Computer Tools for Efficient Science

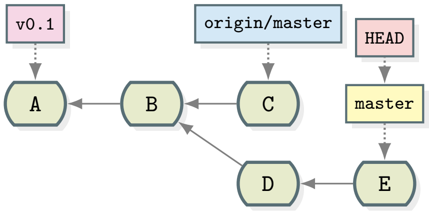
Daan van Vugt daanvanvugt@gmail.com

2/12/2015

Version control

Git

- 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

> 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

Use LaTeX

Why?

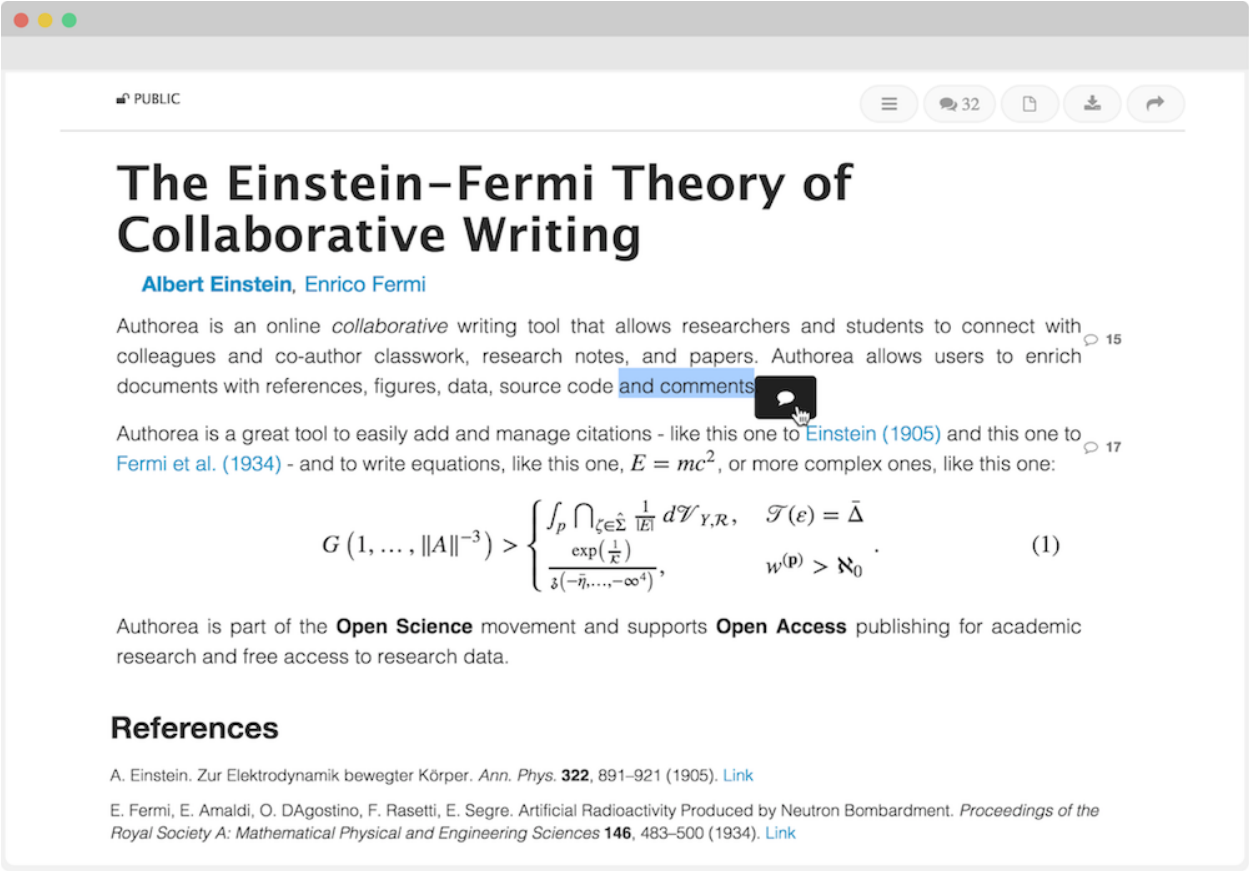
- Easy collaboration
- Version control

- Bibtex integration
- Formulas are easy
- It looks awesome
- Automation

How?

- Web-based: **Authorea**, ShareLaTeX, WriteLaTeX, StackEdit
- Windows/Mac: TeXstudio/TeXshop
- Linux: Gummi, Vim/Emacs

Authorea



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

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:

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

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

References

A. Einstein. Zur Elektrodynamik bewegter Körper. *Ann. Phys.* **322**, 891–921 (1905). Link

E. Fermi, E. Amaldi, O. DAgostino, F. Rasetti, E. Segre. Artificial Radioactivity Produced by Neutron Bombardment. *Proceedings of the Royal Society A: Mathematical Physical and Engineering Sciences* **146**, 483–500 (1934). Link

Alternatives to LaTeX

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

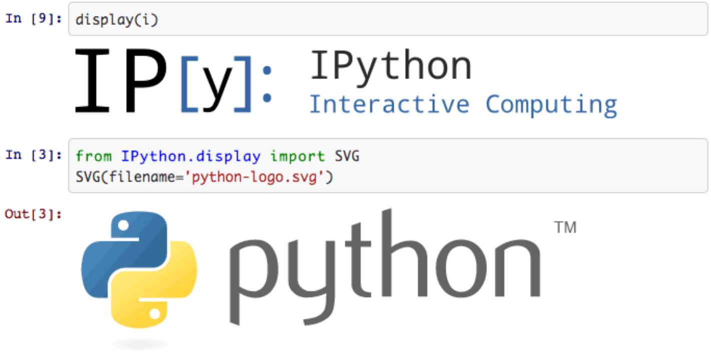
Sharing data accompanying the paper

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

Interactive notebooks

Write your code and documentation in the same place

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



```
In [9]: display(i)
```

```
In [3]: from IPython.display import SVG
        SVG(filename='python-logo.svg')
```

Writing code

Best practices

- **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?

Editors

- Vim (vimtutor & vim-adventures to learn)

- Notepad++
- Emacs
- 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

MATLAB

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

Creating figures

Tools for generating graphics

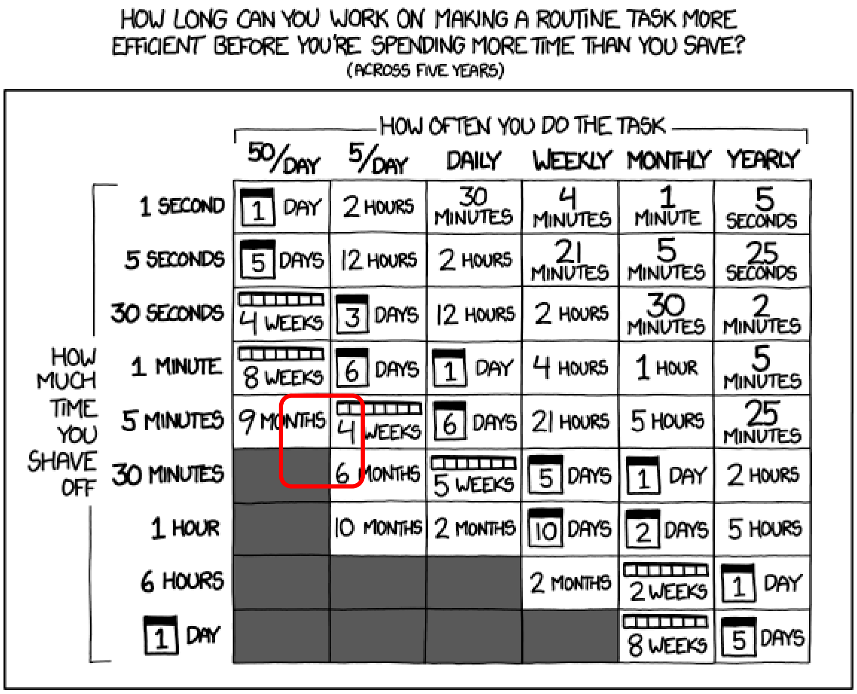
- D3.js (Interactive on webpages)
- Matplotlib (Python, + D3.js)
- Gnuplot
- MATLAB / Octave + Matlab2tikz (for LaTeX)
- Ggplot2 (R)
- Mathematica / Maple
- Paraview (3D figures)

Why not Excel/Origin?

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

Automation

Is It Worth The Time? (XKCD)


HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?
(ACROSS FIVE YEARS)

HOW MUCH TIME YOU SHAVE OFF	HOW OFTEN YOU DO THE TASK					
	50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
30 MINUTES		6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS
1 HOUR		10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS
6 HOURS				2 MONTHS	2 WEEKS	1 DAY
1 DAY					8 WEEKS	5 DAYS

How to automate

- Makefiles
- 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

Presenting research

Posters

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

Presentations

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

Searching / reading papers

Mendeley/Zotero/JabRef/Bibdesk

Mendeley:

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

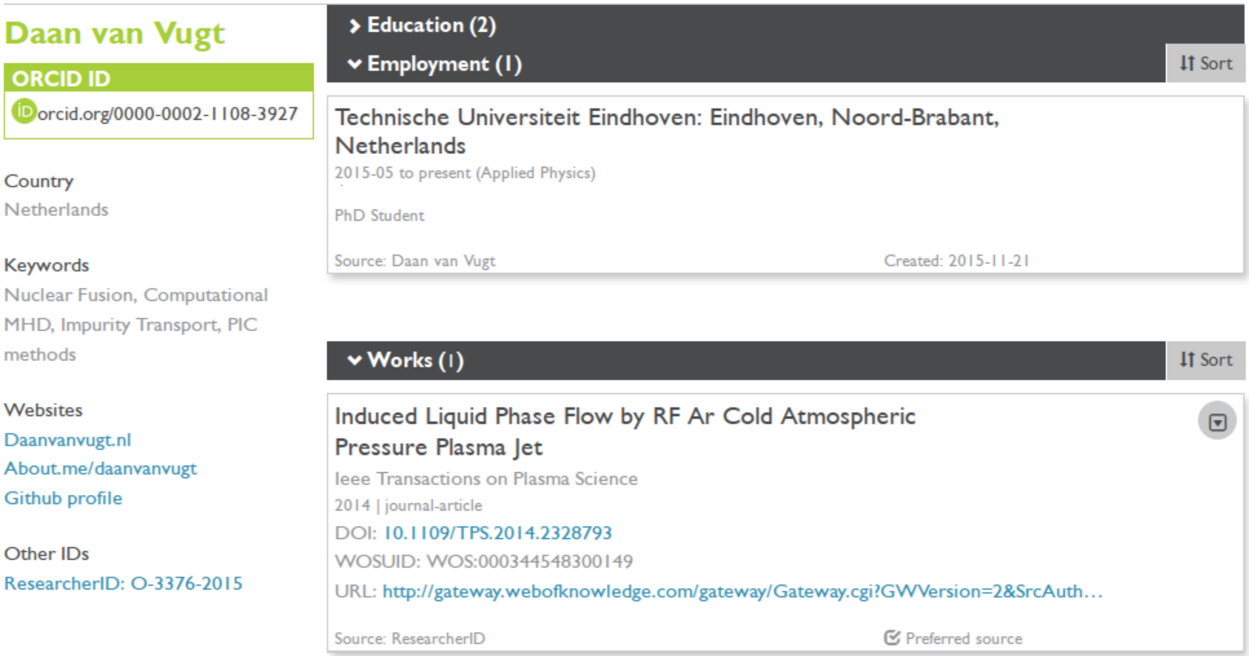
Other tools, sites for finding papers

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

Online presence

- 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)

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)

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