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

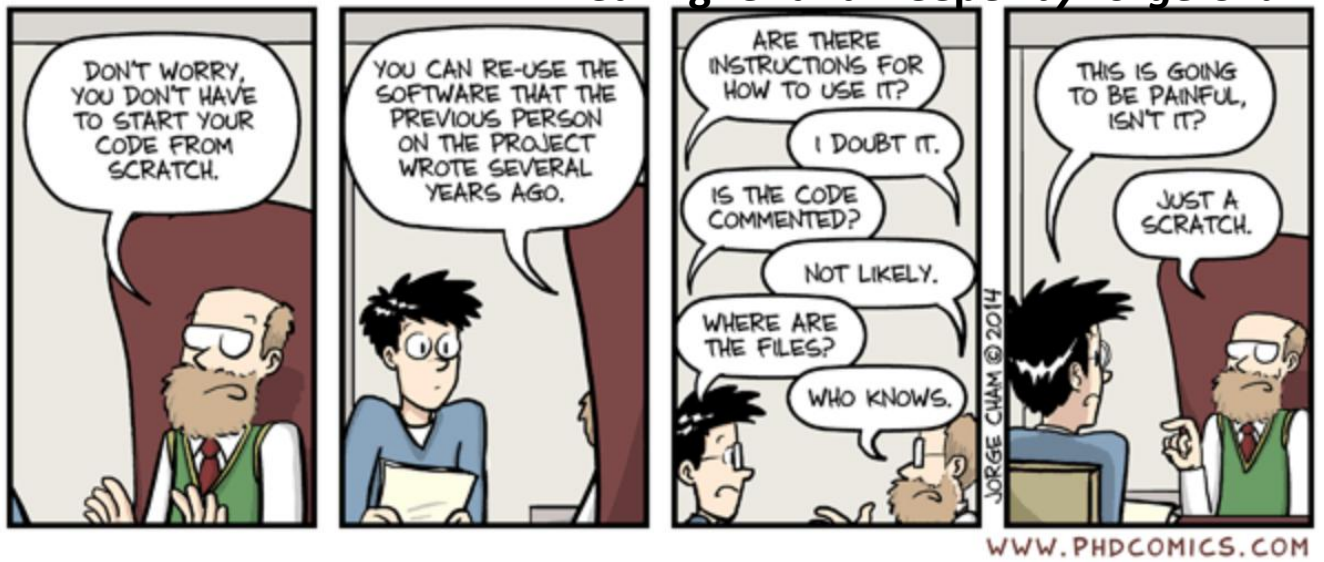
	UK[1]	US[2]
Researchers using research software	92 %	95 %
Fundamental to their research	69 %	66 %

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*



http://phdcomics.com/comics.php?f=1689

Science Depends on Software Correctness

SCIENTIFIC PUBLISHING

A Scientist's … Prob

Until recently … a trajectory … about. In 19… crystallograp… the prestigio… San Diego, … emony at the… Presidential… for Scientist… country's hig… researchers. … stream of h… detailing the… of important… cell membra…

Then the… nightmare. … researchers … *Nature* that … protein stru… had describ… paper. Whe… Chang was … that a homem… gram had fli… data, invertin… map from … derived the f… Unfortunatel… the program…

Chara… Cyan… "Will…

Jayanti … and Phi…

Departme…

Sup…

ABSTR… led to t… along w… nuclear… configu… on the… scripts … unreco…

…ne of code sinks

…om the … …ne … Shifts

…ted States

…*JCO*), explained … …traction or a cor-

Nature Protocols

72.4 (Incorrect)

73.2 **Different Calculated Chemical Shifts!**

73.2

72.7 (Incorrect)

Current Biology

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

RETRACTED: How birds outperform humans in multi-component behavior

Sara Letzner • Onur Güntürkün • 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.

RETRACTED

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*

One solution: RSE

1. Establish a position dedicated to research <u>software</u>
   - 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

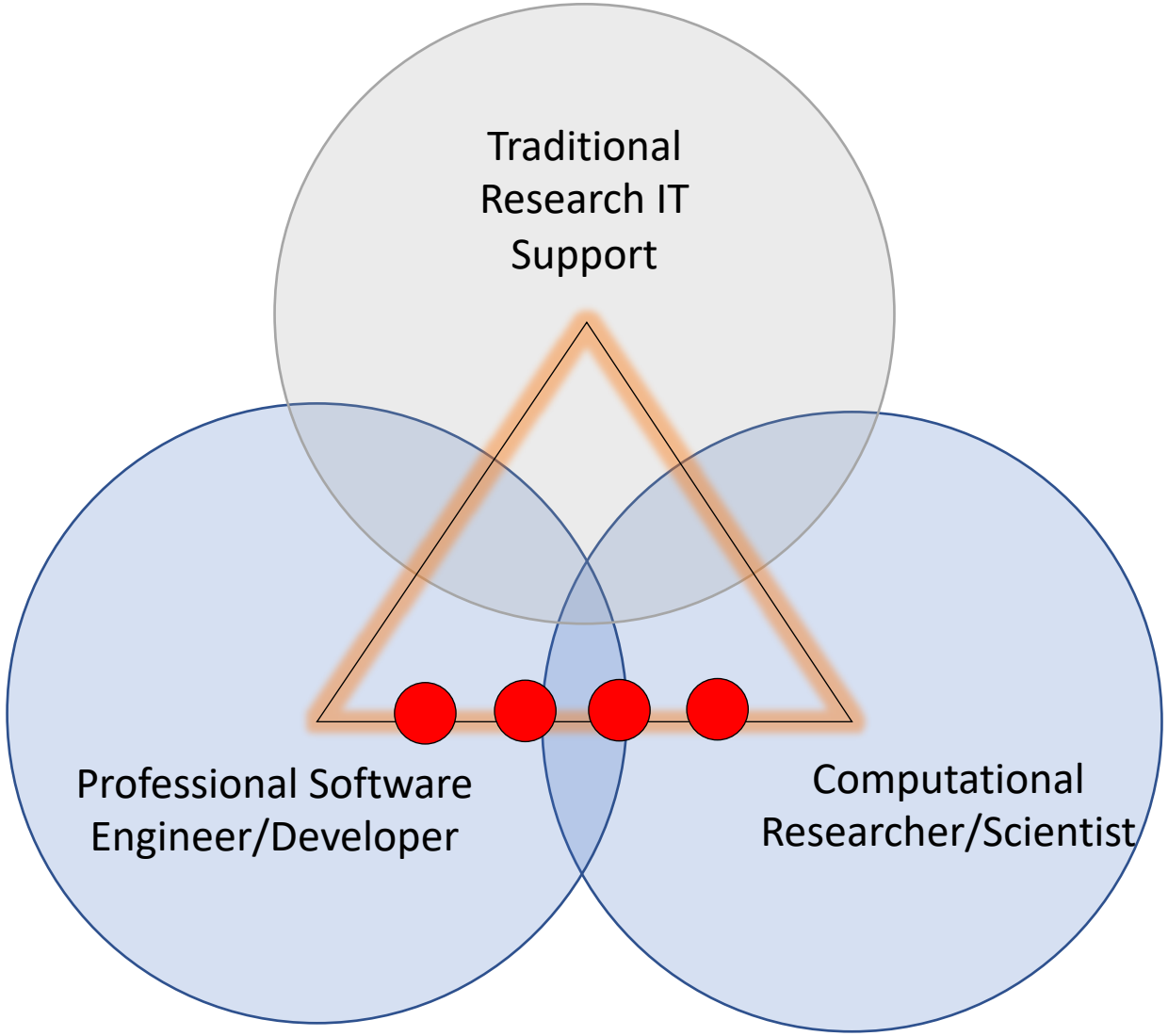
- Goal: Help researchers create the most efficient, scalable, and sustainable research code possible in order to enable new scientific advances.

- 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

Princeton RSE Group, 2021

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

Vacant

Genomics

Molecular Biology

Civil & Environmental Engineering

Operations Research & Financial Engineering

Vacant

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

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

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

RSE NL-RSE

RSE Society of Research Software Engineering (UK)

de RSE

Photo by Antonia Cozacu, Jan Philipp Dietrich, de-RSE e.V. (CC BY 4.0).

RSE Nordic-RSE

RSE RSE-AUNZ

RSE BE

US RSE US RESEARCH SOFTWARE ENGINEER ASSOCIATION

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

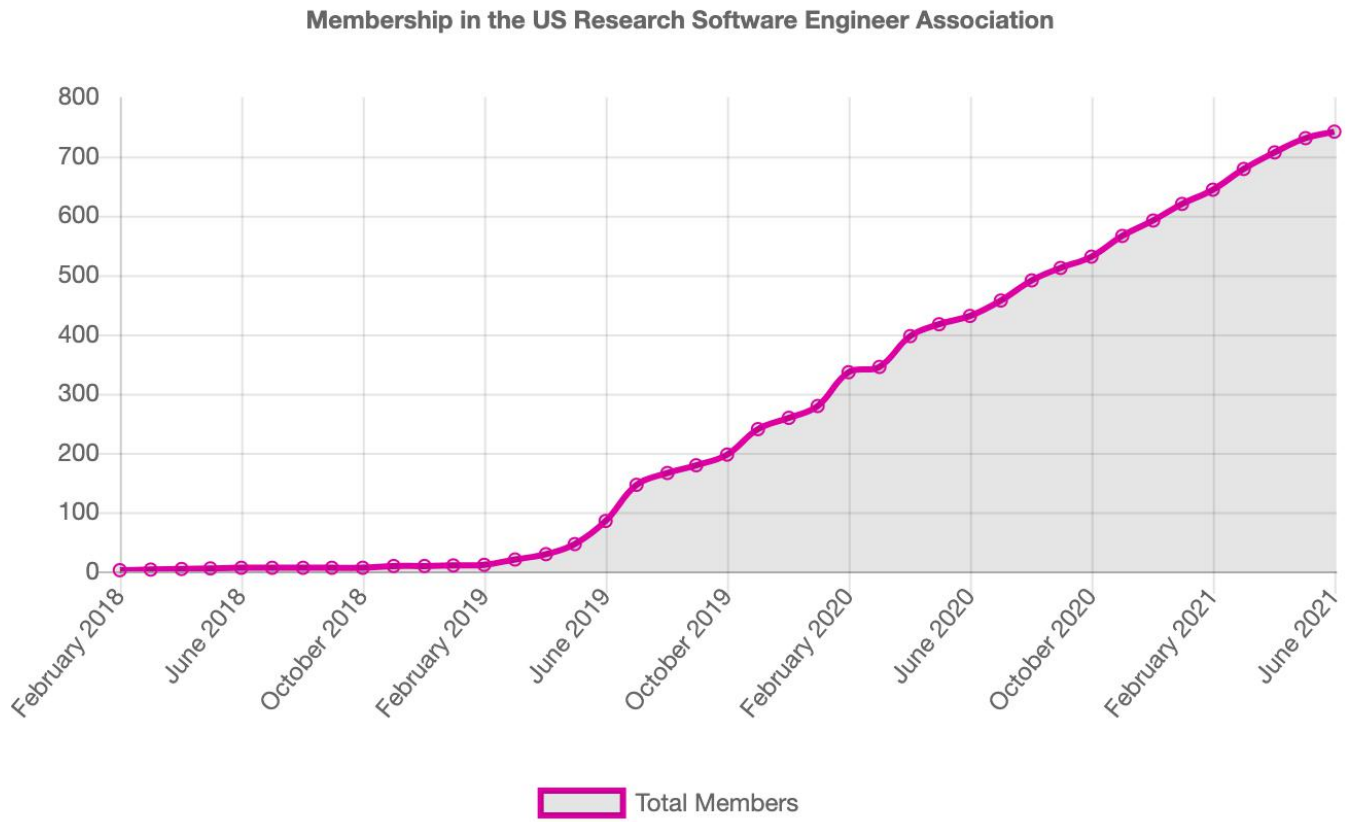
@us_rse

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



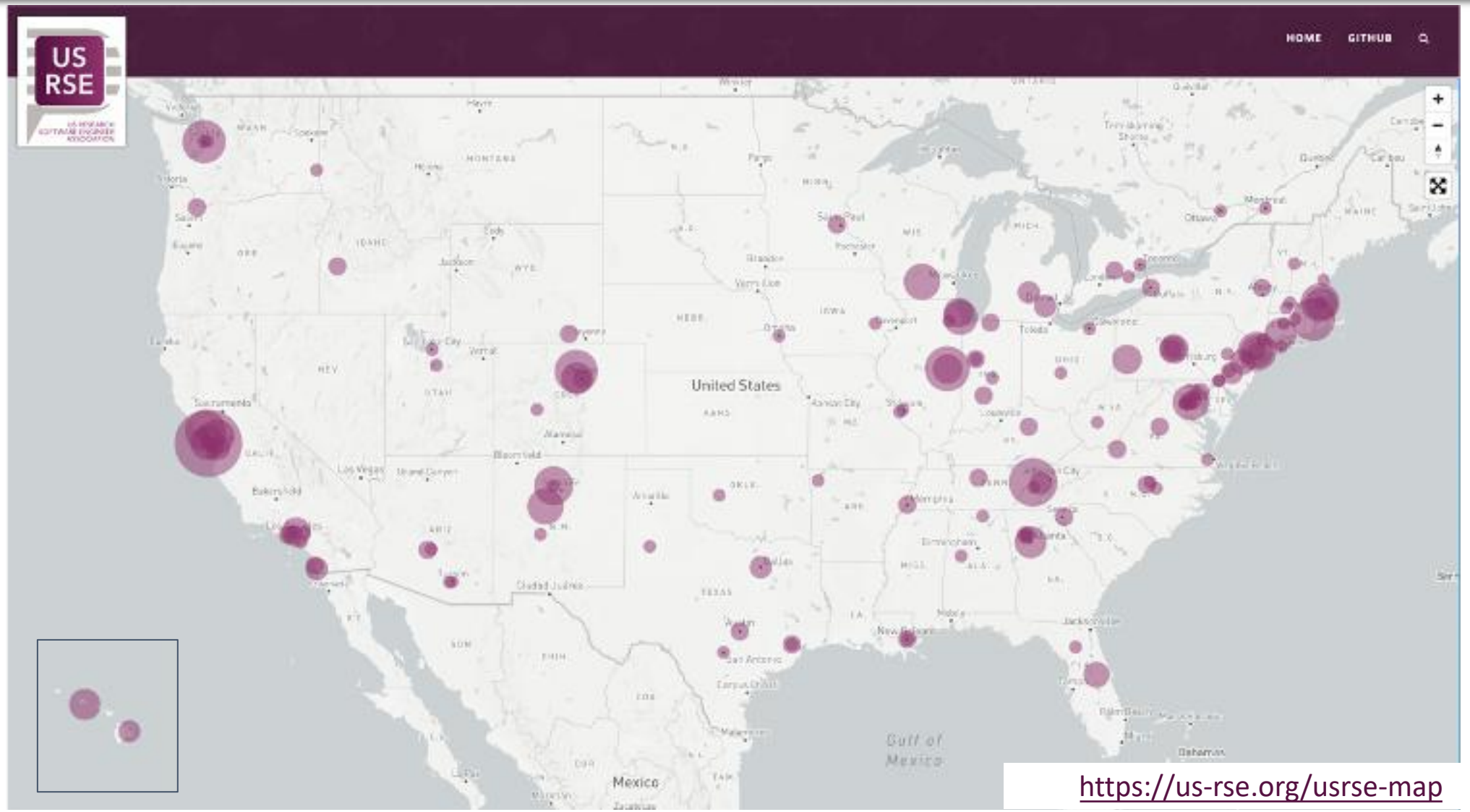
https://us-rse.org

@us_rse

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

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

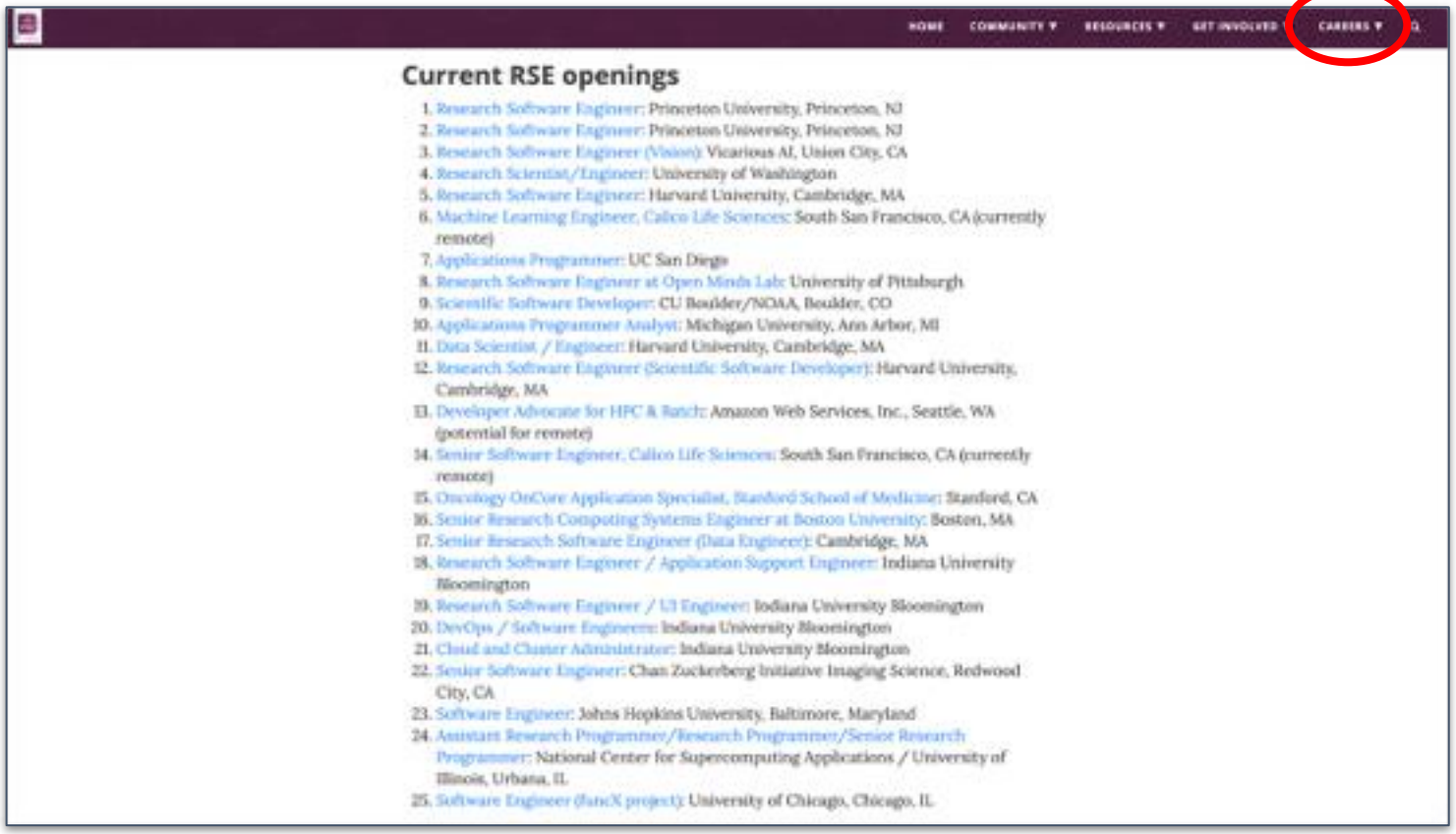
Sept 2020 Community Call

https://us-rse.org

RSE Job Board

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

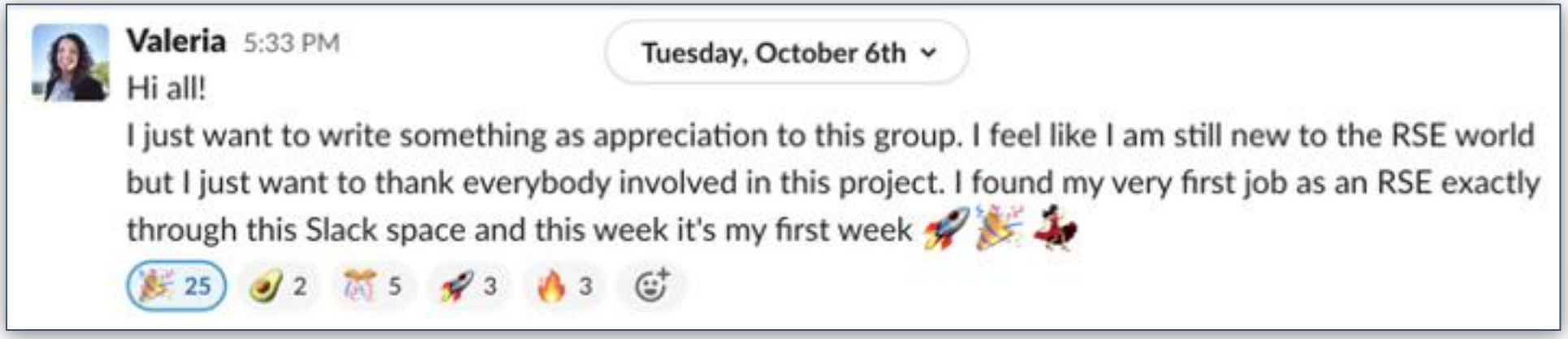
https://us-rse.org/jobs



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!

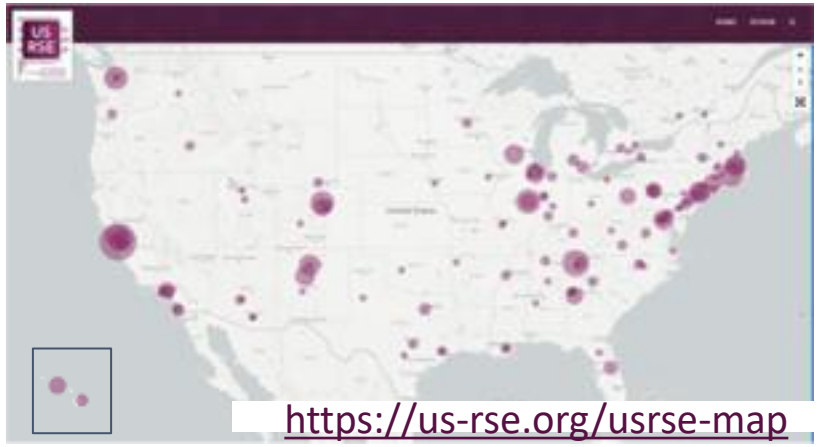


https://us-rse.org

@us_rse

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/usrse-map

#US-RSECon2023

https://us-rse.org

@us_rse

Join us!

Free signup for newsletter and slack.

https://us-rse.org/join

Get involved
&
Lurkers welcome!

contact@us-rse.org

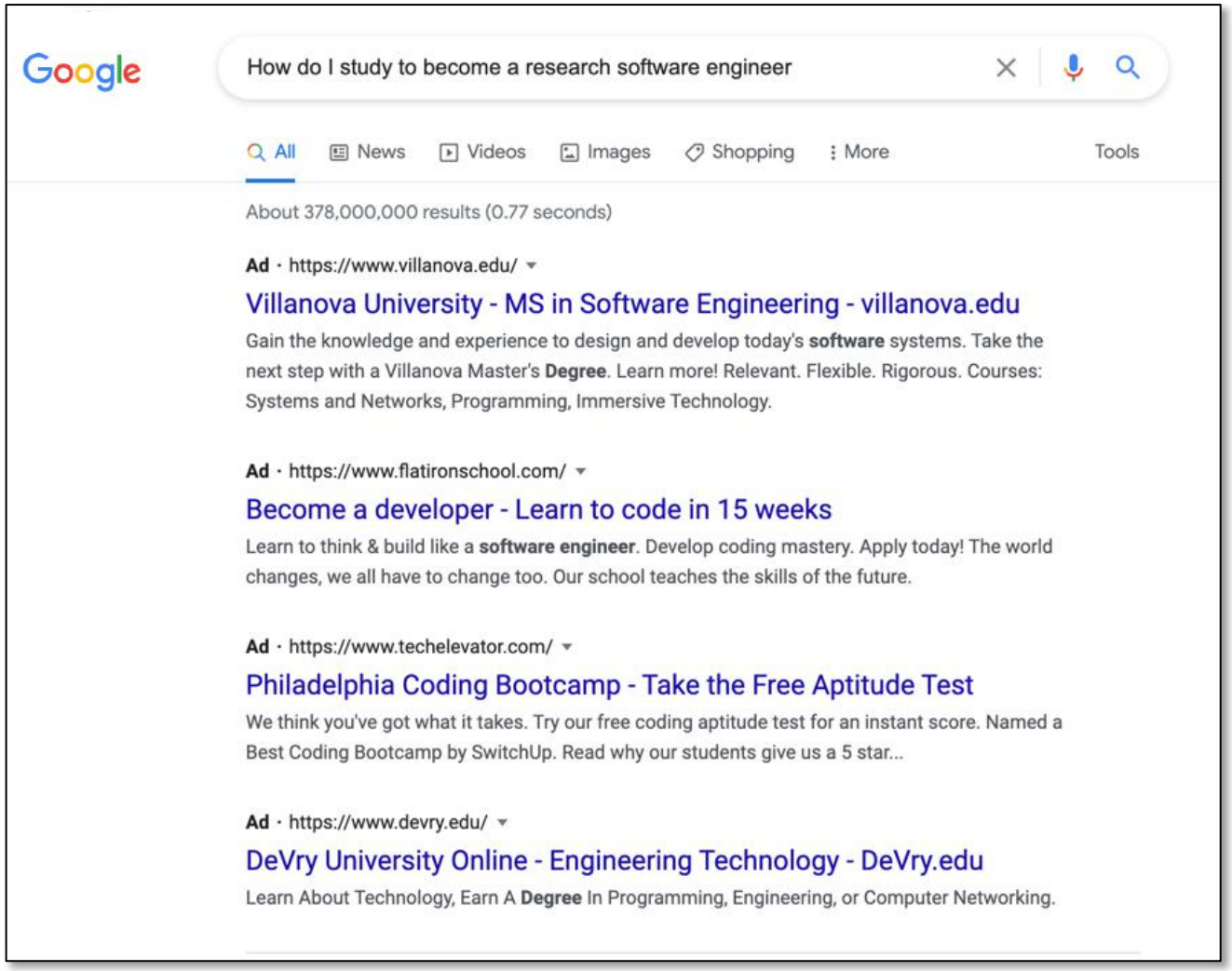
https://us-rse.org

@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?



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

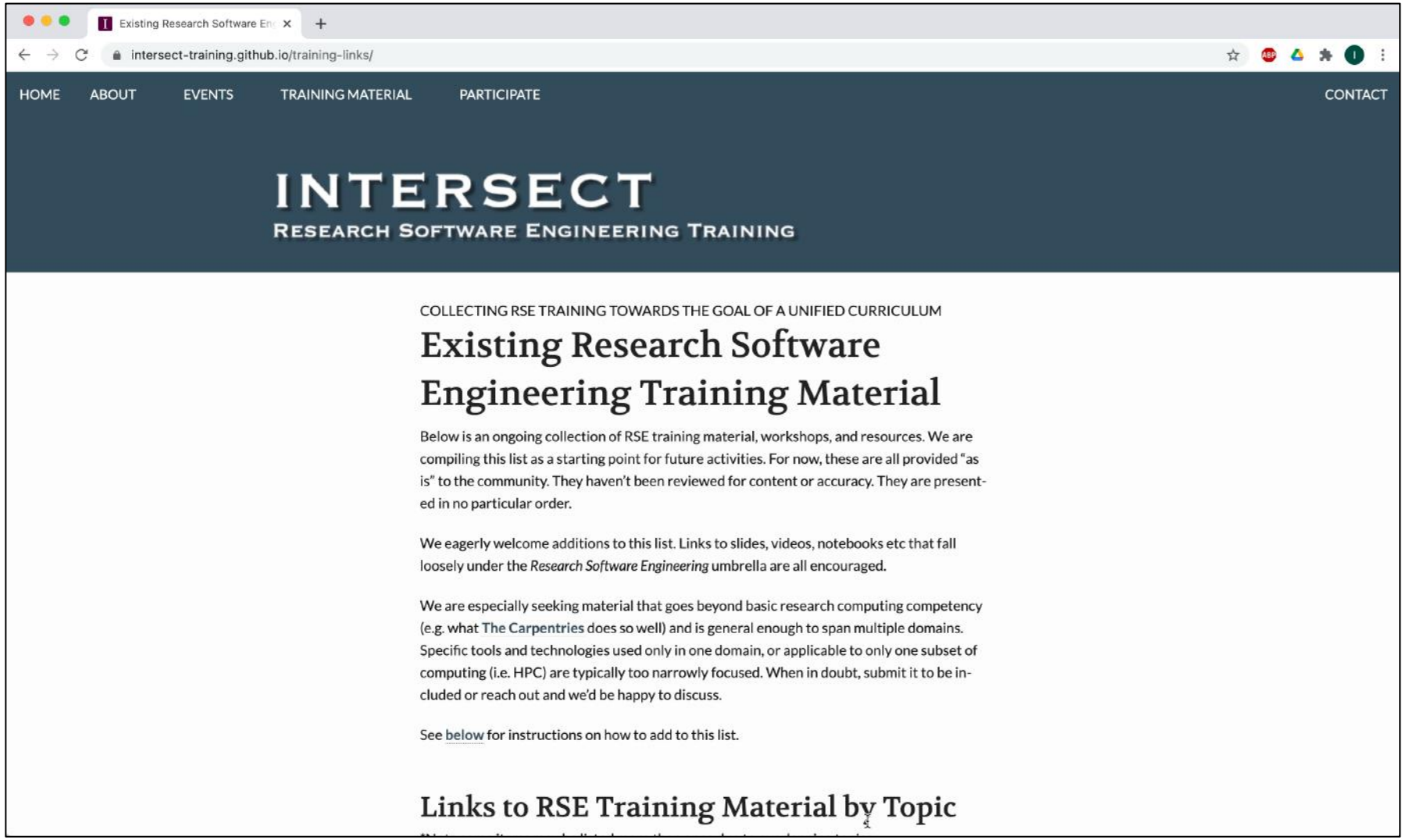
Details to be established in workshops and collaboratively:

- **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 & 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

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

US RSE
US RESEARCH SOFTWARE ENGINEER ASSOCIATION

INTERSECT
RESEARCH SOFTWARE ENGINEERING TRAINING

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