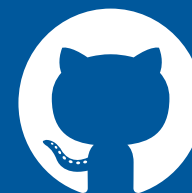


Freie Software in der Wissenschaft: Prototypen, Reproduzierbarkeit, Infrastruktur

A. Huebl

ORCID:0000-0003-1943-7141



@ax3l

Helmholtz-Zentrum Dresden - Rossendorf

Technische Universität Dresden

Ringvorlesung WS2018

Freie Software und Freies Wissen als Beruf

Dresden, November 27th 2018



HZDR



HELMHOLTZ
ZENTRUM DRESDEN
ROSSENDORF

Lecture Content

- Prototypes
- Reproducibility
 - Crediting
 - Decision Processes
- Infrastructure
 - Digital Scientific Tools
 - Productivity
- Self-Help: Licenses, FOSS & OA as a student

Open Science!

- Machines, Blueprints, Setups
- Data
- Workflows, Algorithms
- Presentations & Posters
- Manuals, Thesis, Papers
- **People!**
- Code: snippets, libraries, frameworks, applications

My Background

Natural Sciences

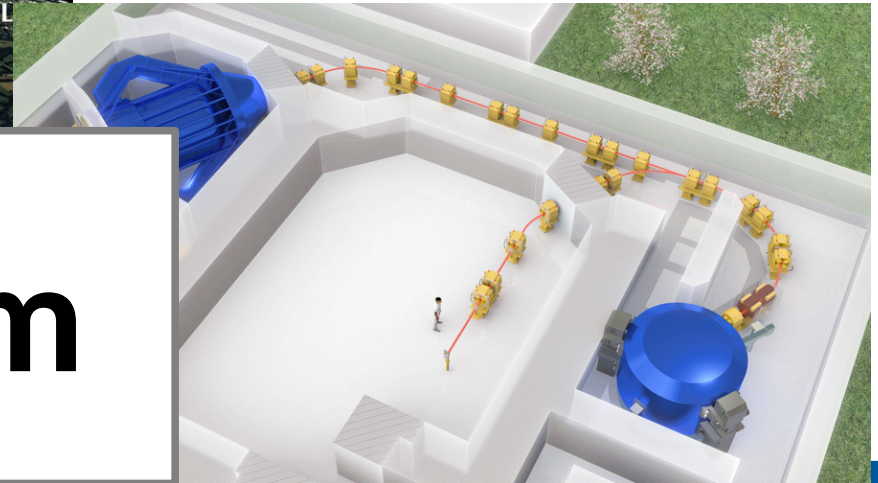
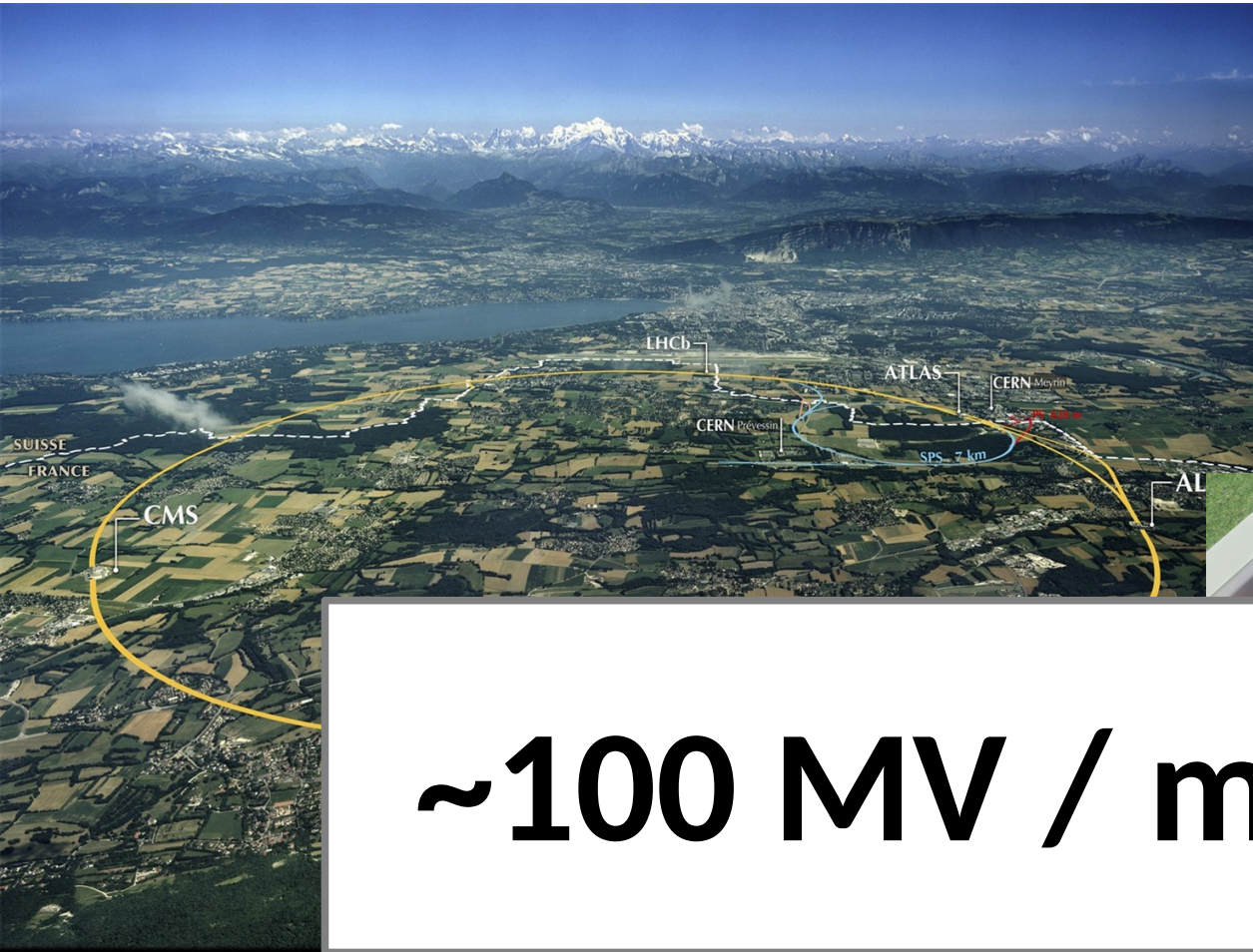
My Background

PhD Student

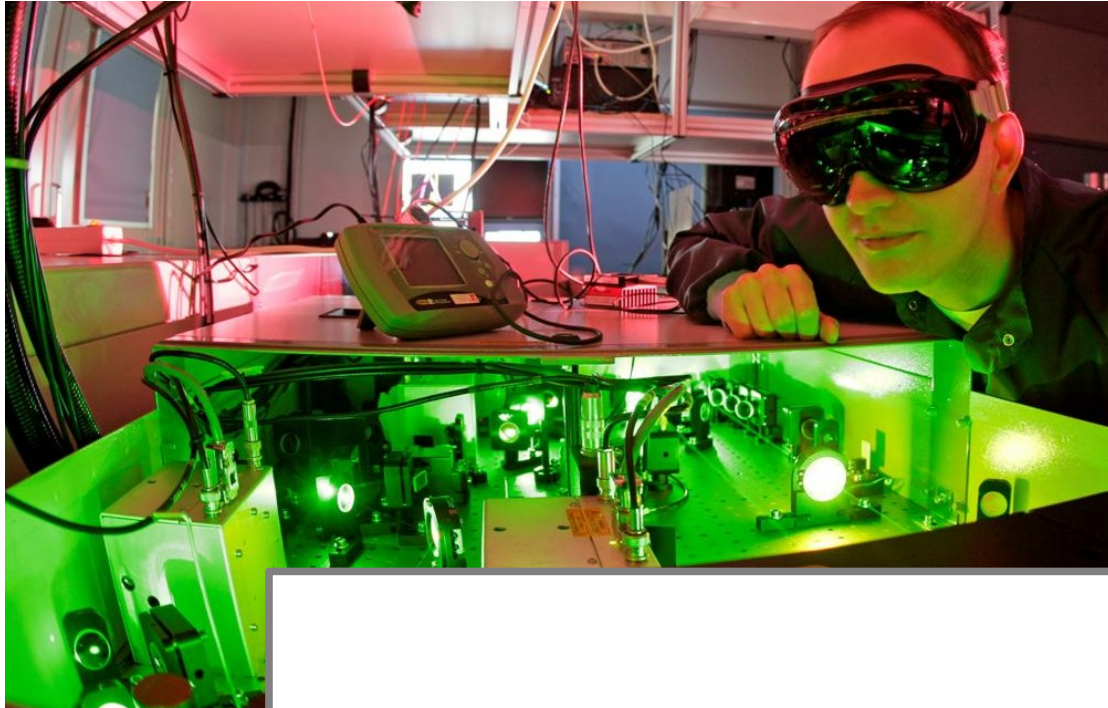
- Axel Huebl
 - *Dipl.-Phys. (TU Dresden)*
 - *Computational Physicist*
 - *PhD Student (should be writing)*
 - *Working at*
Helmholtz-Zentrum Dresden – Rossendorf e.V.
- *FOSS software developer & maintainer:*
 - *High-Performance Computing*
 - *Multi-PByte Data Analysis*



Conventional Particle Accelerators



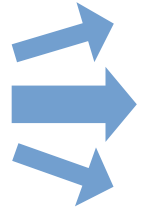
(Laser-) Plasma Acceleration



30 – 500 fs
800 – 1053 nm
45 – 200 J

50 nm
-
10 μ m

$$I = \frac{\text{PW}}{\text{cm}^2}$$



1 000 000 MV / m

RESSEN
concept



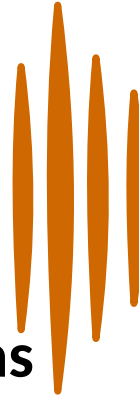
HZDR

Member of the Helmholtz Association

Simulations for Laser-Plasma Interaction

Particle-in-Cell Simulations

PICon GPU



+ review
+ education



- Ab initio, electro-magnetic plasmas
- Runs on the world's largest clusters
- Open sourced 2013



M. Bussmann et al., SC'13, DOI:10.1145/2503210.2504564

Member of the Helmholtz Association
Computer Assisted Radiation Physics | picongpu.hzdr.de

Helmholtz Association

“solving the grand challenges of science, society and industry”

- Union of 18 German research centers
 - 38'733 employees
 - 4.38 billion € budget annually
- Large scale facility research
 - *distributed data acquisition & labs*
 - *distributed, heterogeneous computing*
 - *complex eco-systems, interdisciplinary*

HELMHOLTZ
SPITZENFORSCHUNG FÜR
GROSSE HERAUSFORDERUNGEN

Prototypes

Quick hacks live forever

Scientific Software

Common Tasks

- Analysis, Modeling
 - *generate knowledge from data*
- Machine Control
 - *highly tunable, logging, sensitive, multi-protocol & platform, remote, ...*
- Demonstrator
 - *see, algorithm X works!*

Typical Design

Does this look familiar?

- Analysis, Modeling
 - *StackOverflow and prev. colleague*
- Machine Control
 - *plug on a plug on a ... semi-manual workflow*
- Demonstrator
 - *for results, see plot <N>*

Modern software development



First Order Problems

- Planning
 - *good: re-use existing solutions*
 - *bad: existing (& new) solution is unmaintained*
- Execution
 - *setup of environment, data, dependencies, ...*
- Reproduction
 - *can we repeat this after $\langle N \rangle$ months?*

Reproducibility

Is this science yet?

The Scientific Method

- principles and procedures for the **systematic pursuit** of knowledge involving the recognition and formulation of a **problem**, the collection of **data** through observation and experiment, and the formulation and testing of **hypotheses**

"Scientific Method." Merriam-Webster.com.

Merriam-Webster, 27th Nov 2018.

- **Reproducibility** is the *closeness of the agreement* between the results of measurements.
- **Replicability** is the ability to independently achieve *non-identical conclusions* that are at least *similar*.

"Reproducibility." en.wikipedia.org, 27th Nov 2018

Scientific Replication Crisis

Yet another Crisis?

- Methodological problem
 - *first discussed in social sciences but universal*
- Replicating own and *third party* results
 - Failure of replication does *not* always mean conclusions are wrong. But they are not reliable.
- Nature 2016 poll (1'576 researchers):
 - Tried and failed to reproduce other (70%) and own (>50%) experiments

Baker, Monya. DOI:10.1038/533452a

Scientific Replication Crisis

2018 Status

- Still ongoing
 - *cultural process*
- Great efforts sprung off
 - *pre-registered trials*
 - *open workflows & codes*
 - *open access journals & data*
 - ***improved funding guidelines & author crediting***

Excursus: What's a DOI

- unique resource identifier
 - *For any kind of resource*
 - *DOI:prefix/someNumber*
- ID fixed
- Data fixed (long-term archives)
- *Meta-data adjustable*

Excursus: What's a ORCID



- unique author identifier
 - 1:n
 - *an author can have N ORCIDs*
 - *orcid.org/XXXX-YYYY-ZZZZ-WWWW*
- ID fixed
- Author fixed, meta-data **adjustable**:
 - *Name*
 - *Publication list*
 - *Affiliation*

Ways towards Reproducibility

Software: Minimum

- versioning
 - DOIs & long-term available archives
- open data
 - calibration, input/output
- reproducible environments
 - dependencies, installed variant
 - hardware ...

Ways towards Reproducibility

Software Development

- open development
 - if you can't freeze time, move forward
 - documented decision processes
 - open issues
 - changelogs
- semantic versioning
- problem benchmarks are tests
- automated testing (CI) and deployment (CD)
- reviews

Infrastructure For Free Scientific Software

Infrastructure For Free Scientific Software

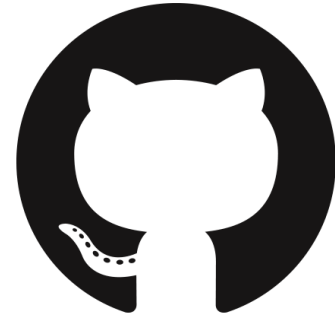
- Tools
- Services
- Workflows

Code Hosting Criteria

- Ownership
 - **must stay yours**
 - Exportable?
 - Extras exportable?
- Financing
 - should be stable
 - avoid ad-based
(privacy, seriousness, focus)

Code Hosting Platforms

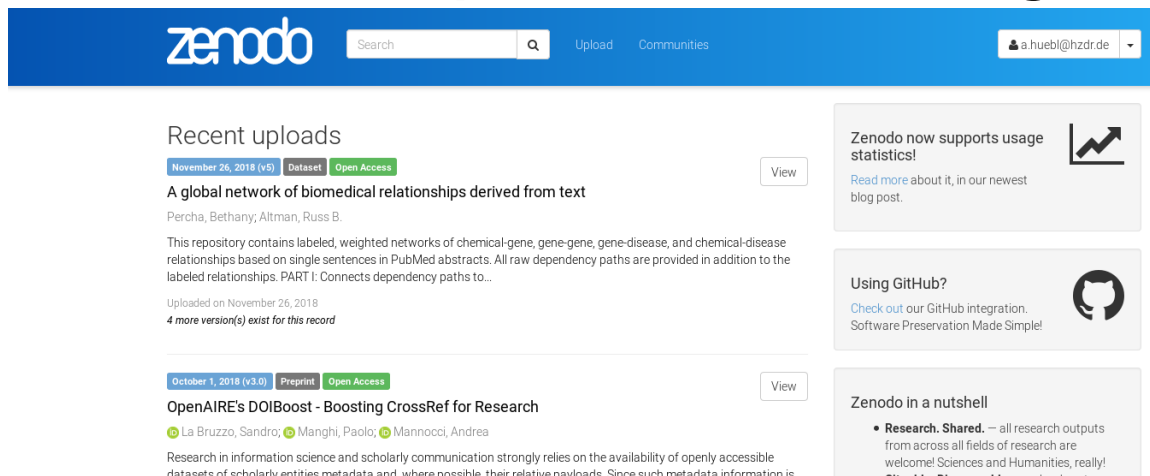
- GitHub.com
 - defacto standard
 - MS owned
 - Free if code is public
 - Non-free server SW
- GitLab.com/org
 - open-core business
 - self-hostable
 - **time-to-contribution! (SSO)**



Free DOIs

Software, Data, Talks, Posters, Preprints

- Zenodo.org
 - powered by free software: Invenio (self-host)
 - CERN and EU sponsored hosting



- Alternatives (US): Figshare



PICongGPU 0.4.2: CPU Plugin Performance

 Huebl, Axel;
  Widera, René;
  Worpitz, Benjamin;
  Pausch, Richard;
  Burau, Heiko;
  Garten, Marco;
  Starke, Sebastian;
  Grund, Alexander;
  Debus, Alexander;
  Matthes, Alexander;
  Bastrakov, Sergei;
  Steiniger, Klaus;
  Göthel, Ilja;
  Rudat, Sophie;
  Kelling, Jeffrey;
  Bussmann, Michael

This release fixes a performance regression for energy histograms and phase space plugins on CPU with our OpenMP backend on CPU. At least OpenMP 3.1 is needed to benefit from this. Additionally, several small documentation issues have been fixed and the energy histogram python tool forgot to return the first iteration.

Please refer to our [ChangeLog](#) for a full list of features, fixes and user interface changes before getting started.

Thanks to Axel Huebl, René Widera, Sebastian Starke, and Marco Garten for contributions to this release!

Preview

 picongpu-0.4.2.zip

 The previewer is not showing all the files

ComputationalRadiationPhysics-picongpu-7a6eeeb

-  .gitignore
 309 Bytes
-  .readthedocs.yml
 86 Bytes
-  .travis.yml
 6.7 kB
-  zenodo.json
 3.4 kB
-  CHANGELOG.md
 91.3 kB
-  CONTRIBUTING.md
 14.3 kB
-  COPYING
 35.1 kB
-  COPYING.LESSER
 7.7 kB
-  INSTALL.rst
 13.5 kB
-  LICENSE.md
 3.8 kB
-  README.md
 11.4 kB
-  REFERENCE.md
 773 Bytes
-  USAGE.rst
 7.8 kB
-  bin
 -  cuda_memtest.sh
 1.5 kB

Files (4.4 MB)

Name	Size
------	------

206

views

43

downloads

[See more details...](#)

Available in



Publication date:

November 19, 2018

DOI:

DOI [10.5281/zenodo.1491926](https://doi.org/10.5281/zenodo.1491926)

Keyword(s):

[PICongGPU](#)
[CUDA](#)
[OpenMP](#)
[manycore](#)
[Plasma Physics](#)
[GPU](#)

Grants:

[European Commission:](#)

- [EUCALL - European Cluster of Advanced Laser Light Sources \(654220\)](#)

Related identifiers:

Supplement to:

<https://github.com/ComputationalRadiationPhysics/picongpu/tree/0.4.2>

License (for files):

[GNU General Public License v3.0 only](#)

Free User ID: ORCID

For all kinds of authors

- orcid.org
 - Powered by free software & open APIs
 - NSF and funding by scientific societies
- Alternatives: less free or legacy...

The screenshot shows the ORCID iD profile for Axel Huebl. The header includes the ORCID logo and navigation links: FOR RESEARCHERS, FOR ORGANIZATIONS, ABOUT, HELP, and SIGN OUT. Below the header, it displays the user's name, biography, and a list of works. The profile also includes links to the user's ORCID ID, public record print view, and QR code. The works section lists three publications, each with a title, journal, DOI, and source information.

ORCID ID
https://orcid.org/0000-0003-1943-7141
View public version

Display your ID on other sites
Public record print view
Get a QR Code for your ID

Also known as
Axel Huebl, Axel Hübl

Country
Germany

Keywords
Laser-Plasma Based Accelerators, HPC, PIconGPU, Many-Core, GPGPU

Websites
HZDR Publication Database
PIConGPU Project Page
GitHub Profile
ResearchGate Profile

Other IDs

Emails
a.huebl@hzdr.de

Biography

Employment (7)

Education and qualifications (1)

Invited positions and distinctions (1)

Membership and service (0)

Funding (0)

Works (28 of 28) + Add works Export works Bulk edit Sort

Isolated proton bunch acceleration by a petawatt laser pulse
Nature Communications
2018-12 | journal-article
DOI: 10.1038/s41467-017-02663-1
Source: Crossref Preferred source

PIConGPU 0.4.0: CPU Support, Particle Filter, Probes & Merging
2018-10-19 | software
DOI: 10.5281/zenodo.1467038
Source: Axel Huebl Preferred source

Quantitatively consistent computation of coherent and incoherent radiation in particle-in-cell codes—A general form factor formalism for macro-particles
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment
2018 | journal-article
DOI: 10.1016/j.nima.2018.03.033

DRESDEN concept HZDR

Automated Tests

Continuous Integration

- Travis-CI
- Circle-CI
- AppVeyor
 - free for FOSS
- Drone.io
- GitLab-CI
- Jenkins
 - self-hostable

openPMD / openPMD-api build passing

Current Branches Build History Pull Requests > [Build #1507](#)

✓ **dev** Merge pull request #396 from franzpoeschel/fix-mpiDeprecated → #1507 passed

Replace MPI::

→ Commit f1dc0b7 [↗](#)

🔍 Compare 9b4c34f...f1dc0b7 [↗](#)

🌿 Branch dev [↗](#)

👤 Axel Huebl

⌚ Ran for 33 min 4 sec
⌚ Total time 1 hr 39 min 48 sec
📅 12 days ago

[Build jobs](#) [View config](#)

🕒 **Style**

✓ # 1507.1	👤	📄 PY@2.7 +PEP8
✓ # 1507.2	👤	📄 PY@3.6 +PEP8
! # 1507.3	👤	📄 clang-format@5.0.0

Your build matrix was set to allow the failure of job 1507.3 so we continued this build to the next stage.

🕒 **Static code analysis**

✓ # 1507.4	👤	📄 pyflakes 2.7
✓ # 1507.5	👤	📄 pyflakes 3.6

🕒 **C++ unit tests**

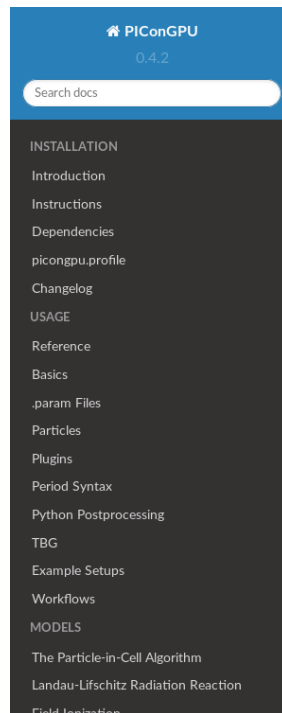
✓ # 1507.6	👤	📄 clang@5.0.0 -MPI -PY +H5 -ADIOS1
✓ # 1507.7	👤	📄 clang@5.0.0 +MPI -PY +H5 +ADIOS1

Documentation

Continuously built

- APIs
 - Doxygen (C++, Java, ...)
 - Autodoc (py)
 - Rustdoc (rs)
- Manuals:
 - Sphinx + Language Extensions
- Service: readthedocs.org
 - Powered by free software



Building and Shipping

User-Level Package Managers

- Most users are not developers
 - Compile? Dependencies?
- Most users are not admins
 - `pip install --user matplotlib`
- Reproducible, complex environments
 - Containers: **Singularity**, Docker, Shifter
 - Publish recipes!

Building and Shipping

User-Level Package Managers

- User-Level Package Managers
 - PyPI, Conda (-forge)
 - Spack, EasyBuild, Conan.io, cargo, npm, vcpkg, flatpak, nix
 - Linux Distributions & Homebrew
- Containers: all in one "binary"
 - **Singularity**, Docker, Shifter
 - Publish recipes!

Building and Shipping Container Recipes

Singularity example

- Developer:
 - writes this →
 - publishes build
- Users:
 - executes like a single binary
 - not necessarily isolated

```
Bootstrap: docker
From: ubuntu:16.04

%post
    apt-get -y update
    apt-get -y install fortune cowsay lolcat

%environment
    export LC_ALL=C
    export PATH=/usr/games:$PATH

%runscript
    fortune | cowsay | lolcat
```

Summary: Infrastructure & Tools

Curated List

- Better Scientific Software
 - BSSw.io



- community-driven resources
- powered by free software
- contributable

Self-Help

Licenses, FOSS & OA as a Student

How to Choose A License

GNU Free Software Definition

Users have the freedom to:

- Run,
 - Copy,
 - Distribute,
 - Study,
 - Change and Improve the software.
- (and then re-distribute)

“Free software” does not mean “non-commercial”!

www.gnu.org/philosophy/free-sw.en.html

Licensing

First Things First

- I am not a lawyer.
- This is not legal advise.

These is just the personal rumbling of someone working scientifically with and producing on a daily basis open software and open data.

DRESDEN
concept



HZDR

How to Choose A License

General Aspects

- You publish/upload/share something
 - **protection** for yourself (use *and* contributions)
 - allow to use it
 - re-share
 - re-use parts, e.g. for follow-up problems, comparisons, reproducing, ...
 - **improve and adjust**
- **choosealicense.com/licenses/**

Resource

choosealicense.com

GNU LGPLv3

Permissions of this copyleft license are conditioned on making available complete source code of licensed works and modifications under the same license or the GNU GPLv3. Copyright and license notices must be preserved. Contributors provide an express grant of patent rights. However, a larger work using the licensed work through interfaces provided by the licensed work may be distributed under different terms and without source code for the larger work.

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- Disclose source
- License and copyright notice
- Same license (library)
- State changes

Limitations

- Liability
- Warranty

[View full GNU Lesser General Public License v3.0 »](#)

Mozilla Public License 2.0

Permissions of this weak copyleft license are conditioned on making available source code of licensed files and modifications of those files under the same license (or in certain cases, one of the GNU licenses). Copyright and license notices must be preserved. Contributors provide an express grant of patent rights. However, a larger work using the licensed work may be distributed under different terms and without source code for files added in the larger work.

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- Disclose source
- License and copyright notice
- Same license (file)

Limitations

- Liability
- Trademark use
- Warranty

[View full Mozilla Public License 2.0 »](#)

Apache License 2.0

A permissive license whose main conditions require preservation of copyright and license notices. Contributors provide an express grant of patent rights. Licensed works, modifications, and larger works may be distributed under different terms and without source code.

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- License and copyright notice
- State changes

Limitations

- Liability
- Trademark use
- Warranty

[View full Apache License 2.0 »](#)

How to Choose A License

Copyleft

- Examples:
 - (A)GPL, MPL (Software)
 - CC BY-SA (Data, Texts, Figures, Art, ...)
- Pros:
 - modifications stay open
 - protects “smaller” contributors
- Cons:
 - author’s pseudo-fear of ...?
 - industry: partially no-go (pseudo-fear)

How to Choose A License

Without Copyleft

- Examples:
 - new BSD, MIT, ISC (Software)
 - CC-BY (Data, Texts, Figures, Art, ...)
- Pros:
 - minimal hurdles, maximum compatible
- Cons:
 - As a smaller party, one can get "swallowed"

How to Choose A License

Additional Aspects

- Even copyleft can struggle with:
 - SaaS, e.g. MongoDB (SSPL)
- Open Standards - risks:
 - "Embrace, extend, and extinguish"
- Forks
 - **Are good** (sign of need)
 - Contributions can not be enforced, but fostered. *Open culture! Credits!*

DO NOT Role Your Own License

Really, don't. Tell your employer. Don't.

- There are plenty.
- All of them are tested and have known compatibility.
- All of them are tested against common standards (e.g. GNU Free Software Definition).
- They are ***trusted*** by very large communities.
- If you do:
 - **First**, get approval/compatibility from/with: **Open Source Initiative (OSI)**, GNU FSD, Fedora- and Debian-legal
 - **Then** use it.

How to Open Source as a Student

In case you are the first in your group/lab/department

- BA, MA, diploma, etc.:
 - If unpaid, usually ***your copyright, your decision.***
 - Ask your adviser, then Prof. → ok? Done!
- Student work, graduate work (paid):
 - Definitely ask your Prof. & department head.
 - Discuss maintenance! (Estimate additional workload)
 - Try to connect to same-level groups.
 - Integrate in existing publication cycle (e.g. at HZDR).

How to Open Source as a Student

In case you are the first in your group/lab/department

- No solution in sight? Nobody heard of FOSS?
 - Do not get lost in bureaucracy.
 - Do not escalate decision level unnecessarily, unless clear strategy & success stories.
- Instead, go one step back and **connect**.
 - “Academic” user groups with FOSS in Dresden: FSFW DD, GCoE, C++/Rust/Python user groups, students/colleagues in your department, ZIH (TUD), C3D2, ...

What to Expect after a FOSS Release

First, Nothing Happens

- Congrats, you are awesome!
- Over time:
 - Bookmarks/Stars/Likes
 - Users
 - Questions / Year
 - Contributors
- *All metrics are empiric.
E.g.: re-distribution.*
- Typical ratio:
 - 30:
 - 10:
 - 5:
 - :1

Summary

Free Software in Science

Summary

First, Nothing Happens

- Great resources are out there.
- Ask for help. Help others.
- **Use the freedom you are given in academia.
As an author, you decide.**
- *Be kind: leave things in a better place as you found them. Contribute and help if you see things can be improved.*

Thank you for your attention

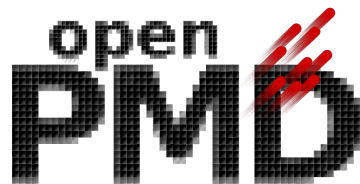
Our Scientific FOSS C++ Projects

Try, Reuse, Complain ;-) and Improve

ComputationalRadiationPhysics/



@ax3l



root-project/cling: -x **CUDA**



mallocMC



PICongPU

PMacc



This project has been enabled by many people in open-source and open-science communities. Great thanks to the communities and developers of: PICongPU, Alpaka, Jupyter, the SciPy ecosystem, ADIOS, HDF5, CMake, openPMD, Spack, Xeus, Cling, ...

Talk by **A. Huebl** (HZDR), a.huebl@hzdr.de



HZDR

Member of the Helmholtz Association

Axel Huebl | HZDR - Research Group Computer Assisted Radiation Physics | picongpu.hzdr.de