Democratizing Data

#GCCBOSC 2018

Tracy K. Teal, PhD
The Carpentries, Executive Director
Think – Pair

What makes you proud or excited to work in bioinformatics?
Share

What ideas or themes came up?

Enter them here:
https://tinyurl.com/bosc2018-1
Share

Word cloud results
Think - Pair

What frustrates you about working in bioinformatics?
Share

What ideas or themes came up?

Enter them here:
https://tinyurl.com/bosc2018-2
Share

Word cloud results
Themes

• We’re excited about the impact bioinformatics can have
• We like solving interesting problems
• We are frustrated by the environment where we do the work
• We are frustrated by how we can work and be valued for our role and work
• We are disappointed in who is able to do this work
Data and tools are not what limits progress...
but how and who does the work
Research driven software

Software developed in pursuit of a research question
<table>
<thead>
<tr>
<th>Software as a Service</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of people with problem</td>
<td>Many</td>
</tr>
<tr>
<td>Understanding of problem</td>
<td>Good</td>
</tr>
<tr>
<td>Need for users to understand software implementation</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Software as a Service</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Number of people with problem</td>
<td>Many</td>
</tr>
<tr>
<td>Understanding of problem</td>
<td>Good</td>
</tr>
<tr>
<td>Need for users to understand software implementation</td>
<td>Low</td>
</tr>
</tbody>
</table>
Software as a Service (Not)
We must value the skills and the people as integral to the research process.
In a world of almost infinite data, code and software is what turns data into information & knowledge
If we in academia want to stay in the discovery game, we cannot outsource our software development or have our people’s best career options be analysis of click rates.
What can we do now to support software not as a service?
1. Build local talent

2. Support development

3. Collaborate
Build local talent
Current researchers

- Understand the research problem
- Have the ability to learn new things
- Motivated to answer the questions
- Have computational skills
Most useful thing Bioinformatics Resource Australia could do is offer training
Current unmet needs

Training in the gaps
Training in the Gaps needs to be...

The 4 A’s

• Accessible
• Approachable
• Aligned
• Applicable
Training in the Gaps

- Workshops
- Short courses
- Self-guided learning
- MOOCs
- Integration into existing courses
Community organization training researchers in best practices for data analysis and software development to make research more effective and reproducible.

- Core skills for effective research computing
- Trained instructors
- Two-day hands-on workshops
- Collaboratively developed, openly licensed lesson materials
- Over 1300 trained volunteer instructors on 6 continents
• Core skills for effective research computing
• Trained instructors
• Two-day hands-on workshops
• Collaboratively developed, openly licensed lesson materials
Our Workshops. Our Learners.
Our Reach.
Our Impact.

According to Programmatic & Long-Term Survey Respondents

- 1,480 Instructors badged
- 1,332 Workshops taught
- 37,000 Learners reached
- 44 countries
- ...and growing every year!

- 77% are more comfortable with tools
- 54% have made their analysis more reproducible
- 65% have gained confidence working with data
- 74% have recommended The Carpentries workshops to friends or colleagues.
Democratizing data skills
Democratizing data skills

- Answer more questions
- Engage a more diverse set of researchers in computational analysis
- More creative approaches and questions
- Everyone wins!
Our “money ball” is not identifying underappreciated talent, but creating a program that builds talent
2. Supporting development
Why is git telling me I have a detached head!?
Git makes sense to me now that I understand the theory that underlies it!
This is Git. It tracks collaborative work on projects through a beautiful distributed graph theory tree model.

Cool. How do we use it?

No idea. Just memorize these shell commands and type them to sync up. If you get errors, save your work elsewhere, delete the project, and download a fresh copy.
Supporting development

• Software development consulting services
• Information or access to resources on best/better practices (Checklists! Templates!)
• Opportunities to learn from each other (paired programming, lab code review)
• Valuing the time it takes to learn the skills and do software development well
• Giving credit for software & good software development practices
The Journal of Open Source Software (JOSS) is a developer friendly journal for research software packages.

Author Guidelines

If you've already licensed your code and have good documentation then we expect that it should take less than an hour to prepare and submit your paper to JOSS.
Software development consulting services and resources

“Software Project Carpentry”
3. Collaborate
If you want to go fast, go alone.
If you want to go far, go together.
If you want to go fast, go alone.
If you want to go in a lot of different uncoordinated directions at once, go together.
Collaborate

• Open code
• Open data
• Communication
  – Shared knowledge
  – Ideas from different perspectives
• Building a community of practice: in our labs, at our universities, across universities
Software as a Service

1. Build local talent
2. Support development
3. Collaborate
If you want to go fast, go alone.
If you want to go far, go together.