The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Reproducible Research

AND WHY TO DO IT...

Lukasz K Bonenberg

10th May 2016

NGI

Table of Contents[1]

1. Introduction

2. git - Version Control

3. What can we use it for?

4. Reproducible research

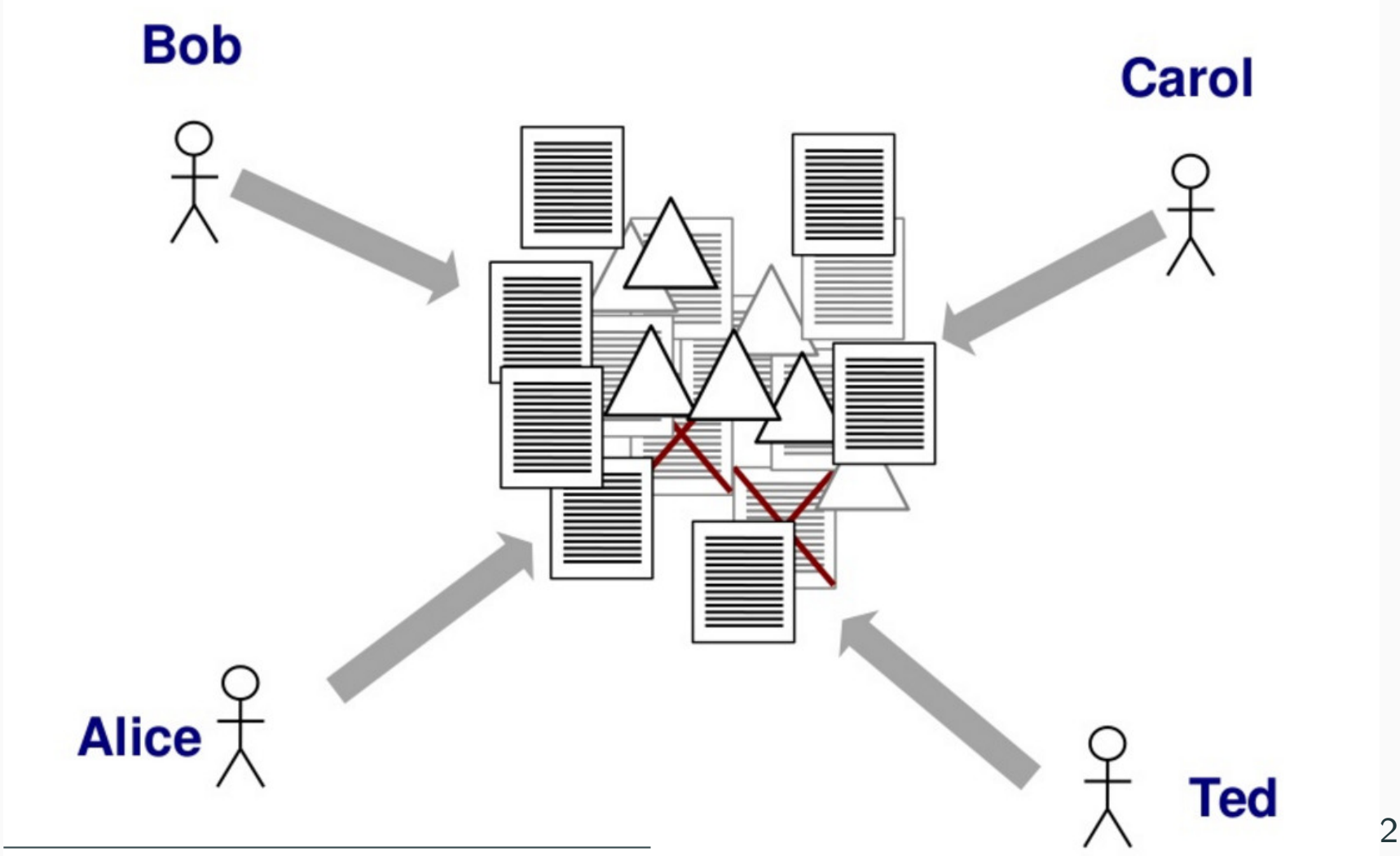
5. Wrap up

[1] History of changes at `https://github.com/DfAC/TeachingSlides/`.

Introduction

- report_01.doc
- report_02.doc
- report_03_revByJim.doc
- report_04_changes.doc
- report_05_final.doc
- report_05_finalFinal.doc
- report_05_finalFinal_FINAL.doc
- report_05_finalFinal_FINAL_send.doc

Entropy

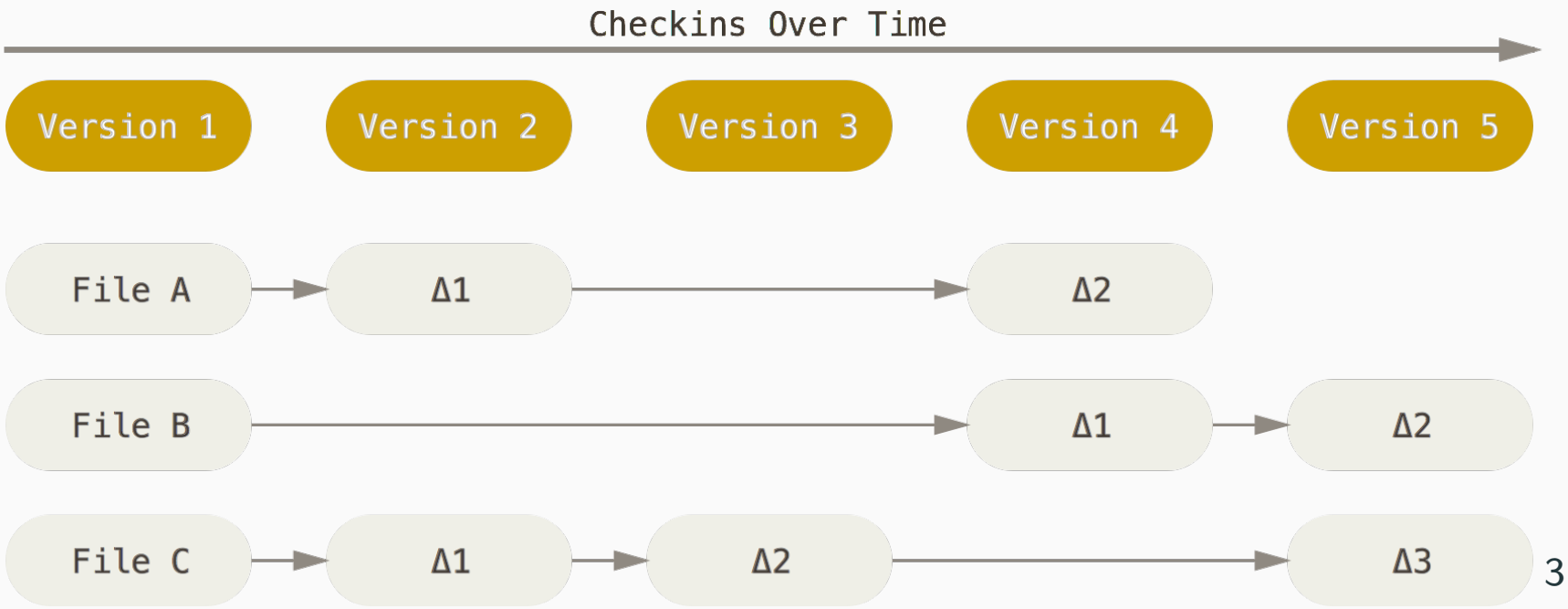


[2]

[2] http://www.slideshare.net/jomikr/quick-introduction-to-git

git - Version Control

Control the time



3

[3] https://git-scm.com/book/ch1-3.html

Examples

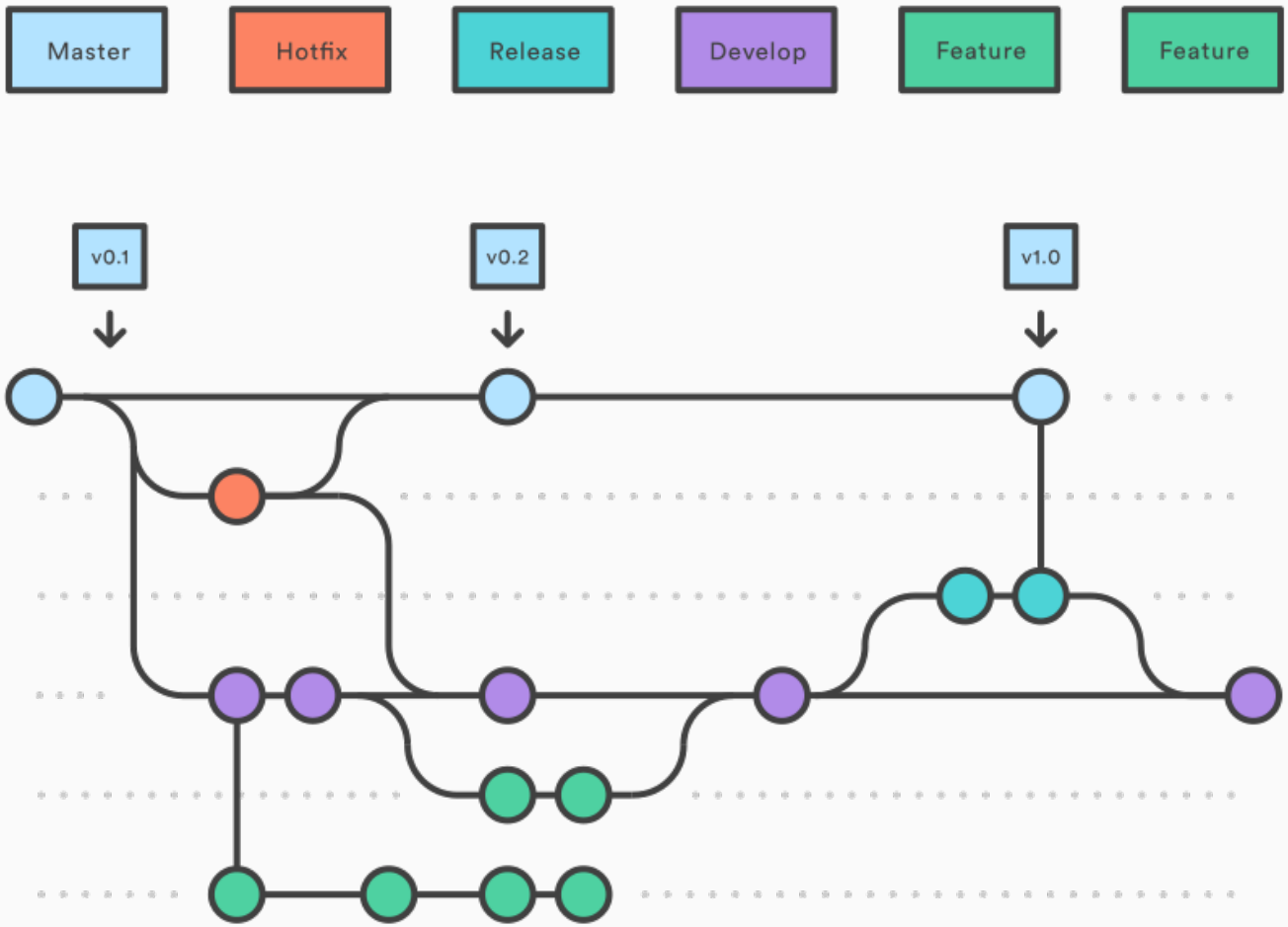
- https://github.com/tomojitakasu/RTKLIB
- https://github.com/DfAC/TEQCSPEC

Lets try it

- http://rogerdudler.github.io/git-guide/
- https://try.github.io/levels/1/challenges/10

What can we use it for?

Team work



[4]

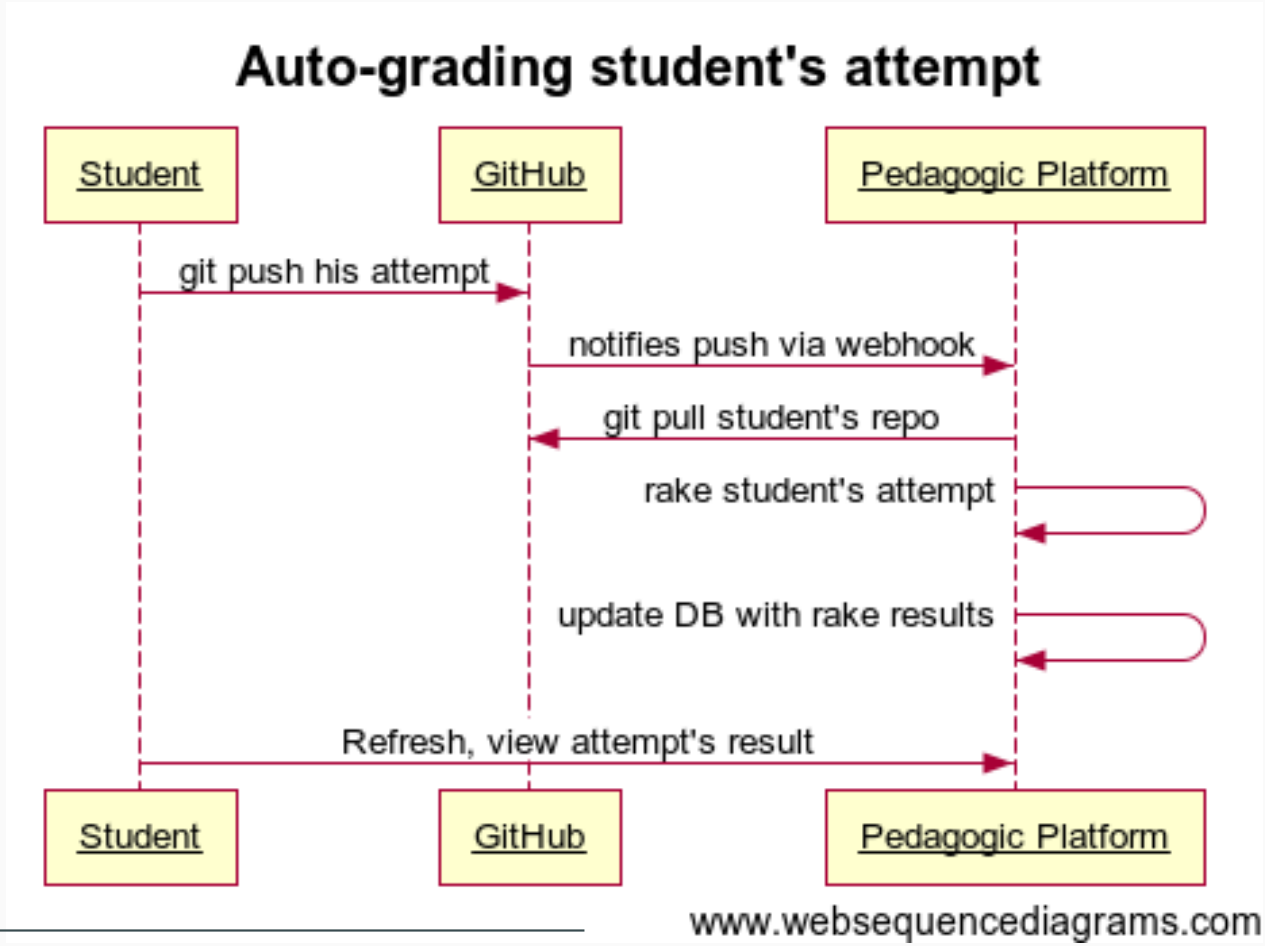
[4] https://www.atlassian.com/git/tutorials/

More teaching

- https://github.com/FOSS4GAcademy
- https://github.com/DfAC/TeachingSlides

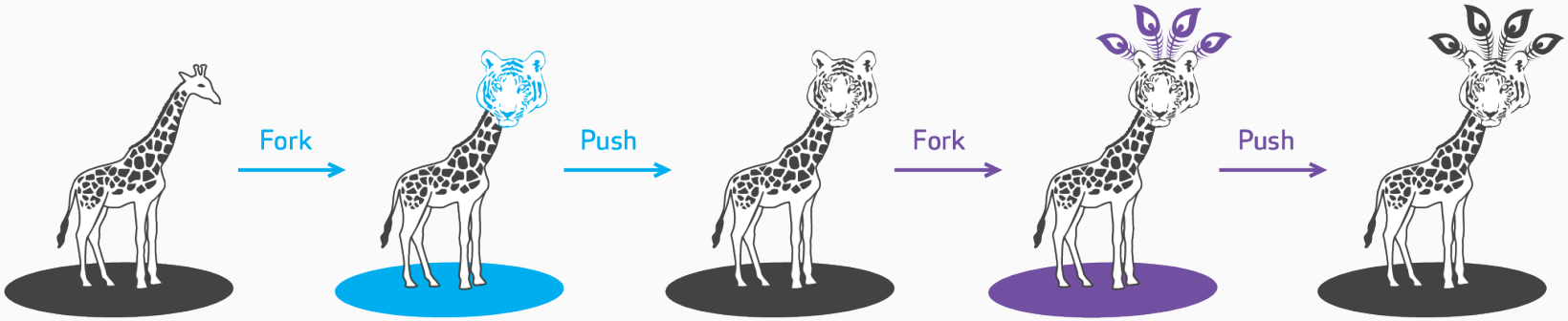
More teaching

Sebastien Saunier[5] discussed auto-marking using git.



[5] http://bit.ly/1MQLSo9

Open Source

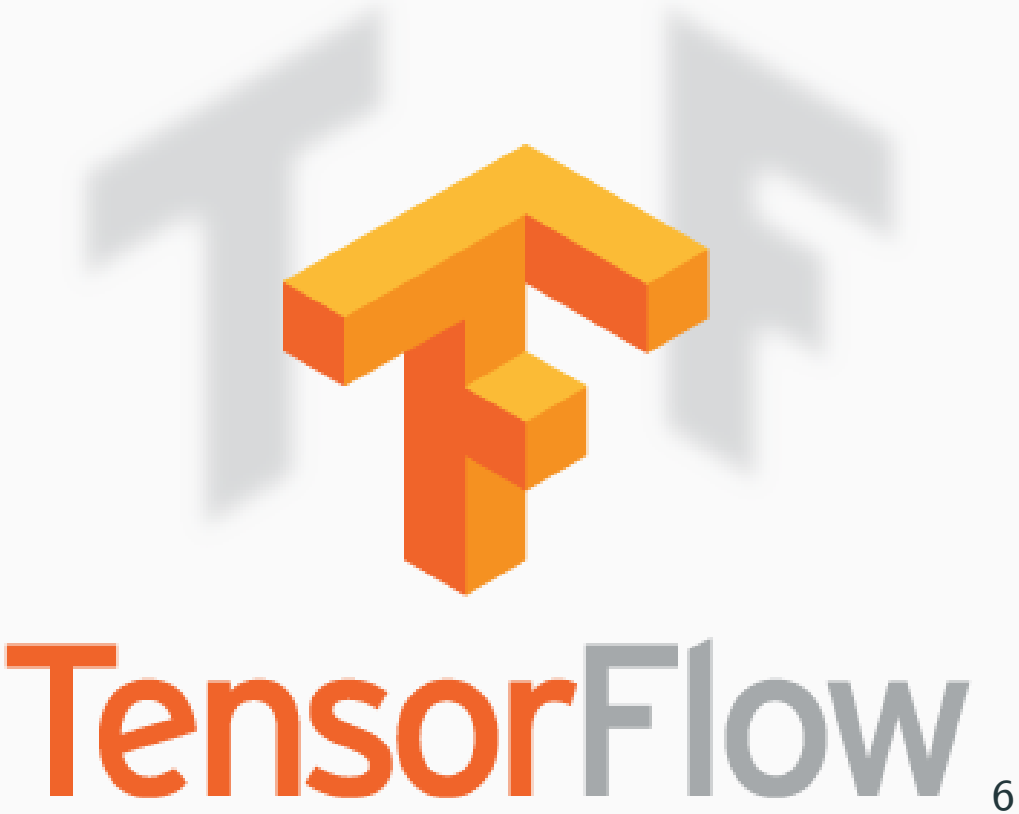


peacock feather, tiger, giraffe from The Noun Project

Social space

- https://github.com/tomojitakasu/RTKLIB/issues
- https://gist.github.com/forked

Big guys' open source



[6]

[6] https://www.tensorflow.org/

Reproducible research

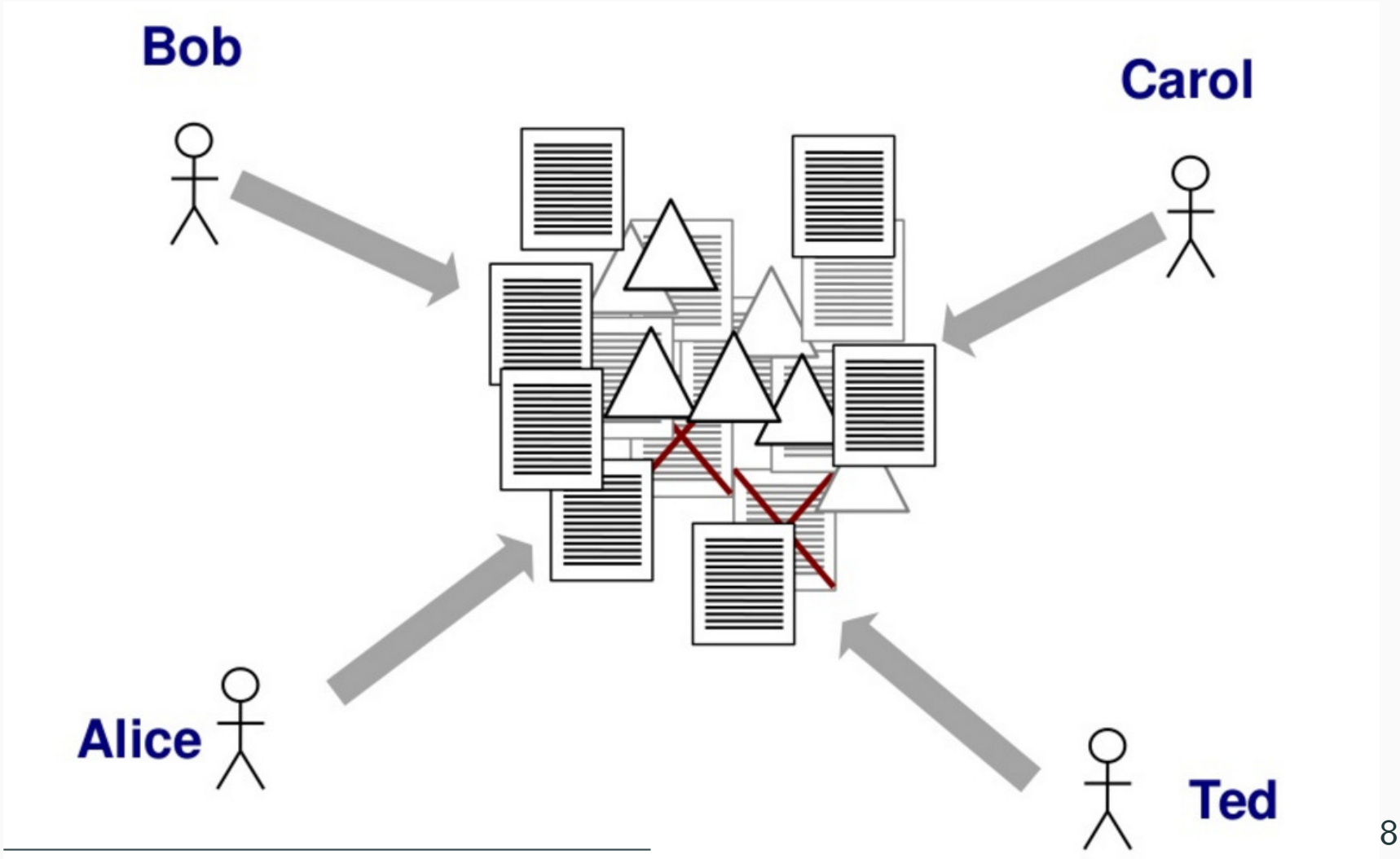
Reproducible research

Reproducible research is the idea that data analyses, and more generally, scientific claims, are published with their data and software code so that others may verify the findings and build upon them[7].

This means that both data and the analysis code are available allowing others to easily execute the same analysis to obtain the same results.

[7] Quoted after Roger Peng (Johns Hopkins University)

Why?



[8]

[8] http://www.slideshare.net/jomikr/quick-introduction-to-git

Maintaining data

Prof Keith A. Baggerly discussed sampling errors on cancer research[9]. A number of research institutions The concept is very popular in computer science and engineering as well:

- Stanford Exploration Project - `http://sepwww.stanford.edu/`
- Computer vision - `http://www.csee.wvu.edu/~xinl/`
- Wavelab - `http://finmath.stanford.edu/~wavelab/`

There is even Coursera course on the topic[10].

[9] Cancer Bioinformatics Workshop, Cambridge 2010 - `http://bit.ly/235DoBa`
[10] `https://www.coursera.org/learn/reproducible-research`

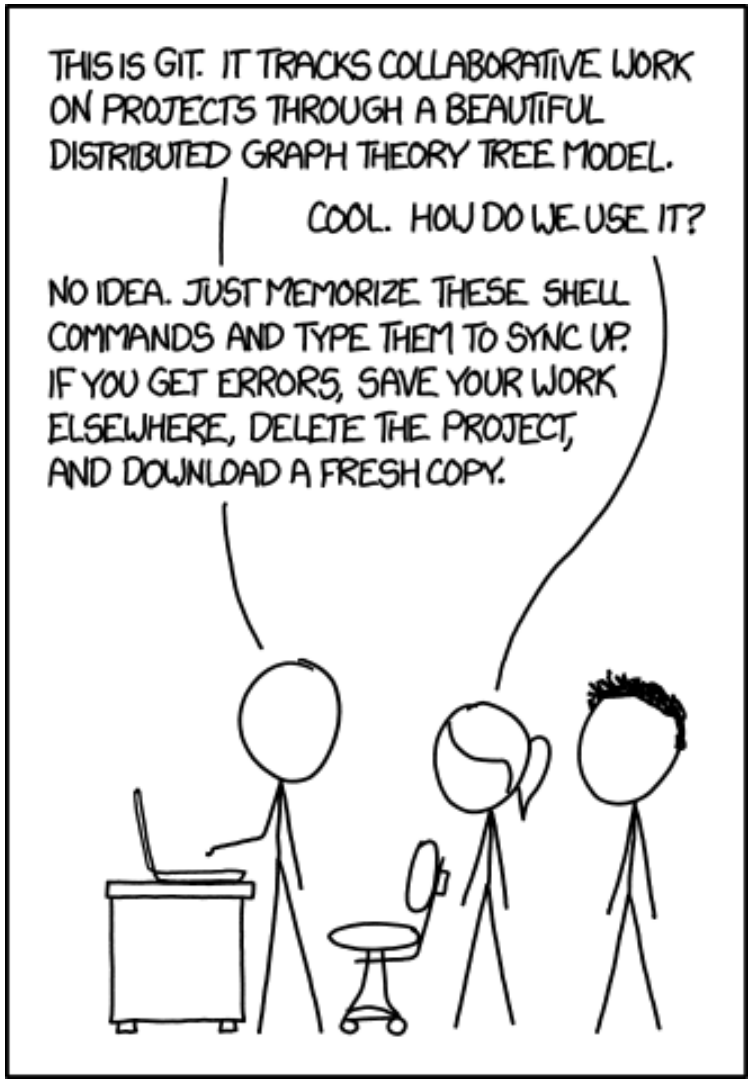
Understanding science

- Gravitational Wave - `http://bit.ly/LIGO_OS`
- Earthquakes - `http://bit.ly/1MbL6C9`
- Heroes interaction - `http://bit.ly/1RDJ4Lv`
- Faces recognition - `http://bit.ly/1XgqjxS`

Wrap up

- git
  - backup & maintain your work;
  - has application in teaching, research or learning;
  - a social tool.
- Reproducible research
  - make it easy to return to the data and analyses;
  - make it easy to share information internally & externally;
  - provide research scrutiny and transparency;
  - can be great self-marketing tool.

Thank you


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.