COMPUTER TOOLS FOR EFFICIENT SCIENCE

Daan van Vugt <daanvanvugt@gmail.com>

2/12/2015

TU/e

OUTLINE

Version control

Writing papers/reports

Writing code

Analyzing data

Creating figures

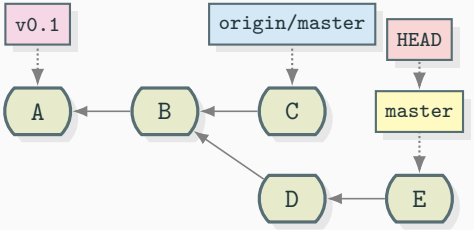
Automation

Presenting research

Searching / reading papers

VERSION CONTROL

- Git manages changes to a tree of files over time
- Distributed development, many branches
- Excellent integration with many sites and services (GitHub, GitLab, Bitbucket)



See also: Mercurial (hg), svn

> *GitHub is how people build software. With a community of more than 11 million people, developers can discover, use, and contribute to over 29 million projects using a powerful collaborative development workflow.*
> *(source: github.com/about)*

Alternatives

- Bitbucket (with free private projects)
- GitLab (self-hosted)

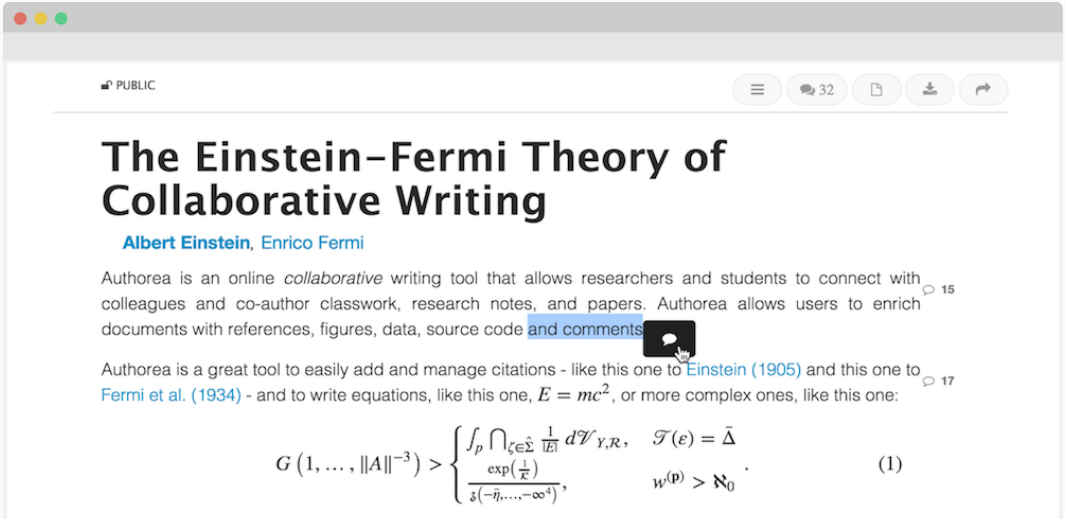
WRITING PAPERS/REPORTS

Why?

- Easy collaboration
- Version control
- Bibtex integration
- Formulas are easy
- It looks awesome
- Automation

How?

- Web-based: **Authorea**, ShareLaTeX, WriteLaTeX, StackEdit
- Windows/Mac: TeXMaker/TeXstudio/TeXshop
- Linux: Gummi, Vim, Emacs (+Auctex)



PUBLIC

32

The Einstein-Fermi Theory of Collaborative Writing

Albert Einstein, Enrico Fermi

Authorea is an online *collaborative* writing tool that allows researchers and students to connect with colleagues and co-author classwork, research notes, and papers. Authorea allows users to enrich documents with references, figures, data, source code and comments.

15

Authorea is a great tool to easily add and manage citations - like this one to Einstein (1905) and this one to Fermi et al. (1934) - and to write equations, like this one, $E = mc^2$, or more complex ones, like this one:

17

$$G\left(1, \ldots, \|A\|^{-3}\right) > \begin{cases} \int_p \bigcap_{\zeta \in \hat{\Sigma}} \frac{1}{|\mathcal{E}|} \, d\mathscr{V}_{Y,\mathcal{R}}, & \mathscr{T}(\varepsilon) = \bar{\Delta} \\ \frac{\exp\left(\frac{1}{\kappa}\right)}{\hat{s}\left(-\vec{\eta}, \ldots, -\infty^4\right)}, & w^{(\mathbf{p})} > \aleph_0 \end{cases} . \tag{1}$$

Authorea is part of the **Open Science** movement and supports **Open Access** publishing for academic research and free access to research data.

- Writefull (checks text for correct language)
- Draft & Typewrite (Real-time collaborative writing)
- Hackpad
- Etherpad
- Google Drive
- Microsoft Word (eww)

SHARING DATA ACCOMPANYING THE PAPER

- 3TU datacenter
- DataCite
- Dryad
- Figshare
- OpenScienceFramework
- Slideshare
- Zenodo

Write your code and documentation in the same place

- Jupyter (IPython/Ruby/Julia, link to Authorea, SageMathCloud)
- Mathematica
- rCharts + Slidify + shiny (R)

WRITING CODE

- **Version control!**
- Unit tests
- Documentation
- Read about best practices online
- Version control (again, because it's important)
- Check the style guide for your language/project
- Did I mention version control?

- Vim (vimtutor & vim-adventures to learn)
- Emacs
- Notepad++
- Sublime Text
- Many more, choose one you like and pimp it, add syntax highlighting etc

Many offer integration with syntax checkers and build tools

(Hard mode) Learn a better keyboard layout: (Programmer) Dvorak or Colemak

ANALYZING DATA

Good features

- Contains a unit test framework since 2013
- Some integration with git
- Nice GUI, tools like profiler and parfor

Open alternatives:

- Octave
- Python + Numpy + Scipy
- Linux tools: sed, awk, grep, gnuplot etc.
- R, paraview
- C/C++/Fortran shlib + python (for speed)

CREATING FIGURES

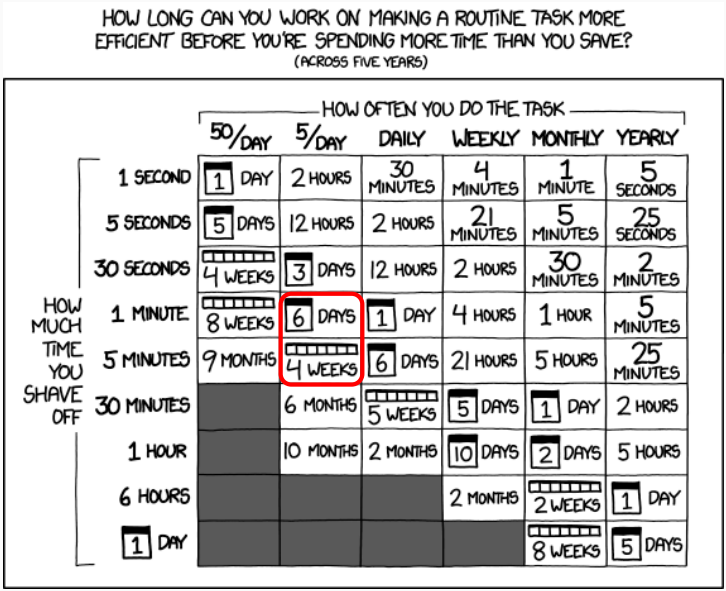
- D3.js (Interactive on webpages)
- Matplotlib (Python, + D3.js)
- Gnuplot
- MATLAB / Octave + Matlab2tikz (for LaTeX)
- Ggplot2 (R)
- Mathematica / Maple
- Paraview (3D figures)

WHY NOT EXCEL/ORIGIN?

- Not easily scriptable / automated
- Hard to create publication-quality graphics

AUTOMATION

IS IT WORTH THE TIME? (XKCD)



- GNU Make
- Bash scripting
- Ruffus (Python, computational pipelines)
- Snakemake, Pegasus (Workflow management system)
- Vistrails (Workflow and provenance management system)
- PyRDM (Research Data Management)
- Sumatra, Elabftw, Wings Workflow (Electronic Lab Notebook)
- Digital lab notebooks: Evernote, Onenote


ungzip
file: reads.1
ungzip
file: reads.2
flash_merge
trim_primers_fivep
trim_primers_threep
read_length_histogram
fastq_to_fasta
makeblastdb
gene: J
organism: rhesus_monkey
makeblastdb
gene: V
organism: rhesus_monkey
makeblastdb
gene: D
organism: rhesus_monkey
dereplicate
abpipe_igblast

PRESENTING RESEARCH

- LaTeX (poster package)
- Scribus
- Inkscape
- Powerpoint

- LaTeX Beamer class (this presentation, + Pandoc)
- Prezi
- Powerpoint

SEARCHING / READING PAPERS

Mendeley:

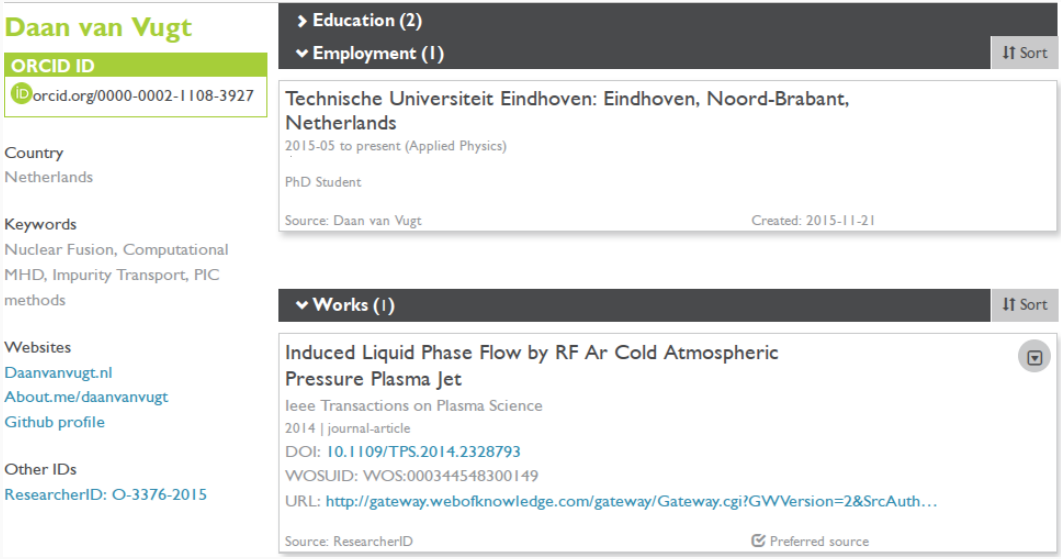
- Import papers, automatically gets name and title right
- Share libraries with colleagues
- Sync bibtex files with LaTeX
- Full-text search

OTHER TOOLS, SITES FOR FINDING PAPERS

- Web of Science
- Google scholar
- Webplotdigitizer
- CiteULike
- ResearchGate
- Scopus
- Lazyscholar.org

- ORCID
- ResearcherID
- Academia.edu
- ResearchGate
- About.me
- Twitter

ORCID



Daan van Vugt

ORCID ID

orcid.org/0000-0002-1108-3927

Country
Netherlands

Keywords
Nuclear Fusion, Computational MHD, Impurity Transport, PIC methods

Websites
Daanvanvugt.nl
About.me/daanvanvugt
Github profile

Other IDs
ResearcherID: O-3376-2015

› Education (2)

⌄ Employment (1)

Sort

Technische Universiteit Eindhoven: Eindhoven, Noord-Brabant, Netherlands
2015-05 to present (Applied Physics)

PhD Student

Source: Daan van Vugt
Created: 2015-11-21

⌄ Works (1)

Sort

Induced Liquid Phase Flow by RF Ar Cold Atmospheric Pressure Plasma Jet
Ieee Transactions on Plasma Science
2014 | journal-article
DOI: 10.1109/TPS.2014.2328793
WOSUID: WOS:000344548300149
URL: http://gateway.webofknowledge.com/gateway/Gateway.cgi?GWVersion=2&SrcAuth…

Source: ResearcherID
Preferred source

NOW IT IS YOUR TURN!

- Read 5-10 minutes about some of these programs
- Try one (or a few)
- Let me know how it goes, and if you find something interesting
- View the source of this presentation on https://github.com/Exteris/tools-for-science