Research Software Sustainability

Sandra Gesing
UC Santa Barbara
October 10, 2019

sciencegateways.org
University of Notre Dame

- In the middle of nowhere of northern Indiana (1.5 h from Chicago)
- 4 undergraduate colleges
- ~35 research institutes and centers
- ~12,000 students
Center for Research Computing

- Software development and profiling
- Cyberinfrastructure/science gateway development
- Computational Scientist support
- Collaborative research/ grant development
- System administration/ prototype architectures
- Computational resources: 25,000 cores+
- Storage resources: 3 PB
- National resources (e.g., XSEDE)
- ~50 researchers, research programmers, HPC specialists

http://crc.nd.edu

CRC and OIT building

CRC HPC Center (old Union Station)
• Shift from disciplinary teams to multidisciplinary teams
• A plethora of computational methods and available data
• Diversity of funding necessary
• Learning beyond the curriculum

“We like to bring together people from radically different fields and wait for the friction to produce heat, light and magic. Sometimes it takes a while.”
• Genomics
• Proteomics
• Metabolomics
• Immunomics
• System biology
• Molecular simulations
• Docking
• Epidemiology
• ...

Life Sciences

Black Swallowtail – larvae and butterfly
The Genomics Boom

February 16, 2001
biotech company Celera

February 15, 2001
The Human Genome Project
The Genomics Boom

Craig Venter (left) and Francis Collins (right)
Big Data

• Explosion in the quantity, variety and complexity of data

• Questions can be answered impossible to even ask about 10 years ago

• Costs far reduced (e.g., Human Genome project, 15 years, ~$2 billion; today ~3 days, $1000)
Research Software

Use
90%

Can’t continue without
70%

95%

63%

http://doi.org/10.5281/zenodo.843607
> 50% neither formal nor informal training in software engineering

<table>
<thead>
<tr>
<th>Use</th>
<th>90%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can't continue without</td>
<td>70%</td>
<td>63%</td>
</tr>
</tbody>
</table>
Lack of career paths

<table>
<thead>
<tr>
<th>Use</th>
<th>90%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can't</td>
<td>70%</td>
<td>63%</td>
</tr>
<tr>
<td>continue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>without</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://doi.org/10.5281/zenodo.843607
## How to cite software?

<table>
<thead>
<tr>
<th>Use</th>
<th>90%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can't continue without</td>
<td>70%</td>
<td>63%</td>
</tr>
</tbody>
</table>

Sustainability

University of New Hampshire:
“We define sustainability as what sustains us as diverse people and communities—from clean air and water to healthcare, education and art—and making decisions in our individual and collective lives with this big picture in mind.

Sustainability is both local and global. It requires of us that we consider both the past and the future in terms of current and best practices.

At UNH, we use the sustainable learning community model developed by Dr. Tom Kelly. Sustainability involves maintaining the long-term health of biodiversity, climate, food, and culture, and where these four systems interact. “

https://sustainableunh.unh.edu/whatissustainability
Sustainability of software, teams…

hard skills and soft skills

with researchers, institutions’ key people, funding bodies, community, …
Sustainability of software, teams...

Climate between Research and Facilitation

Culture: Innovation vs. Maintenance

hard skills and soft skills

software, algorithms, domain research

Diversity in STEM

Funding and resources

with researchers, institutions’ key people, funding bodies, community, ...
“After all, usability really just means that making sure that something works well: that a person … can use the thing - whether it's a Web site, a fighter jet, or a revolving door - for its intended purpose without getting hopelessly frustrated.”

(Steve Krug in “Don't make me think!: A Common Sense Approach to Web Usability”, 2005)
Re-Usability

- Sharing of knowledge and data
- Re-Using of „recipes“ and workflows
- Re-Usability of software

“The key to productivity is reusability. The easiest way to produce code is obviously to have it already!”

Reproducibility

- **rerun**: variations on experiment and set up
- **repeat**: same experiment, same set up, same lab
- **replicate**: same experiment, same set up, independent lab
- **reproduce**: variations on experiment, on set up, independent labs
- **reuse**: different experiment

**Computational analyses?**

**Repeatability:**
- Same experiment
- Same set up
- Same result

**Reproducibility:**
- Similar experiment
- Similar set up
- Similar result
- > 1 lab
- > 1 experiment
Researchers
“the local academic community struggles to effectively manage its assets which manifested itself in a number of challenges, and as for researchers, they lacked storage capacity and data curation processes, and the institution lacked standard metadata and indexing technologies, as well as tools that would support the whole research workflow” - Digital Asset Strategy Committee, DigitalND, 2011

Libraries
Typically, data curation happens retroactively, and as a result data is either not captured at all or available metadata is incomplete.

Pressures from the Outside
“...digitally formatted scientific data resulting from unclassified research supported wholly or in part should be stored and publicly accessible to search, retrieve, and analyze.” - White House OSTP Public Access Memo, Feb. 2013
Technology-Enhanced Research

- Increased complexity of today’s research questions
- Hardware and software skills required
- Greater need for openness and reproducibility
  - Science increasingly driving policy questions
- Opportunity to integrate research with teaching
  - Better workforce preparation

We need end-to-end solutions that provide broad access to advanced resources and allow all to tackle today’s challenging science questions

➡️ Science Gateways
Data- and compute-intensive problems

Web-based agile frameworks

Distributed data and computing infrastructures

Users generally not IT specialists

Tools and workflow engines

High-speed networks
Data and compute-intensive problems

Web-based agile frameworks

Distributed data and computing infrastructures

Need for science gateways!

Tools and workflow engines

Users generally not IT specialists

High-speed networks
Data and compute-intensive problems

Web-based agile frameworks

Distributed data and computing infrastructures

Tools and workflow engines

Users generally not IT specialists

High-speed networks
It’s a Fan!

It’s a Wall!

It’s a Rope!

It’s a Spear!

It’s a Snake!

It’s a Tree!
It's a Science Gateway

It's a Research Portal

It’s a Virtual Research Environment

It’s a Cyber-infrastructure

It’s a Collaboratory

It’s a Virtual Lab
XSEDE Users

Graph showing the number of users from March 2014 to December 2018.

- Open accounts
- Active + Gateway
- Active users
- Gateway users
- New HPC users
- New XUP accounts

Key:
- All users
- Gateways
- Login users

XSEDE user counts
Initiatives and Projects

- Science Gateways Community Institute
- US Research Software Sustainability Institute
- UK Software Sustainability Institute
- US Research Software Engineer Association
- International Research Software Engineer Associations
- Better Scientific Software
- ACI-REF Virtual Residency
- Campus Research Computing Consortium (CaRCC)
- CyberAmbassadors
- SIGHPC Education Chapter
- Software & Data Carpentry
**Elements**: Small groups - create & deploy robust capabilities for demonstrated need to advance science & engineering.

**Framework Implementations**: Larger teams organized around the development and application of common infrastructure aimed at solving common research problems, resulting in a sustainable community framework serving a diverse community or communities.

**Planning Grants for Community Cyberinfrastructure**: Focus on long-term capabilities in cyberinfrastructure to serve a research community of substantial size and disciplinary breadth.

**Community Cyberinfrastructure Implementations**: Focus on long-term hubs of excellence in cyberinfrastructure and technologies, to serve a research community of substantial size and disciplinary breadth.

SI2
Software Infrastructure for Sustained innovation

CSSI
Cyberinfrastructure for Sustained Scientific Innovation
Science Gateways Community Institute

- Diverse expertise on demand
- Longer term support engagements
- Software and visibility for gateways
- Information exchange in a community environment
- Student opportunities and more stable career paths
On-Campus Teams

• Great visibility for the institution’s research activities
• Synergy effects between projects
• Shared resources, costs and expertise across departments
• Lower learning curves
• Expertise that is otherwise difficult for individual projects to obtain

Collaborations make challenges less steep
Gateway Ambassadors

Are you enthusiastic about sharing advanced software or digital products for research or teaching with a community?

Do you see the need for such a community at your institution?
Gateway Ambassadors

Community builders
• creating a network between people, experts, and resources

Members of a professional community
• Community activities
  • learning from peers
  • sharing information

• Ambassador activities
  • meeting with individuals
  • hosting awareness sessions
• Functioning of the individual and team
• Functioning of the research software
• Functioning of the research field itself

Developing a pathway to research software sustainability
Functioning of the Individual and Team

- Training & education
- Ensuring appropriate credit for software development
- Enabling publication pathways for research software
- Fostering satisfactory and rewarding career paths for people who develop and maintain software
- Increasing the participation of underrepresented groups in software engineering
Functioning of Research Software

• Supporting sustainability of the software
• Growing community, evolving governance, and developing relationships between organizations, both academic and industrial
• Fostering both testing and reproducibility
• Supporting new models and developments (e.g., agile web frameworks, Software-as-a-Service)
• Supporting contributions of transient contributors (e.g., students)
Functioning of the Research Field Itself

- Growing communities around research software and disparate user requirements
- Cataloging extant and necessary software
- Disseminating new developments
- Training researchers in the usage of software
- Understanding and improving pipelines of diverse developers and maintainers
## Service Areas

<table>
<thead>
<tr>
<th>Development Support (consulting &amp; short term small project support)</th>
<th>Supporting Software</th>
<th>Supporting People</th>
<th>Supporting the community</th>
<th>Science &amp; research impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incubator (technology advice, business planning, usability advice, etc.)</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Training (courses &amp; guides)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Policy (research &amp; campaigns)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Community (fellowships, workshops, blogs, website)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
What is an RSE?

We like an inclusive definition of Research Software Engineers to encompass those who regularly use expertise in programming to advance research. This includes researchers who spend a significant amount of time programming, full-time software engineers writing code to solve research problems, and those somewhere in-between.
Mission of US-RSE

Community
• Create a professional community to share knowledge, connections, and resources.

Advocacy
• Promote RSEs impact on research, highlighting the critical and valuable role RSEs serve.

Resources
• Access to information and material to support individuals and RSE groups
Lessons Learned from Sustained Projects

Commonalities

- Evangelist
- Diverse mechanisms of funding
- Community building
- Open source and open science
- Collaboration, collaboration, collaboration
• Evangelists for diverse initiatives – not every approach suits all
• Define a roadmap for collaboration and community building
Contact

sandra.gesing@nd.edu

http://sandra-gesing.com/