



Princeton
Research Computing

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



UK¹



US²

Researchers using research software

92 %

95 %

Fundamental to their research

69 %

66 %

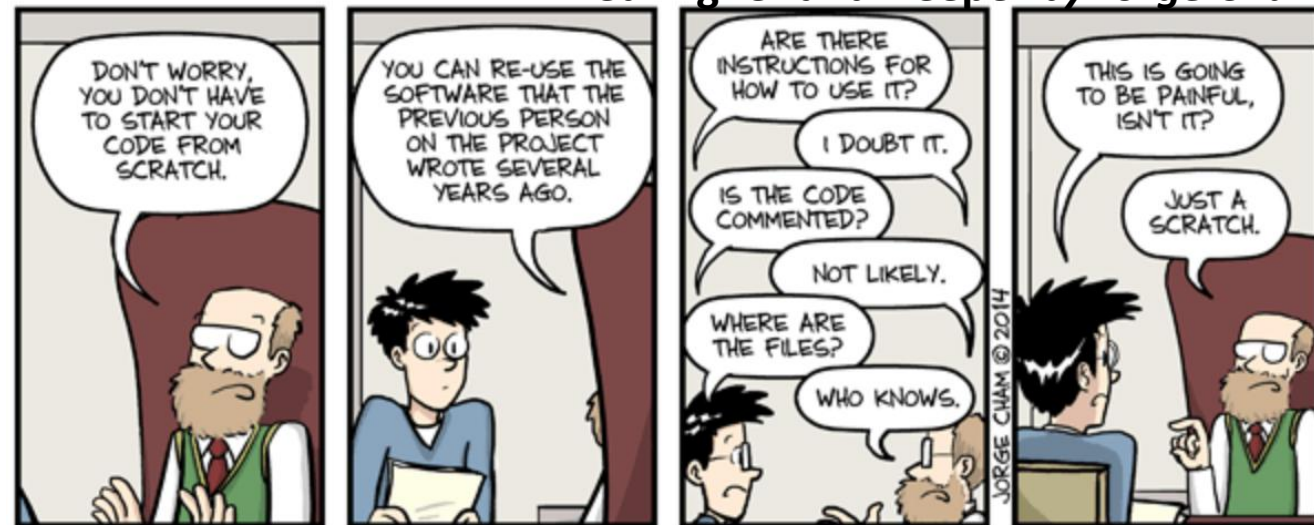
1. S.J. Hettrick et al, UK Research Software Survey 2014, DOI:10.5281/zenodo.1183562

2. U. Nangia and D. S. Katz. Surveying the US National Postdoctoral Association Regarding Software Use and Training in Research. (WSSSPE5.1), 2017.

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*

Piled Higher and Deeper by Jorge Cham



Science Depends on Software Correctness

SCIENTIFIC PUBLISHING

A Scientist's Problem

Until recently, a trajectory of a protein crystallography at the prestigious San Diego, California, ceremony at the Presidential Science Conference for Scientists from the country's highest researchers. A stream of his detailing the of important cell membrane.

Then the nightmare. Researchers at *Nature* that a protein structure had described in a paper. When Chang was [redacted] that a homogram had flipped data, inverted map from a derived the function. Unfortunately, the program other proteins

Characterization
Cyanine
"Will"

Jayanti
and Phil

Department

Sup

ABSTRACT
led to
along with
nuclear
configu
on the
scripts
unrecog

Current Biology

CORRESPONDENCE | VOLUME 27, ISSUE 18, PR996-R998, SEPTEMBER 25, 2017

RETRACTED: How birds outperform humans in multi-component behavior

Sara Letzner • Onur Güntürk • Christian Beste

Open Archive • DOI: <https://doi.org/10.1016/j.cub.2017.07.056> • Check for updates

Article Info

Linked Article

Comments

This article has been retracted: please see Elsevier Policy on Article Withdrawal

(<http://www.elsevier.com/locate/withdrawalpolicy>).

This article has been retracted at the request of the authors.

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.

...ne of code sinks

...om the
ne
Shifts

...ted States



Nature Protocols

72.4 (Incorrect)

73.2
73.2
Different
Calculated
Chemical
Shifts!

72.7 (Incorrect)

JCO), explained
retraction or a cor-

Reproducibility

- Ad-hoc software can make reproducibility *hard*



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

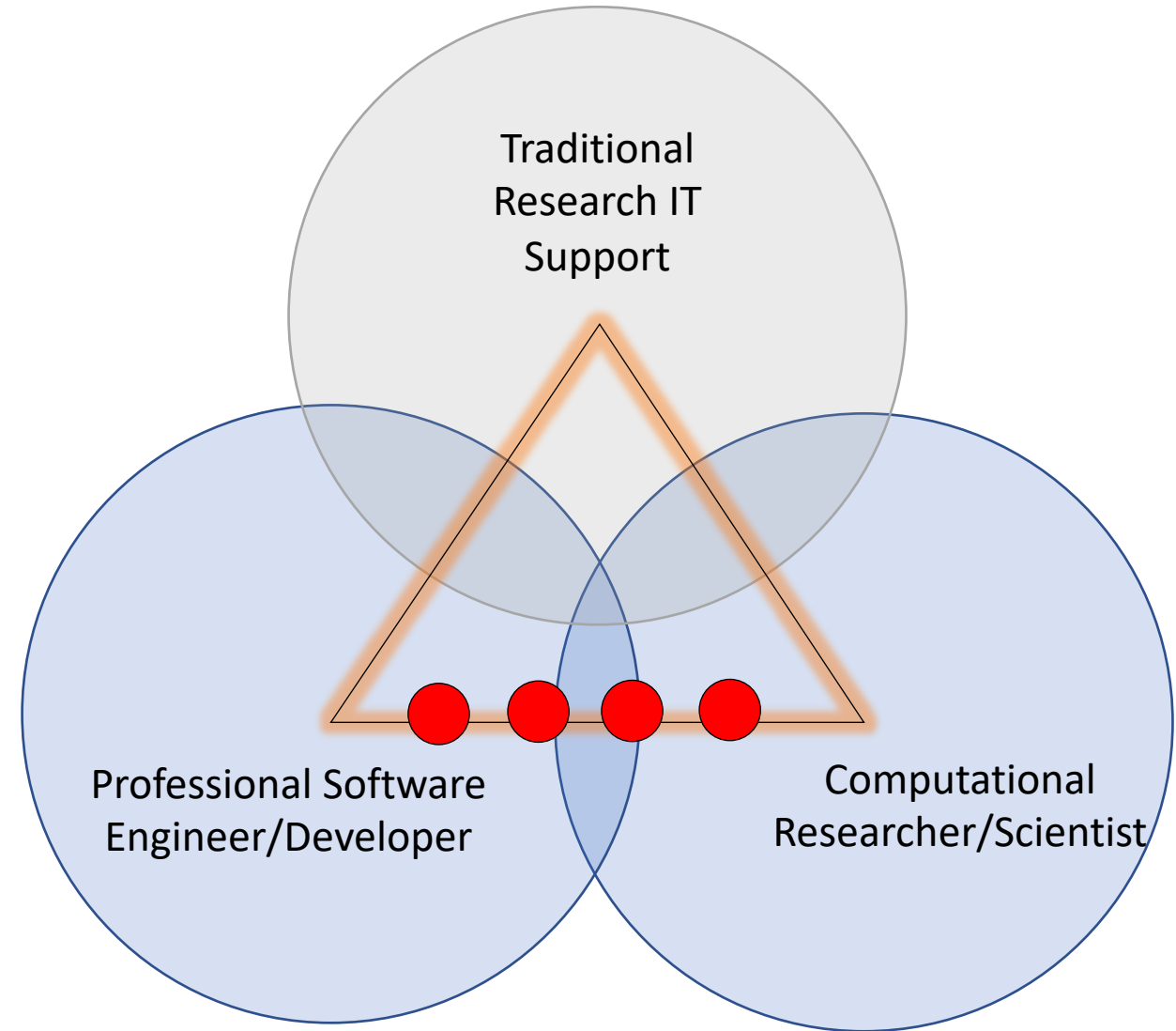
Princeton RSE Group

- 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

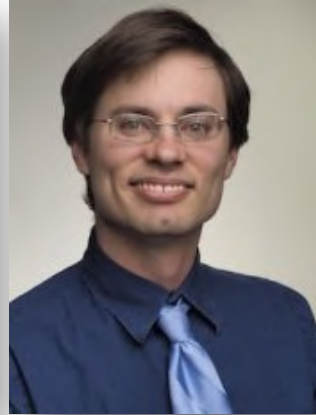


Current RSE Partnerships

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



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.”

-- Jonathan Cohen, Robert Bendheim and Lynn Bendheim Thoman Professor in Neuroscience. Professor of Psychology and the Princeton Neuroscience Institute. Co-Director, Princeton Neuroscience Institute.

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



PRINCETON UNIVERSITY

SPRING 2021 VIRTUAL MINI-COURSES AND WORKSHOPS

researchcomputing.princeton.edu/workshops

PICSciE and OIT Research Computing, with the participation of the University Library and Data & Statistical Services (DSS), are offering instructor-led virtual trainings this Spring, including:

INTRO TO THE LINUX COMMAND LINE	LEVEL-UP YOUR MATLAB
INTRO TO DATA ANALYSIS USING R	CONTINUOUS INTEGRATION & AUTOMATED TESTING WITH GIT AND JENKINS
REPRODUCIBLE HPC RESEARCH WITH CONTAINERS: DOCKER AND SINGULARITY	A PRIMER ON AMAZON WEB SERVICES (AWS)
INTRO TO DATA ANALYSIS USING PYTHON	MANAGING PYTHON WITH CONDA ON THE HPC CLUSTERS
GETTING STARTED WITH THE RESEARCH COMPUTING CLUSTERS	INTRODUCTION TO THE MACHINE LEARNING LIBRARIES
REMOVING THE TEDIUM FROM YOUR RESEARCH WORKFLOW	GEOGRAPHIC INFORMATION SYSTEMS (GIS)
INTRO TO VERSION CONTROL USING GIT	

Workshops on QGIS and ArcGIS Pro:
<https://library.princeton.edu/collections/pumagic/workshops>

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

Casebooks Project Editor (Research Assistant / Associate) Climate Researcher (Research Associate) Clinical Study Programmer CoMPLEX Research Associate Computational Biologist / Bioinformatician Computational Scientist Computational Scientist in Computational Fluid Dynamics & Industrial Applications Computational Scientist in Structural Mechanics and Industrial Applications Computer Scientist Computer Vision Researcher Content Developer / Programmer Control Engineer-IMG - 3 posts CREATE Data Specialist Data Analyst Data Integration Coordinator Data Manager x3 Database and Software Engineer Database Manager / Researcher Database Programmer Digital Media Technician E-Learning Portal Manager (KTP Associate) e-Learning Systems Development Analyst e-Learning Systems Development Analyst (Moodle, SQL) E-Learning Web Developer E-Portfolio Learning Technologist Embedded Systems Engineer Engineering Technician Environmental Scientist EPSRC Studentship on Algorithmic Construction of Finsler-Lyapunov Functions Experimental Officer in Bioinformatics Experimental Psychologist Finance Assistant Gaia Alerts Software Developer Gaia Software Developer (Gaia UK Team) GIS Applications Specialist Graduate Programmer / Software Developer Graphics Programmer Health Data Manager / Scientist High Throughput Bioinformatician High Throughput Sequencing Bioinformatician / Two posts HIVE Manager / Co-ordinator HIVE Senior Researcher and Technical Lead Hydroinformatics Scientific Software Developer Hydroinformatics / Hydroinformatics / Hydroinformatics Systems Developer Instrumentation Engineer Investigator Statistician IT Developer / IT Support Technician (Unix / Windows Systems) Knowledge Transfer Partnership (KTP) Associate / Knowledge Transfer Partnerships (KTP) Associate - Software Developer KTP Associate - Robot Vision Scientist (Research Fellow) KTP Associate (Fixed Term Contract for 24 months) KTP Associate (Precision Agriculture Data Analyst) KTP Associate @t* Graduate Web Developer KTP Associate: Electronics / Robotics Engineer Learning Technologist Leicester Respiratory BRIT Developer Linux / Perl Biologist Maker Space Technician Marie Curie Early Stage Researcher Marie Curie Early Stage Researcher in the field of Earth and Planetary Science Modelling Marine Earth Observation Scientists Medical Statistician Medical Statistician / Senior Biostatistician Mobile Application Developer NASC IT Support - Programmer and Systems Administrator (Fixed Term Research Assistant) NDRA on EU Project on Automated Multisensor Surveillance Planning Officer Policy Modeller 2017 Post - Doctoral Research Assistant INSTRON Post Doctoral Research Worker Post Doctoral Researcher in the application of Digital Technology Post-Doctoral Research Assistant in Simulation and Visualization Post-Doctoral Research Associate Post-Doctoral Research Associate (Pathogen Genomics) Post-Doctoral Research Fellow Postdoctoral Fellow - population genetics / evolutionary genetic Postdoctoral Fellow in Bioinformatics Postdoctoral Fellow in Cancer Therapeutics Postdoctoral Research Assistant Postdoctoral Research Associate Postdoctoral Research Fellow Postdoctoral Research Scientist Postdoctoral Researcher in Declarative (Logic and Functional) Programming Postdoctoral Researcher Postdoctoral Scientist Postdoctoral statistician Postdoctoral Training Fellow - Statistical and Computational Genetics of Autism Principal / Senior Bioinformatician Principal Bioinformatician Product Development Engineer (Rail) Publishing Portal Web Developer Radio Frequency Engineer Reader in Computer Science Reporting Analyst Research (Software) Engineer Research Assistant Research Associate Research Fellow Research Image Data Manager, Biomedical Engineering Research Officer Research Officer @t* Social Protection Research postgraduate Research Programmer Research Scientist Research Scientist / Senior Research Scientist Research Scientist in Machine Learning and Computer Vision Research Software Developer Research Software Developer for the Herchel Smith Professor of Organic Chemistry Research Software Engineer Research Studentship Research Worker Researcher SAP Trainee Technical Analyst Scientific Officer with Michela Garofalo Scientist SEAH Studentship: Extracting epidemiological data from collections SEEG Data Archive Manager Senior / Research Associate in Clinical Integration and Image Analysis for Fetal Surgery Senior Analyst Programmer (Business Analysis) Senior Analyst / Programmer Senior Bioinformatician Senior Bioinformatician / Bioinformatician Senior Computational Statistician - Spatial Models Senior Data Acquisition Scientist / Data Acquisition Scientist Senior Data Manager Senior Database Administrator Senior IT Developer Analyst Senior Mathematical Modeller Senior Media Developer Senior Postdoctoral Researcher - Evolutionary and Computational Analysis of Infectious Disease (Phylogenetics) Senior Research Assistant Senior Research Associate Senior Research Associate @t* Molecular Modelling & Simulation Senior Research Associate in Quantitative Clinical MRI Senior Research Fellow Senior Research Fellow / Research Fellow in Vibration Diagnostics and Prognostics / Digital Signal Processing Senior Research Laboratory Technician Senior Research Technician Senior Software Developer in Bioinformatics Senior Software Engineer - Software Engineer Senior Statistical Epidemiologist Senior Systems

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 Cozacu, 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



Mission of the US-RSE Association

1. **Community**

- Create a professional community to share knowledge, connections, and resources

2. **Advocacy**

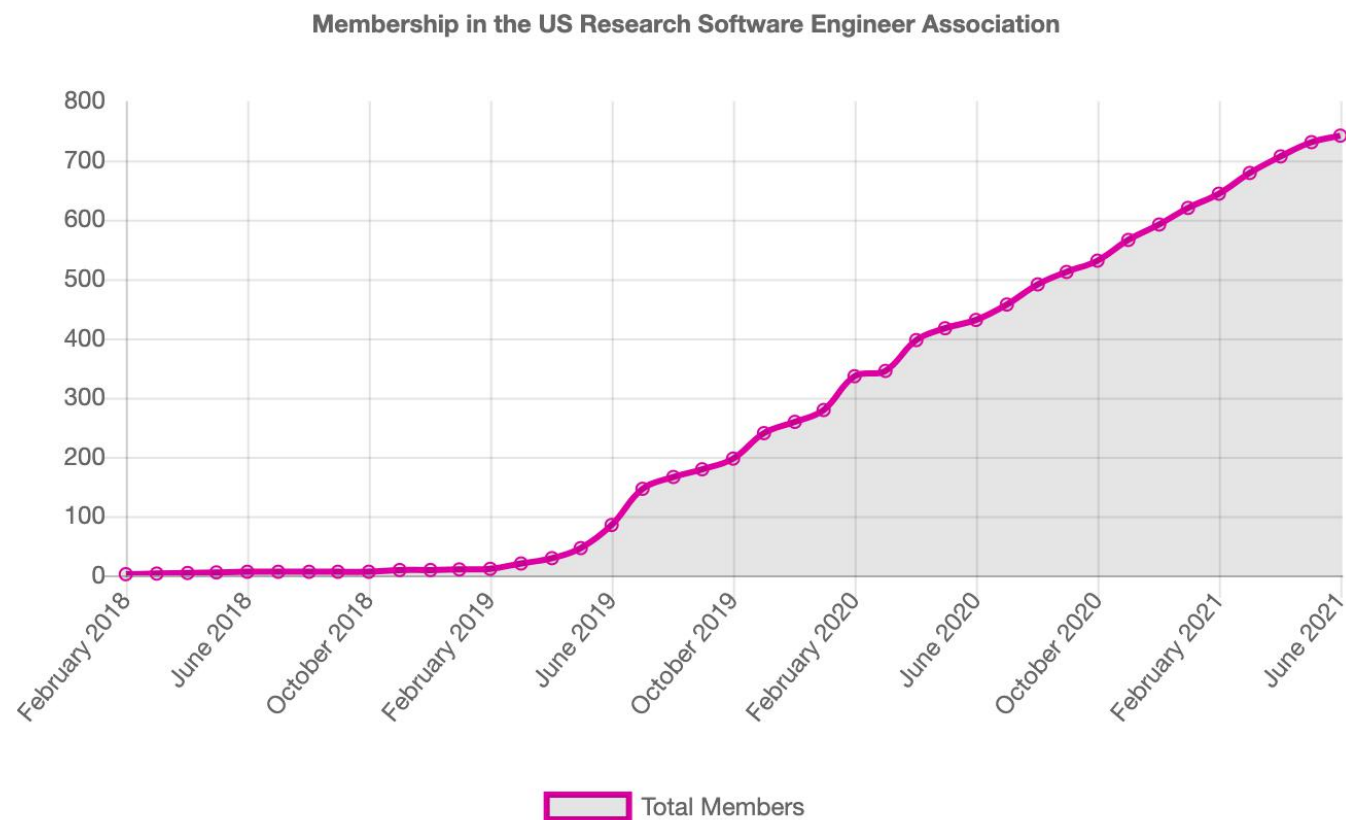
- Promote RSEs impact on research, highlighting the critical and valuable role RSEs serve

3. **Resources**

- Access to information and material to support individuals and RSE groups

Growth of the US-RSE

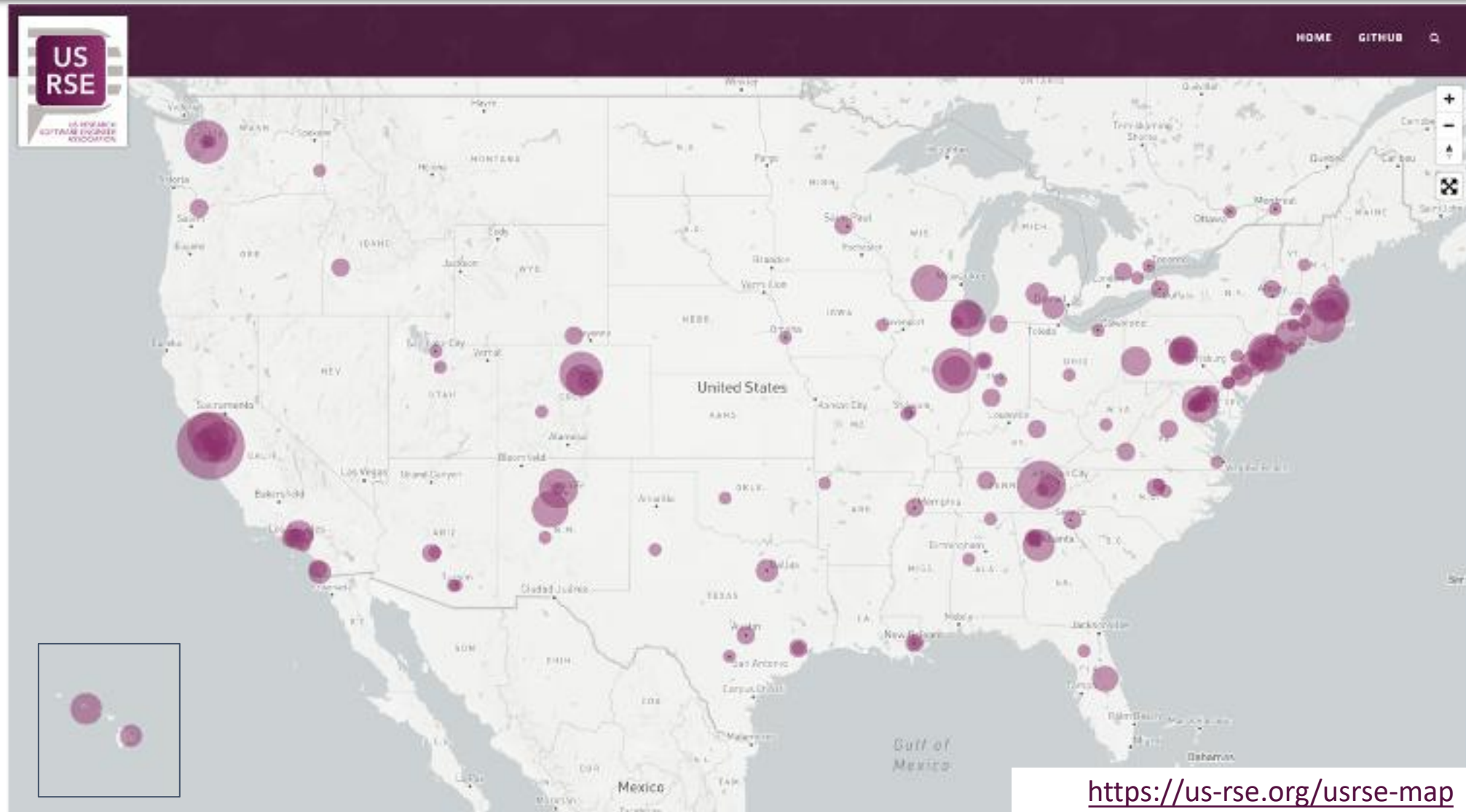
- Founded in early 2018
- Mid-2019 – now: significant growth



Where are our members located?

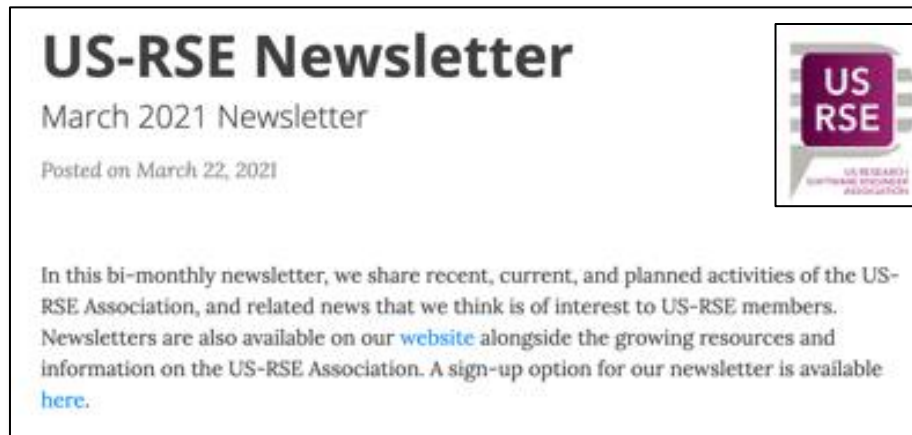


Where are our members located - US?



US-RSE Resources

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



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

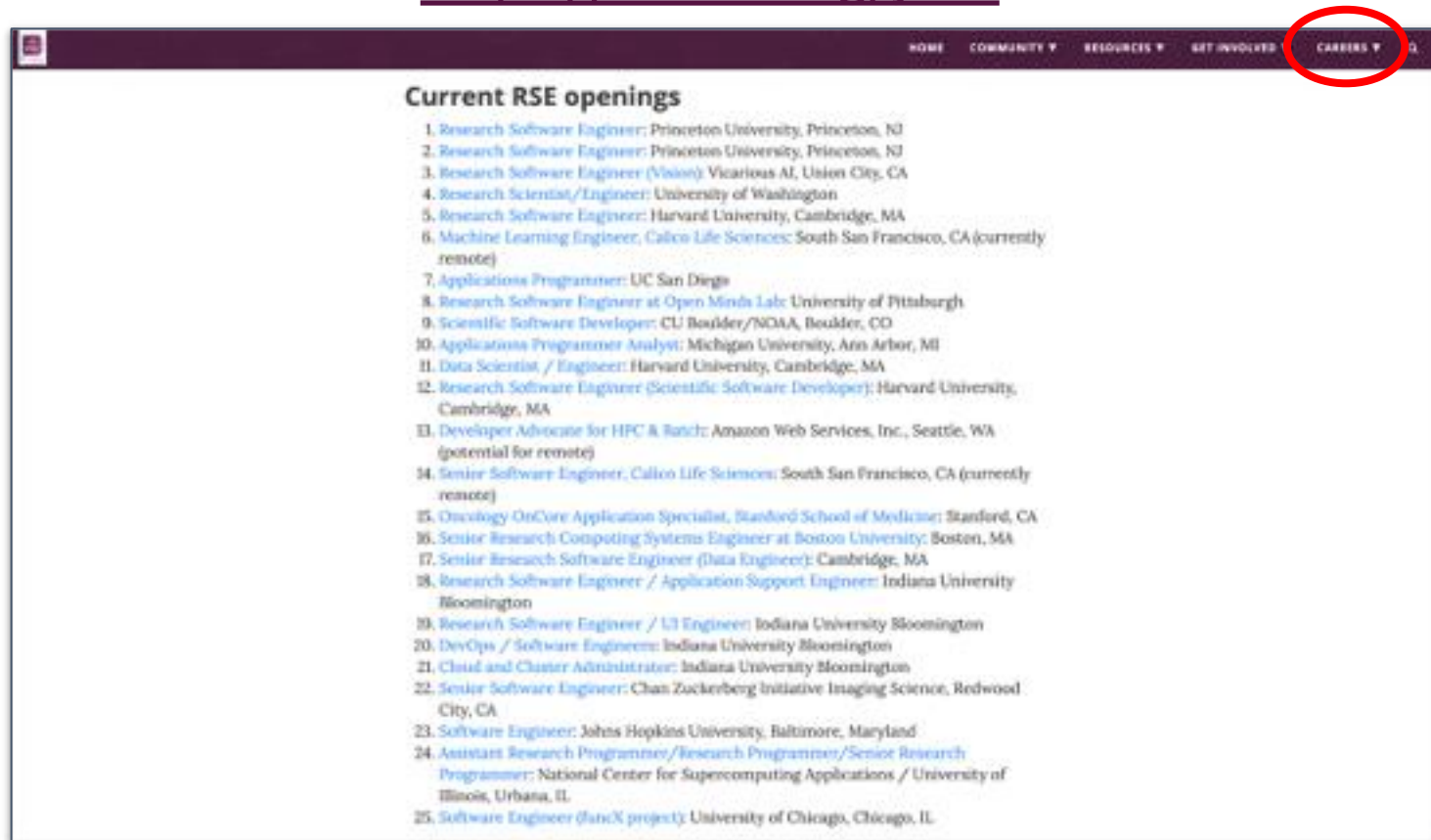


Sept 2020 Community Call

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>

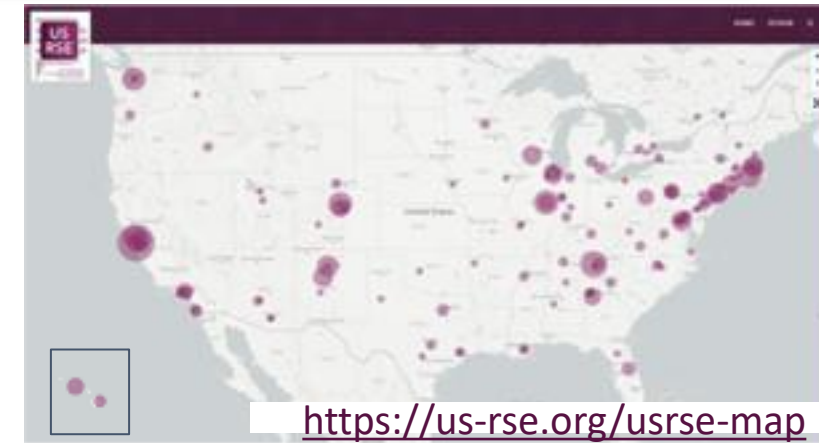
RSE Job Board

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



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



#US-RSECon2023

Join us!

Free signup for
newsletter and slack.

<https://us-rse.org/join>

Get involved
&
Lurkers welcome!



contact@us-rse.org

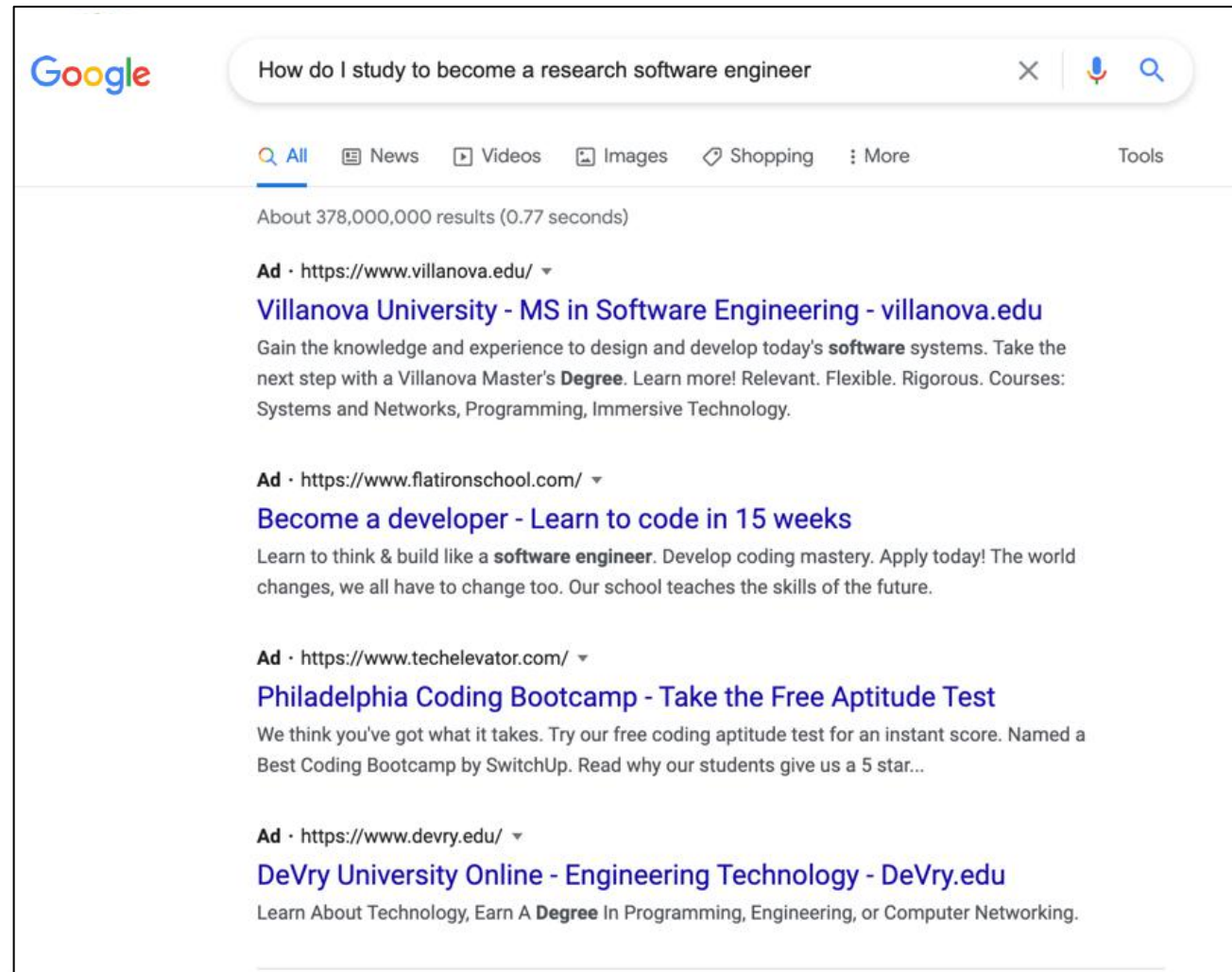
<https://us-rse.org>



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?



The image is a screenshot of a Google search results page. The search bar at the top contains the text "How do I study to become a research software engineer". Below the search bar, there are tabs for "All", "News", "Videos", "Images", "Shopping", and "More". The "All" tab is selected. Below the tabs, it says "About 378,000,000 results (0.77 seconds)". There are four search results listed, each starting with "Ad" and a URL. The first result is from Villanova University, the second is from Flatiron School, the third is from Philadelphia Coding Bootcamp, and the fourth is from DeVry University Online.

Google

How do I study to become a research software engineer

All News Videos Images Shopping More Tools

About 378,000,000 results (0.77 seconds)

Ad • <https://www.villanova.edu/>

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.

Ad • <https://www.flatironschool.com/>

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.

Ad • <https://www.techelevator.com/>

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...

Ad • <https://www.devry.edu/>

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

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.

Collecting & Linking to Training Material



<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>

Thank you!

- <https://www.princeton.edu/researchcomputing>
- <https://us-rse.org>
- <https://intersect-training.github.io>
- Questions, comments, discussions:
 - icosden@princeton.edu
 -  @IanCosden



Princeton
Research Computing

"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>

