

# CAS Applied Data Science

## Ethics and best practices for scientific computing and data science

[cas-applied-datascience.unibe.ch](https://cas-applied-datascience.unibe.ch)

PD Dr. Sigve Haug, [sigve.haug@math.unibe.ch](mailto:sigve.haug@math.unibe.ch)

2020-06-15

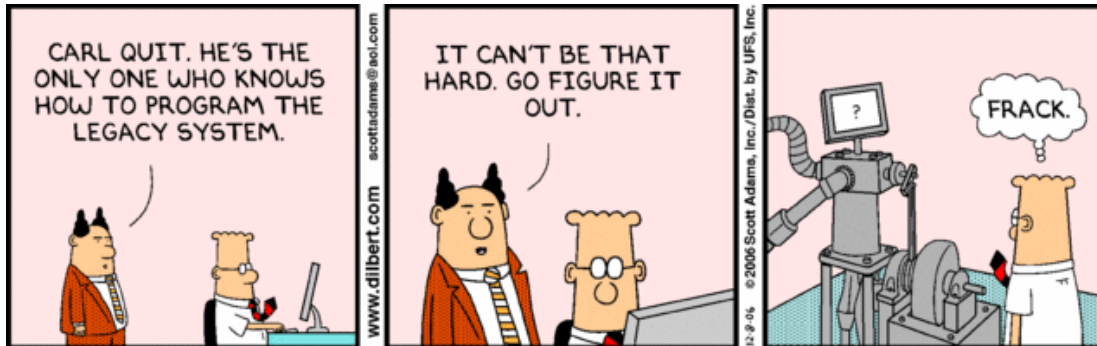
# Why ?

---

- Boeing 737 Max 2018/19
- Ariane 5 1996
- Oil and gas platform Sleipner 1991
- Therac 25 1985/87
- World war III 1983

Poor software costs billions and maybe one day the humanity

# VCS - git



# Why Version Control Software (VCS) ?

- Our research sits to a large extent in the code
- The publication is just the advertisement
- Research (code) should be reproducible
- Research (code) should be open (open science)
- Research (code) should be high quality
- Research (code) should be understandable
- Research (code) should be efficient
- Research (code) should be / is often team work
- Research (code) should be FAIR



VCS can help us with all this !

# Cyber Security

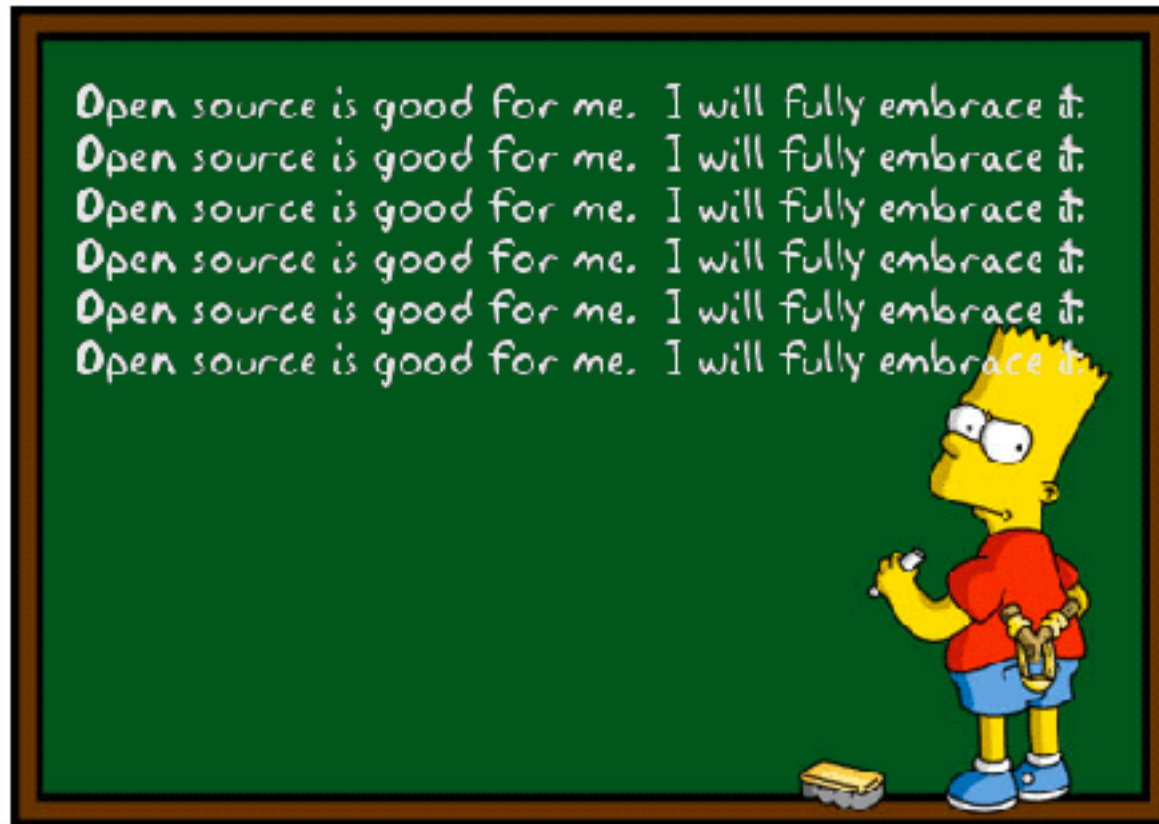
Top 25 most common passwords by year according to SplashData

Rank	2011 <sup>[4]</sup>	2012 <sup>[5]</sup>	2013 <sup>[6]</sup>	2014 <sup>[7]</sup>	2015 <sup>[8]</sup>	2016 <sup>[3]</sup>	2017 <sup>[9]</sup>	2018 <sup>[10]</sup>	2019 <sup>[11]</sup>
1	password	password	123456	123456	123456	123456	123456	123456	123456
2	123456	123456	password	password	password	password	password	password	123456789
3	12345678	12345678	12345678	12345	12345678	12345	12345678	123456789	qwerty
4	qwerty	abc123	qwerty	12345678	qwerty	12345678	qwerty	12345678	password
5	abc123	qwerty	abc123	qwerty	12345	football	12345	12345	1234567
6	monkey	monkey	123456789	123456789	123456789	qwerty	123456789	111111	12345678
7	1234567	letmein	111111	1234	football	1234567890	letmein	1234567	12345
8	letmein	dragon	1234567	baseball	1234	1234567	1234567	sunshine	iloveyou
9	trustno1	111111	iloveyou	dragon	1234567	princess	football	qwerty	111111
10	dragon	baseball	adobe123 <sup>[a]</sup>	football	baseball	1234	iloveyou	iloveyou	123123
11	baseball	iloveyou	123123	1234567	welcome	login	admin	princess	abc123
12	111111	trustno1	admin	monkey	1234567890	welcome	welcome	admin	qwerty123
13	iloveyou	1234567	1234567890	letmein	abc123	solo	monkey	welcome	1q2w3e4r
14	master	sunshine	letmein	abc123	111111	abc123	login	666666	admin
15	sunshine	master	photoshop <sup>[a]</sup>	111111	1qaz2wsx	admin	abc123	abc123	qwertyuiop

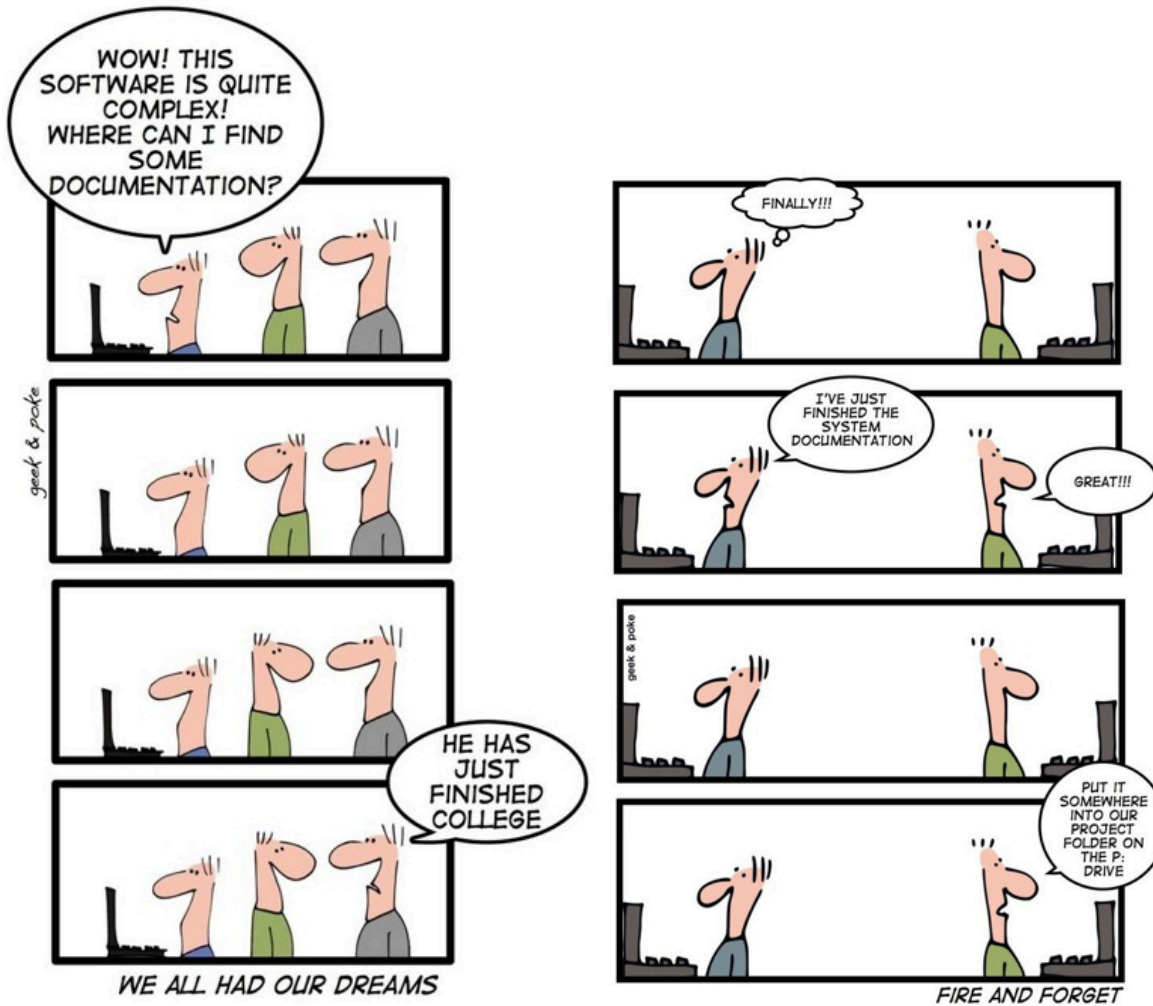
## Ranked passwords from leaks

# FOSS Legal Aspects

---



# Documentation



# Our background

---

- Who is already working with a Version Control Software (VCS) ?
- Which VCS ?
- Alone or in team ?
- Who has some cyber security training ?
- How do you document your software ?
- What do you know about legal aspects around FOSS ?



# Programme

---

## Monday

09:00 - 09:30 Introduction (PD Dr Sigve Haug, SciTS University of Bern)

09:30 - 10:30 Best practices - introduction and distribution of tasks - team work

10:30 - 11:00 Break

11:00 - 12:30 Presentations

12:30 - 13:30 Lunch

13:30 - 14:00 Software Errors - how to write good SW

14:00 - 15:00 Standards of scientific computing and data science -

Perspectives from ethics and philosophy of science (Prof. Dr. Dr. Claus Beissbart)

15:00 - 15:30 Planning of the module projects and CAS exams

## Tuesday

09:00 - 17:00 Git (Pablo Verges)

## Wednesday

09:00 - 17:00 Cyber Security (SWITCH)

## Thursday

09:00 - 12:30 Free Open Source Software - Legal Aspects

13:30 - 17:00 Documentation of Software

## Friday July 3

09:00 - 17:00 Presentation Day

17:00 - 19:00 Aperero

# Today at 14:00 - Ethics Lecture

---



## Prof. Dr. Claus Beisbart

Extraordinarius mit Schwerpunkt Wissenschaftsphilosophie

Abteilung Theoretische Philosophie

---

Telefon	<a href="tel:+41316313590">+41 31 631 35 90</a>
E-Mail	<a href="mailto:claus.beisbart@philo.unibe.ch">claus.beisbart@philo.unibe.ch</a>
Büro	Unitobler, B 223

---

# Article and task for tomorrow

OPEN ACCESS Freely available online

 **PLOS** | BIOLOGY

## Community Page

### Best Practices for Scientific Computing

**Greg Wilson<sup>1\*</sup>, D. A. Aruliah<sup>2</sup>, C. Titus Brown<sup>3</sup>, Neil P. Chue Hong<sup>4</sup>, Matt Davis<sup>5</sup>, Richard T. Guy<sup>6</sup>, Steven H. D. Haddock<sup>7</sup>, Kathryn D. Huff<sup>8</sup>, Ian M. Mitchell<sup>9</sup>, Mark D. Plumbley<sup>10</sup>, Ben Waugh<sup>11</sup>, Ethan P. White<sup>12</sup>, Paul Wilson<sup>13</sup>**

<sup>1</sup> Mozilla Foundation, Toronto, Ontario, Canada, <sup>2</sup> University of Ontario Institute of Technology, Oshawa, Ontario, Canada, <sup>3</sup> Michigan State University, East Lansing, Michigan, United States of America, <sup>4</sup> Software Sustainability Institute, Edinburgh, United Kingdom, <sup>5</sup> Space Telescope Science Institute, Baltimore, Maryland, United States of America, <sup>6</sup> University of California, Berkeley, California, United States of America, <sup>7</sup> University of California, Berkeley, California, United States of America, <sup>8</sup> University of California, Berkeley, California, United States of America, <sup>9</sup> University of California, Berkeley, California, United States of America, <sup>10</sup> University of California, Berkeley, California, United States of America, <sup>11</sup> University of California, Berkeley, California, United States of America, <sup>12</sup> University of California, Berkeley, California, United States of America, <sup>13</sup> University of California, Berkeley, California, United States of America

Available from Ilias :

[https://ilias.unibe.ch/goto\\_ilias3\\_unibe\\_file\\_1182580\\_download.html](https://ilias.unibe.ch/goto_ilias3_unibe_file_1182580_download.html)

Task for 4 groups:

- 5 minutes presentation of (own code with respect to) two best practices. Don't forget example(s)
- Mail pdf file to [h.zedan@outlook.com](mailto:h.zedan@outlook.com)

**Questions ? Enjoy !**

# Various VCS

V • T • E

## Version control software

[hide]

Years, where available, indicate the date of first stable release. Systems with names *in italics* are no longer maintained or have planned end-of-life dates.

### Local only

#### Free/open-source

RCS (1982) • SCCS (1972)

#### Proprietary

PVCS (1985) • QVCS (1991)

### Client–server

#### Free/open-source

CVS (1986, 1990 in C) • CVSNT (1998) • QVCS Enterprise (1998) • Subversion (2000)

#### Proprietary

AccuRev SCM (2002) • ClearCase (1992) • CMVC (1994) • Dimensions CM (1980s) • DSEE (1984) • Endeavor (1980s) • Integrity (2001) • Panvalet (1970s) • Perforce Helix (1995) • SCLM (1980s?) • Software Change Manager (1970s) • StarTeam (1995) • Surround SCM (2002) • Synergy (1990) • Team Concert (2008) • Team Foundation Server (2005) • Visual Studio Team Services (2014) • Vault (2003) • Visual SourceSafe (1994)

### Distributed

#### Free/open-source

ArX (2003) • BitKeeper (2000) • Codeville (2005) • Darcs (2002) • DCVS (2002) • Fossil (2007) • Git (2005) • GNU arch (2001) • GNU Bazaar (2005) • Mercurial (2005) • Monotone (2003) • SVK (2003) • Veracity (2010)

#### Proprietary

TeamWare (1990s?) • Code Co-op (1997) • Plastic SCM (2006) • Team Foundation Server (2013) • Visual Studio Team Services (2014)

### Concepts

Branch • Fork • Changeset • Commit (Gated commit) • Interleaved deltas • Delta compression • Data comparison • Merge • Repository • Tag • Trunk

Category • Comparison • List



# Why Git ?

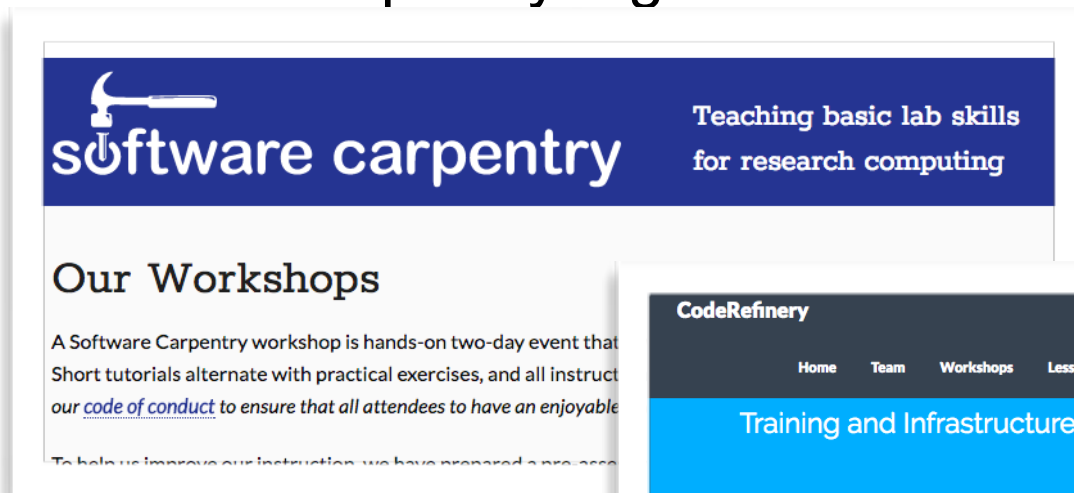
---

- Distributed
- Open source and free
- Platforms:
  - GitHub (more than 15 million users)
    - Free for public content
    - <https://octoverse.github.com/>
  - c4science (why not GitLab ?)
    - Swiss platform for science
  - CERN:
    - [gitlab.cern.ch](https://gitlab.cern.ch)
  - Biology in CH
    - [https://gitlab.isb-sib.ch/users/sign\\_in](https://gitlab.isb-sib.ch/users/sign_in)

Some UniBe groups have their own Git repos

# Efforts similar to this training

software-carpentry.org



[CodeRefinery.org](https://coderefinery.org)

