

Humans in Scaling HPC Facilitation and Education

IEEE-HPEC
September 2019

*Bob Freeman, PhD
Director, Research Technology Operations
Harvard Business School*



HARVARD | BUSINESS | SCHOOL

Why is Scaling with People an Important Problem?

- Science and research programs are no longer an isolated, work-in-silos problem
- Increase in complexity is driving an increased need for better trained workforce
 - "HPC was evolving from traditional modeling & simulation into complex workflows of data collection, modeling, simulation, analytics, and AI" – *Michael Rosenfeld, IBM*
- Changes in technology, research approaches, business landscape, and staffing
- RC Consultants, Educators, & Facilitators
 - Work with users – researchers and educators – to help them improve their research and/or education productivity and aspirations via advanced cyberinfrastructure.
- Briefly discuss 4 topic areas on
 - Create Community in Education
 - Bring HPC/HTC Computing to All Disciplines
 - Bring Facilitation Approaches to Everyone
 - Building Communities for Enabling Research

Create Community in Education

The screenshot shows the IEEE LearningNetwork Public Catalog search results for 'Computing'. The page displays 176 results in the catalog. A sidebar on the left lists categories such as Software Engineering (70), Cloud Computing (45), Cyber Security (23), Ethical Hacking (20), and Blockchain (15). The main content area features a search result for '(CSDA) Computing Foundations', which is an online course in Software Engineering, Computing. The course description states: 'This course is part of a series of eLearning courses designed to help you prepare for the examination to become a Certified Software Development Associate (CSDA) or to learn more about specific software engineering topics. Courses... English (United States)'. Below this, there is a section for '(CSDA) Engineering Foundations'.

The screenshot shows the website 'High Performance Computing Educational and Training Resources' from sighpc.org. The page features a navigation menu on the left with links for Home, Membership, By-laws, Events, Seminar Series, Training/Education Resources, Blog and Forum, and Contacts. The main content area is titled 'High Performance Computing Educational and Training Resources' and includes a paragraph: 'The links below provide access to HPC Educational resources that include, but are not limited to, courses, tutorials, reference guides, repositories and websites. The material has been roughly sorted and is grouped by experience level.' Below this, there is a section titled 'High Performance Resources' with a list of links: 'Getting Started', 'Everyone', 'Novices', 'Intermediate', and 'Advanced'. A final paragraph states: 'In addition to the HPC Resources, we have gathered general Computational Science Resources. The Computational Science resources have been sorted into two broad groups, one for general resources and one for resources that are domain specific, as noted in the lists below.'

The screenshot shows the XSEDE User Portal Overview page. The page features a navigation menu with tabs for MY XSEDE, RESOURCES, DOCUMENTATION, ALLOCATIONS, TRAINING, USER FORUMS, HELP, and ABOUT. Below the navigation menu, there is a section titled 'Having trouble finding your way around high performance computing? Our overview of XSEDE training will help you get your research done. We're here for you.' This section includes a diagram of a network and a grid of links for 'Start', 'Explore', 'Visualize', and 'Research'. The 'Start' column includes links for 'Getting Started', 'Get an XSEDE account', 'Get an Allocation', and 'Access Resources'. The 'Explore' column includes links for 'Working with XSEDE', 'XSEDE Tools', 'Beginning Programming', and 'Software Packages'. The 'Visualize' column includes links for 'Visualization Resources', 'Visualization Tools', 'Analyze Data', and 'Visualization Software'. The 'Research' column includes links for 'XSEDE Resources', 'Running Jobs', 'Data Resources', and 'More Help'.

The impact of MOOC methodology on the scalability, accessibility and development of HPC education and training

Julia Mullen
MIT Lincoln Laboratory
Lexington, MA
jrm@ll.mit.edu

Lauren Milechin
Massachusetts Institute of Technology
Cambridge, MA
lauren.milechin@mit.edu

Veronika Filinger
EPCC, The University of Edinburgh
Edinburgh, United Kingdom
v.filinger@epcc.ed.ac.uk

David Henty
EPCC, The University of Edinburgh
Edinburgh, United Kingdom
d.henty@epcc.ed.ac.uk

ABSTRACT

This work explores the applicability of Massively Open Online Courses (MOOCs) for scaling High Performance Computing (HPC) training and education. Most HPC centers recognize the need to provide their users with HPC training; however, the current educational structure and accessibility prevents many scientists and engineers who need HPC knowledge and skills from becoming HPC practitioners. To provide more accessible and scalable learning paths toward HPC expertise, the authors explore MOOCs and their related technologies and teaching approaches. In this paper

fall into two primary areas: a limited pool of trainers leading to a limited number of workshop offerings over a calendar year and the diversity of the subdomains of interest within the HPC community. Consider the diversity of subdomains. While there is some overlap of content and skills that HPC practitioners must learn, each subdomain within the HPC ecosystem has a different focus. Additionally, as research applications in medicine, social science and biology become more complex and require more extensive computing power, a new cohort of students with limited computer literacy are searching for a pathway to HPC expertise. This variety

Create Community in Education

Harvard University Research Computing*
Trainers group (HURT)

TORCH: Trainers Optimizing Research
Computing across Harvard



FAS RC



FAS HumRC



HMS RITG



HBS RCS



HSPH BioInfo



FAS / SEAS



HARVARD IT ACADEMY

Three course cards are displayed:

- IQSS**
R Data Wrangling (University-Wide)
Version: 1.0
Course (1 class)
Statistics courses usually use clean and well-behaved data for examples and homework. This leaves many unprepared for the messiness and chaos of data in the real world. This workshop aims to prepare you for dealing with messy data by walking you through r...more
0 USD | [VIEW CLASSES](#)
- Harvard Business School**
Introduction to R (HBS)
Version: 1.0
Course (0 class)
R is a programming language that is especially powerful for data exploration, visualization, and statistical analysis. With hands-on exercises, we will learn how to import and manage datasets, create R objects, install and load R packages, conduct basic s...more
[SHOW INTEREST](#)
- FAS Research Computing**
Setting environment - C/Fortran/Python/R on FASRC Cluster (FAS)
Course (1 class)
This session will cover setting up your own environment on FASRC cluster like R, Python user environments, compiling C and Fortran process. Online details: Join Zoom meeting <https://harvard.zoom.us/j/189036384> Join by telephone +1 929 436 2866 US (New York ...more
0 USD | [VIEW CLASSES](#)

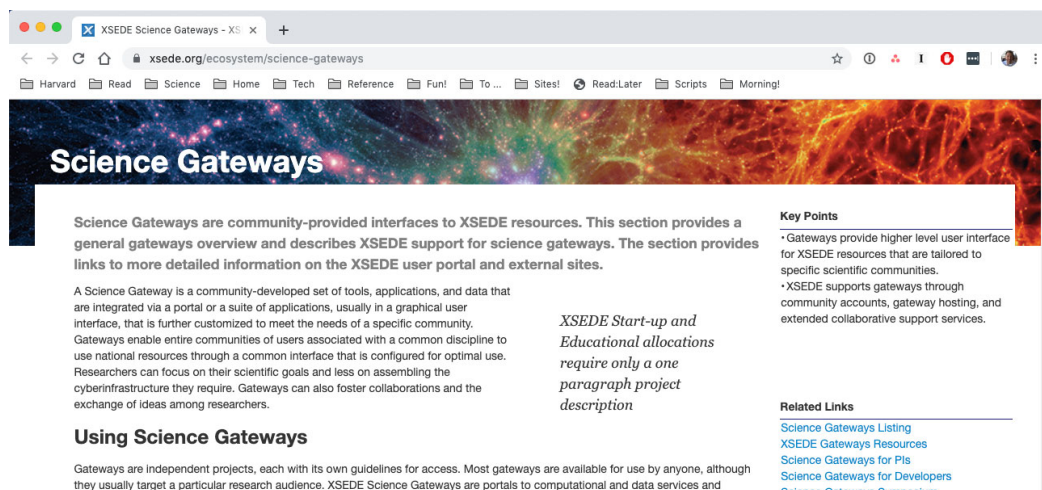
Article page for "Building a local community of practice in scientific programming for life scientists" on PLOS Biology.

Metadata: 0 Save, 1 Citation, 4,450 View, 17 Share

Published: November 28, 2018 • <https://doi.org/10.1371/journal.pbio.2005561>

Navigation: [Article](#) | [Authors](#) | [Metrics](#) | [Comments](#) | [Media Coverage](#) | [Download PDF](#)

Bring HPC/HTC Computing to All Disciplines



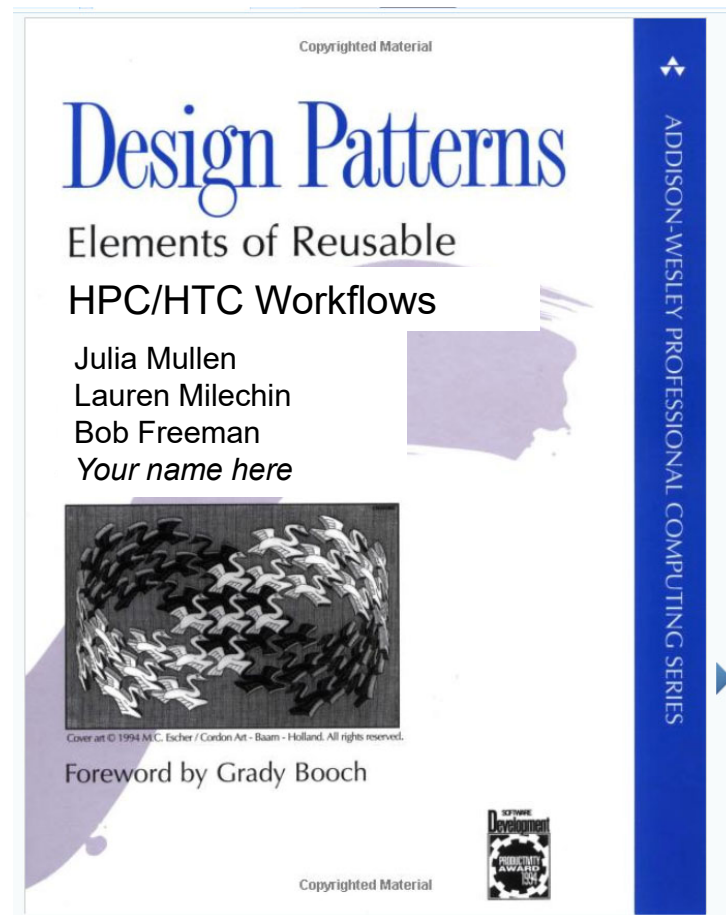
Using Science Gateways

Gateways are independent projects, each with its own guidelines for access. Most gateways are available for use by anyone, although they usually target a particular research audience. XSEDE Science Gateways are portals to computational and data services and resources

- High-1
- Work
- Gener
- Collat
- Job si
- Educ

A Science Gateway is a **community-developed set of tools, applications, and data** that are **integrated** via a portal or a suite of applications, usually in a graphical user interface, that is further customized to **meet the needs of a specific community**. Gateways **enable entire communities of users** associated with a common discipline to use national resources through a common interface that is configured for optimal use. **Researchers can focus on their scientific goals** and less on assembling the cyberinfrastructure they require. Gateways can also **foster collaborations and the exchange of ideas** among researchers.

XSEDE Start-up and Educational allocations require only a one paragraph project description



HPC Carpentry: Teaching basic skills for high-performance computing.

HPC Carpentry is a set of teaching materials designed to help new users take high-performance computing systems. No prior computational experience is needed. These lessons are ideal for either an in-person workshop or independent study.

NOTE: This is the draft HPC Carpentry release. Comments and feedback are welcome.

INTRO TO HIGH-
PERFORMANCE
COMPUTING

THE SHELL AND
HIGH-PERFORMANCE
COMPUTING

ANALYSIS PIPELINES
WITH PYTHON

Next steps & How can you get involved??

In progress:

- Tech wizardry to facilitate scheduler-agnostic lessons with site-specific (forked & rendered) scheduler & resource details
- Updated *Outline.md* document with lesson topics, objectives, & goals to help direct lesson development

If you wish to help develop lessons, join discuss-hpc <https://carpentries.topicbox.com/groups/discuss-hpc>

If you've run these lessons, please give us feedback! (discuss-hpc)

Look for upcoming details about an HPC-Carpentry Sprint!!

IEEE-HPEC, Bob Freeman, PhD, 25-Sep-2019

<https://hpc-carpentry.github.io/> <https://github.com/hpc-carpentry/>

WHAT IS HPC CARPENTRY?

Each of these lessons are designed to teach **how to interact with a compute cluster in a specific way**. The first two lessons, [Intro to HPC](#), and [Shell on HPC](#) are meant to be used together as a **general introduction to the bash command line and submitting jobs on a typical HPC cluster**. The second two lessons focus on two separate use cases for HPC: **running large numbers of compute jobs** (often with complex inter-job dependencies), or **creating parallel programs** that execute across one or more compute nodes. Each lesson takes roughly a full day to teach (or work through independently). It is intended that the Intro to HPC lesson be taught together with either the Python or Chapel component to form a two-day workshop or the Intro to HPC + Shell can be taught as a standalone.

H A R V A R D | B U S I N E S S | S C H O O L

Bring Facilitation Approaches to All Disciplines

NSF-Funded Project – ACI-REF

\$5.3M NSF Award supports the project leadership team and 2 Facilitators for each of the 6 partner sites for 2 years.



PI: Jim Bottum, Clemson

Project Leadership:

- James Cuff, Harvard
- Maureen Dougherty, USC
- Gwen Jacobs, Hawaii
- Paul Wilson, Wisconsin
- Tom Cheatham, Utah (PI Chair)
- Marcin Ziolkowski, Clemson

Facilitator Lead: Bob Freeman, Harvard

Chief Scientist: Miron Livny, Wisconsin

- "Work smarter, better, faster": be more efficient
- ... and to Think Differently: ask bigger questions and not be constraint by current (desktop) resources
- Push for answers – always find out why; just don't "fix the problem"

Empowering Would-Be Computational Researchers

We are dedicated to forging a **nationwide alliance of educators to empower local campus researchers to be more effective users of advanced cyberinfrastructure (ACI)**. In particular, we seek to work with the **"long tail" of ACI users**—those scholars and faculty members who traditionally have not benefited from the power of massively scaled cluster computing but who recognize that their research requires access to more compute power than can be provided by their desktop machines.

To do so, the consortium is building a **coordinated network of ACI-REFs**, campus champions whose mission is to **leverage existing resources and "make a difference" in supporting their local campus researchers**, while also unifying member institutions under common objectives.

Aligned with strategies and models for advancing the nation's research capabilities, the ACI-REFs will be computational scientists with broad backgrounds in computer instruction, uniquely qualified to **bridge the gap between researchers and campus IT**.

A Startup Framework for Building Digital Research Capacity and Community at UCLA

Annelie Rugg, UCLA Humanities Technology

Building Communities for Enabling Research



Our community of over 600 Campus Champions **promotes and facilitates the effective participation** of a diverse national community of academic and not-for-profit institutions in the **application of advanced digital resources** and services to accelerate discovery, enhance education, and foster scholarly achievement.

There are over **600** Campus Champions: Including directors, faculty, researchers, students, as well as research-enabling and systems professional At over **300** academic, non-academic, and not-for-profit research-focused institutions

<https://www.xsede.org/community-engagement/campus-champions>

IEEE-HPEC, Bob Freeman, PhD, 25-Sep-2019



Virtual Residency Introductory/Intermediate Workshop: Overview



Henry Neeman, University of Oklahoma
Director, OU Supercomputing Center for Education & Research (OSKER)
Associate Professor, College of Engineering
Adjunct Faculty, School of Computer Science
XSEDE Campus Engagement Joint Co-Manager
Virtual Residency Introductory/Intermediate Workshop 2019
Sunday June 2 2019



This Is So New, We Don't Know How to Teach It

- For the Introductory workshops, we were able to find speakers for most of the topics we covered.
- For this combined Introductory/Intermediate workshop, very few of the topics are issues that any of us know enough about to be able to teach it to others at the Intermediate level.
- So, most of the Intermediate sessions are panels – we'll learn from each other!

Building Communities for Enabling Research

NSF ACI-1620695: “RCN: Advancing Research and Education Through a National Network of Campus Research Computing Infrastructures - The CaRC Consortium”



Initially, a group of 28 institutions intent on developing charter, membership rules, etc.

Now producing deliverables for the community and open to all interested and willing to contribute.

<https://carcc.org/> info@carcc.org



The Campus Research Computing Consortium (CaRCC) is an organization of dedicated professionals developing, advocating for, and advancing campus research computing and data* and associated professions

Ecosystem Workshop Participating Orgs

- Association of Research Libraries (ARL)
- Big Data Hubs
- Campus Champions (CC)
- Campus Research Computing Consortium (CaRCC)
- Carpentries
- Coalition for Academic Scientific Computation (CASC)
- Coalition for Networked Information (CNI)
- Education Opportunities (HPC University, SIG HPC Education)
- EDUCAUSE
- EPOC/CI Engineers
- Global Environment for Network Innovations (GENI)
- HPC Systems Professionals
- Midscale Experimental Research Infrastructure Forum (MERIF)
- Minority Serving Institution/ Historically Black Colleges & Universities
- Open Science Grid (OSG)
- Quilt (Regional Networks)
- Research Data Access & Preservation Association (RDAP)
- Women in HPC (WHPC)
- XSEDE (Extreme Science and Engineering Discovery Environment)

Building Community: the People Network



Initially organized around the five “facings”, each track provides for focused discussion. Each has (or will have) monthly videoconferences and mailing list.

- **Researcher-facing:** Launched Spring 2018
- **Systems-facing:** Launched Jan 2019
- **Data-facing:** Launched May 2019
- **Emerging-Centers:** Launching Sept 2019!
- **Software-Facing:** Launching at a future date
- **Sponsor/Stakeholder-facing:** To be developed

More info and join: <https://carcc.org/people-network/>

Researcher- & Staff-facing, & Professional Development

- Effective Meetings
- Promoting Good User Behavior
- Crowdsourcing Training Friction Points / Open Mic
- Communicating a Problem / Open Mic
- Getting the researcher turnout you want
- Software, containers, libraries, ... oh my!
- XSEDE ECSS: How we can help you
- Lightning talks / Open Mic
- Creating effective training materials
- PEARC 2019 Follow-up
- "A Startup Framework for Building Digital Research Capacity and Community at UCLA"

Facilitation beyond the RC Center:

- Partnerships & collaborations with other service groups
- IT Department
- Research Administration
- Behavioral Research
- Security & Compliance

Questions?

