Developing bioinformatics community and infrastructure in a diverse research environment

Toby Hodges
Zeller Team
European Molecular Biology Laboratory

🔗 tobyhodges
✉️ toby.hodges@embl.de
🐦 @tbyhdgs

https://bio-it.embl.de
“Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, bioinformatics combines computer science, statistics, mathematics, and engineering to analyze and interpret biological data.”

Bioinformatics is a Growing Field

Increasing Data Volume - EMBL-EBI Databases
Bioinformatics is a Growing Field

Increasing Data Volume - EMBL-EBI Databases

Requests per day - EMBL-EBI web services
Example 1: GitLab - git.embl.de

Setup

• migrated from gitolite in 2015
• 1000s lines to migrate
• served on a locally-managed VM

Benefits

• encouraged good development practices (version control & docs)
• internal & external collaboration
• easier sharing of code and teaching materials
• data privacy & security
Example 2: Bioinformatics Training

- volunteer-driven training
- topics suggested by volunteers & community
- ~500 total attendees since Dec ‘14

Aggregated Feedback (since Dec 2014)

Would recommend to colleague

Courses & Workshops

Year

2012  2013  2014  2015  2016  2017*
Project Organisation

Community
Project Organisation

- Dedicated Personnel
- Computational Centres
- Support Staff
- Volunteers

Taskforce
Contributors
Community
Project Organisation

Dedicated Personnel
Computational Centres
Support Staff
Volunteers

Taskforce
Contributors
Community

Bio-IT Activities
- Training & Consulting
- Resource Coordination
- Information Dissemination
- Community Building
How did it all start?

Motivating Factors
- 2009: First bioinformatics survey
- 2010: SCB unit review

Preceding activity
- 1990s: 'Hot Topics, Cold Beer' (HUB precursor)
- 2001: First biocomputing retreat
- 2008: First information web portal

Training & Consulting
Resource Coordination
Information Dissemination
Community Building

Strategic Head of Bioinformatics, project coordinator, and 0.5 FTE assigned
HUB series launched
Beer sessions launched
Q&A forum launched
Bio-IT Portal redeveloped

<table>
<thead>
<tr>
<th>Year</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio-IT courses</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>11</td>
<td>9*</td>
<td></td>
</tr>
<tr>
<td>Taskforce established</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unified Linux software repository created</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bio-IT Portal redeveloped
Team Leader with Bio-IT mandate appointed
3 new computational Centres
EMBL GitLab migration & launch
Coding club established
Chat system launched
Team Leader with Bio-IT mandate appointed
Bio-IT Portal redeveloped

Aidan Budd
What makes it work? - Structural

- People who want to be involved
- Collaborative spirit
- Critical mass
- Organisational support
- Flexibility (time, budget, personal)
- Dedicated “core” staff
What makes it work? - Active development

Connect
Encourage
Enable
Acknowledge
Advice

- Don’t try to get people to do things they don’t want to do
- Take a “user”-focussed approach from the start
- Start collecting data
- Welcome contributions
- Allow time for community to get established
Acknowledgments

- Malvika Sharan
- Aidan Budd
- Georg Zeller
- Frank Thommen, Jean-Karim Hériché, Venkata Satagopam, Reinhard Schneider
- Clemens Lakner
- Holger Dinkel, Grischa Toedt
- Charles Girardot, Jelle Scholtalbers, Markus Fritz
- Thomas Hoffmann
- Yan P Yuan
- Marc Gouw
- Tom Schwarzl
- Mike Smith
- Christian Arnold
- Sergej Andrejev
- Luis Pedro Coelho
- Karin Sasaki
- Jonas Hartmann
- Vitaly Kovalev, Artem Tarasov, Ivan Protsyuk
- Laurent Gatto, Robert Stojnic
- Michael R Crusoe
- Michael Wahlers, Jure Pečar, Carlos Fernandes
  San Millan, EMBL IT Services
- Eva-Maria Geissen, Matt Rogon, Bernd Klaus
- Konrad Zych, Alessio Milanese, Jakob Wirbel, Nicolai Karcher, Alex Mattausch
- EMBL Python User Group
- HUB Organisers
- Carpentries Community
- Judith Zaugg, Kiran Patil
- Wolfgang Huber
- Toby Gibson
- Peer Bork

https://bio-it.embl.de