Sharing software, tools & research with Github in hydrosciences

Cas Neyens
23 October 2019
Mini Open Science Fair
Reproducible research

*IS THERE A REPRODUCIBILITY CRISIS?*

- 3% Don’t know
- 52% Yes, a significant crisis
- 38% Yes, a slight crisis
- 7% No, there is no crisis

1,576 researchers surveyed

*Source: Nature, 2016*
Reproducible research

**IS THERE A REPRODUCIBILITY CRISIS?**

- 52% Yes, a significant crisis
- 7% Don’t know
- 38% Yes, a slight crisis
- 3% No, there is no crisis
- 1,576 researchers surveyed

**Source: Nature, 2016**

**HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?**

Most scientists have experienced failure to reproduce results.

<table>
<thead>
<tr>
<th>Field</th>
<th>Someone else’s</th>
<th>My own</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>70%</td>
<td>30%</td>
</tr>
<tr>
<td>Biology</td>
<td>70%</td>
<td>30%</td>
</tr>
<tr>
<td>Physics and engineering</td>
<td>60%</td>
<td>40%</td>
</tr>
<tr>
<td>Medicine</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Earth and environment</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>Other</td>
<td>30%</td>
<td>70%</td>
</tr>
</tbody>
</table>
Sharing code

• Make analysis scripted & public

• Distribute

• Further development

• Reproducibility

Reproducible research

“Most computational hydrology is not reproducible, so is it really science?”

Hutton et al., 2016
Reproducible research

“Most computational hydrology is not reproducible, so is it really science?”

Hutton et al., 2016

• Can be improved by scripted analysis & open source sharing
HYDR code

- HYDR has developed:
  - Software
  - Tools
  - Publications

Source: VUB
GitHub

• Online code hosting

• Open source

• Used for sharing & collaborating on code
GitHub

- Uses Git software at its core

- Version control

- “Keeping track of who changed what where”
GitHub structure

User

Repo A
- Code
- Metadata

Repo B
- Code
- Metadata

Cas Neyens
creynens

Hydrologist - PhD researcher at VUB, Department of Hydrology and Hydraulic Engineering. Working on reactive transport models.

At: VUB
Brussels, Belgium

32 contributions in the last year
GitHub structure

User

Repo A
  - Code
  - Metadata

Repo B
  - Code
  - Metadata
HYDR GitHub

- GitHub organization
- Members
- Centralized repo's
- Easy to collaborate and discuss

https://github.com/VUB-HYDR
HYDR GitHub

- Software
- Stand-alone software
HYDR GitHub

- **Tools**
- **Simple scripts**
• Publications
• Data analysis, not data
• Collaboration
  • Git = version control

Source: McWilliams, 2013
Pros & cons

Pro:

• Reproducible research
• Code/tool sharing
• Collaboration → improvement
• Analysis should be scripted
Pros & cons

Pro:

• Reproducible research
• Code/tool sharing
• Collaboration → improvement
• Analysis should be scripted

Con:

• Git has a steep learning curve
• Maintenance
• Analysis should be scripted
• Needs quality control
Questions?

cas.neyens@vub.be