HARVARD UNIVERSITY RESEARCH COMPUTING COUNCIL

Harvard Academic Computing Committee
April 20th, 2017
RESEARCH COMPUTING COUNCIL

• Formed in September 2016 in response for growing demand for Research Computing across Harvard

• Includes representatives from Harvard’s major RC organizations – IQSS, FAS RC, HMS, SEAS, HBS

• Short-term, the Council is focusing on increasing alignment across the University

• Longer-term, the Council is focused on increasing investment in RC support for faculty
<table>
<thead>
<tr>
<th>Name</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acacia Matheson</td>
<td>HUIT</td>
</tr>
<tr>
<td>Alan Wolf</td>
<td>HUIT</td>
</tr>
<tr>
<td>Amir Karger</td>
<td>HMS RC</td>
</tr>
<tr>
<td>Anne Marguiles</td>
<td>HUIT</td>
</tr>
<tr>
<td>Arnold Paul</td>
<td>HBS</td>
</tr>
<tr>
<td><strong>Catie Smith</strong></td>
<td>HUIT</td>
</tr>
<tr>
<td>Christian Hamer</td>
<td>HUIT</td>
</tr>
<tr>
<td>Christopher Botka</td>
<td>HMS IT</td>
</tr>
<tr>
<td>Cris Rothfuss</td>
<td>IQSS</td>
</tr>
<tr>
<td>David Sobel</td>
<td>HUIT</td>
</tr>
<tr>
<td>Dustin Tingley</td>
<td>HBS</td>
</tr>
<tr>
<td>Francesca Dominici</td>
<td>HSPH</td>
</tr>
<tr>
<td>Gabriele Fariello</td>
<td>DCE</td>
</tr>
<tr>
<td>Harry Hoffmann</td>
<td>HUIT</td>
</tr>
<tr>
<td>V Judson Harward</td>
<td>HUIT</td>
</tr>
<tr>
<td>Ingrid Skoog</td>
<td>HUIT</td>
</tr>
<tr>
<td>James Cuff</td>
<td>FAS RC</td>
</tr>
<tr>
<td>Jefferson Burson</td>
<td>HUIT</td>
</tr>
<tr>
<td>Jim Waldo</td>
<td>SEAS</td>
</tr>
<tr>
<td>Krista Coleman</td>
<td>HSPH</td>
</tr>
<tr>
<td>Leonard Wisniewski</td>
<td>IQSS</td>
</tr>
<tr>
<td><strong>Merces Crosas</strong></td>
<td>IQSS</td>
</tr>
<tr>
<td>Nicole Breen</td>
<td>CADM HR</td>
</tr>
<tr>
<td>Noah Selsby</td>
<td>HUIT</td>
</tr>
<tr>
<td>Paul Williams</td>
<td>HMS IT</td>
</tr>
<tr>
<td>Rainer Fuchs</td>
<td>HMS IT</td>
</tr>
<tr>
<td>Robert Freeman</td>
<td>HBS</td>
</tr>
<tr>
<td>Rebecca Loose</td>
<td>SEAS</td>
</tr>
<tr>
<td>Sandra Silk</td>
<td>HUIT</td>
</tr>
<tr>
<td>Stephen Gallagher</td>
<td>HBS</td>
</tr>
<tr>
<td>Suzanne Wones</td>
<td>HU Library</td>
</tr>
<tr>
<td><strong>Scott Yockel</strong></td>
<td>FAS RC</td>
</tr>
<tr>
<td>Troy Adair</td>
<td>HBS</td>
</tr>
</tbody>
</table>
The Council formed three working groups, focused on the prioritized areas for alignment.

These working groups include:

- **ACCESS**: Provide seamless access to a set of common technologies
- **TALENT**: Quickly connect researchers with RC facilitators and subject-matter experts
- **DATA**: Establishing common practices and support for research data management

Foundational to these areas, additionally, is exploring ways to better coordinate human resources activities, and, in particularly, recruitment and retention activities.
THREE PILLARS

- DATA
- TALENT
- HR
- ACCESS
- SHARED INFRASTRUCTURE
- RESEARCH
- DATA SCIENCE INITIATIVE
Research Computing Council
Supporting Research Computing Across the University

Our mission is to enhance the ability for Harvard University Faculty, Researchers and Staff to gain access to our world class high performance technical computing resources

Learn More »

Data
Why is research data management important? What services does Harvard offer to manage and share data?
Research Data Management @Harvard »

Access
What resources does Harvard offer for research computing? How do I access them? What are my options?
Research Computing Resources @Harvard »

Talent
Who can help me at Harvard with research computing and analysis? Is there training? What services can I use?
Research Computing Talent @Harvard »
ACCESS SUBGROUP

Scott Yockel
Research Computing Resources @Harvard

Harvard University offers four research computing resources to suit all your needs. Learn more to find out which one works best for you. Our goal is to eventually provide one single access to all solutions.

FAS Research Computing
FAS provides Odyssey as the Research Computing resource for Harvard affiliates at FAS, SPH, and SEAS.
Learn more >>

HMS Research Computing
HMS provides Orchestra as the research computing resource for the Medical School affiliates.
Learn more >>

IQSS Research Computing
IQSS offers RCE as the research computing resource for all Harvard affiliates working in social science research.
Learn more >>

HBS Research Computing
HBS offers research computing resources and services for Business School affiliates.
Learn more >>
Primary Objective

- Develop a technology resource portal (website)
  - PIs will see what is **internally** available to them from their host institution by nature of their appointment and how to obtain access.
  - In addition, PIs will see what **external** resources are available across the University and how to request access to those resources.

Secondary Objectives

- Enable research within the University that otherwise wouldn’t be possible due to technology gaps, or would be done externally.
- Foster collaborative relationships between schools / institutions
  - e.g. JupyterHub for teaching and research, dbGaP (NIH Human Genome data set)
COMMON RESOURCES

- Compute Cluster (FAS, HMS, IQSS, HBS, HSPH, SEAS)
  - AMD/Intel x86_64, RHEL/CentOS, Batch Processing
- Storage (FAS, HMS, IQSS, HBS, HSPH, SEAS)
  - NFS/SMB, Lustre
- Hosted Machines (FAS, HMS)
- Hosted Databases (FAS, HMS, IQSS, HBS)
  - MySQL, Postgres, MongoDB
- Data Security (FAS, HMS, IQSS, HBS)
  - Level 3/4
- Training Platforms (FAS, HMS, IQSS, HBS)
  - Unix, R

NICHE RESOURCES

- Memory (Large Capacity)
  - 512 GB (FAS, IQSS, HSPH)
  - 1 TB (FAS)
- Accelerators
  - NVidia GPUs (FAS, HMS, SEAS)
  - Intel Phi (FAS)
- Storage (High Performance)
  - 50 GB/s r/w (FAS)
- High-Speed Data Transfer
  - 10 GbE (FAS & HMS)
  - 40 GbE (FAS)
  - Globus Endpoint (FAS)
- Training Platforms
  - Python (HMS, IQSS, HBS)
  - MatLab (HMS)
  - Tableau (HBS)
What are your resource needs?

- Compute
- Storage
- Software
- Hosted Machine
- Database

Do you have access to these resources within your host institution (primary appointment)?

- Yes
  - Connect PI with the appropriate contact(s) for their institution.

- No
  - Connect with the appropriate contact(s) for the resource institution. Work with PI and resource owner on needs assessment and negotiation of resource utilization.

**Resource Locations:**
- FAS
- HMS
- IQSS
- HBS
- HSPH/SEAS (via FAS)

**Flowchart Options:**
- Number/size of resources needed
- Duration of use
- Cost
NEXT STEPS

• Build-out the resource portal - efforts will include:
  • Finalizing the decision tree
  • Exploring how to leverage HarvardKey for login/authentication
  • Defining the list of contacts at each school/institute responsible for approving requests
  • Cost structure that would support university wide use of resources, thus promoting the appropriate pairing of users with resources, and would in-turn help shift away from the culture of price shopping and camping out
  • Data transfers - since this is becoming an emerging challenge, it would be wise to develop a university-wide solution
  • Data repositories - since it is costly to support 100+TB data sets, it would be wise to develop cross-institutional solutions
TALENT SUBGROUP

Troy Adair
Across Harvard University, there is a wide and extensive community supporting research resources. Learn more about how to get in touch with and leverage the experience and knowledge of fellow HU researchers that are passionate about the same things you are. Over time, we encourage you to continue to engage in the areas you're interested in, and to help us foster a growing community of fellow enthusiasts.
TALENT SUBGROUP GOALS

Primary Goal

• Create comprehensive database of pertinent organizations and individuals involved in research computing across all schools and organizations.

Supporting Objectives

• Help define and start a collective approach to recruiting, developing, and retaining research computing staff across all schools at Harvard University.

• Create, conduct, and analyze staff and manager surveys/focus groups to supplement existing university HR information with skills inventory, skills need assessments, and career interest and opportunity mappings.

• Use analysis above to assess overall RC workforce training and development focus.

• Create Research Computing Community of Practice to convene physically 2x/year, and set the stage for fostering continuous community interactions via a web portal.
SKILLS INVENTORY

Currently 63 “Skills”, Categorized as:

• Statistics Software (e.g., R, Stata, SAS, SPSS, etc...)
• Programming Software (e.g., Python, Java, etc...)
• Other Research Software (e.g., Qualtrics, NVivo, PCTeX, Tableau, etc...)
• Research Network/Systems Administration
• Research Methodology (e.g., Research Design, Data Management, Data Collection, etc...)
DATA SUBGROUP

Mercè Crosas
Research Data Management @Harvard

Towards FAIR data: Findable, Accessible, Interoperable, and Reusable

“Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process.”


Start Exploring the Data Lifecycle ▶

Data Acquisition and Planning
What do I need to know before bringing research data into Harvard? How do I prepare for a data management plan?
• Data User Agreement, Data Management Plan, Harvard Policies, licensed data.

More »

Data Storage
Where and how should I store my research data? What are the options at Harvard?
• Data files, documentation, logbooks, notebooks, security levels, and permits.

More »

Compute and Analysis
What are the options for research computing at Harvard? Which tools or methods should I use for my research?
• Harvard Research Computing, data science and computational help.

More »

Data Sharing and Archival
What is Data Sharing and why is it important? What do Funders and Journals require? Can I get help on data curation?
• Harvard Dataverse repository, domain repositories, Open Data policies.

More »

Preservation Services
What is long-term preservation? What services do Harvard offer for preservation of data collections?
• Harvard Library services, format migration, suitable medium.

More »

Data Disposal
Are there some cases when I need to destroy my data? How should I do it? What services do Harvard offer?
• Contractual obligations, method of disposal, documentation.

More »