engineering - villanova.edu
Gain the knowledge and experience to design and develop today's software systems. Take the next step with a Villanova Master's Degree. Learn more! Relevant. Flexible. Rigorous. Courses: Systems and Networks, Programming, Immersive Technology.

Become a developer - Learn to code in 15 weeks
Learn to think & build like a software engineer. Develop coding mastery. Apply today! The world changes, we all have to change too. Our school teaches the skills of the future.

Philadelphia Coding Bootcamp - Take the Free Aptitude Test
We think you've got what it takes. Try our free coding aptitude test for an instant score. Named a Best Coding Bootcamp by SwitchUp. Read why our students give us a 5 star...

DeVry University Online - Engineering Technology - DeVry.edu
Learn About Technology, Earn A Degree In Programming, Engineering, or Computer Networking.
INTERSECT
Research Software Engineering Training

https://intersect-training.github.io

Ian A. Cosden
Director, Research Software Engineering for Computational & Data Science
Princeton University

Jeffrey C. Carver
Professor of Computer Science
University of Alabama
Introducing: INTERSECT

- **Innovative Training Enabled by a Research Software Engineering Community of Trainers**
  - 3-year NSF Funded CyberTraining project

- **3 main goals:**
  - Develop an open-source modular training framework conducive to community contribution
  - Deliver RSE-led research software engineering training targeting developers
  - Deepen the connection within the national community of Research Software Engineers

- **Sponsor two annual events:**
  - Workshops for RSEs to develop/refine material
  - RSE-led bootcamp for intermediate/advanced research software developers

https://intersect-training.github.io
Anticipated Curriculum

Details to be established in workshops and collaboratively:

- Software Design
- Testing
- Collaboration Techniques
- Software Packaging and Distribution
- Documentation Techniques
- Performance and Optimization

Many such topics are already covered in local workshops, blog posts, and elsewhere.

https://intersect-training.github.io
COLLECTING RSE TRAINING TOWARDS THE GOAL OF A UNIFIED CURRICULUM

Existing Research Software Engineering Training Material

Below is an ongoing collection of RSE training material, workshops, and resources. We are compiling this list as a starting point for future activities. For now, these are all provided “as is” to the community. They haven’t been reviewed for content or accuracy. They are presented in no particular order.

We eagerly welcome additions to this list. Links to slides, videos, notebooks etc. that fall loosely under the Research Software Engineering umbrella are all encouraged.

We are especially seeking material that goes beyond basic research computing competency (i.e. what The Carpentries does so well) and is general enough to span multiple domains. Specific tools and technologies used only in one domain, or applicable to only one subset of computing (i.e. HPC) are typically too narrowly focused. When in doubt, submit it to be included or reach out and we’d be happy to discuss.

See below for instructions on how to add to this list.

Links to RSE Training Material by Topic

https://intersect-training.github.io/training-links
Participate in INTERSECT

- **Workshops** - If you consider yourself an RSE and
  - you have experience giving research software development training
  - you have never formally served as a trainer
  - you have no interest in being a trainer or giving a workshop, but would like to give your input on the material and skills that are needed in the profession
- **Bootcamps** - If you have a baseline knowledge and are looking to expand your research software engineering skill set
- **Asynchronous Training Material** - If you want to contribute to the mission of curating and developing specific training topics

https://intersect-training.github.io/paticipate

https://intersect-training.github.io
Thank you!

- https://www.princeton.edu/researchcomputing
- https://us-rse.org
- https://intersect-training.github.io

Questions, comments, discussions:
- icosden@princeton.edu
- @IanCosden
"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov