

Science is not open without code

Analysis Preservation Bootcamp, PSI, 22-23 January 2025



Kati Lassila-Perini
Helsinki Institute of Physics - Finland



Hello!

I am **Kati Lassila-Perini**

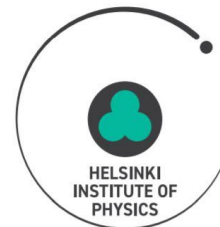
experimental particle physicist

CMS data preservation and open access (DPOA) coordinator (2012 - 2024)

ICFA Data Lifecycle panel chair (2024 -)

Find me at: kati.lassila-perini@cern.ch

[@katilp.bsky.social](https://bsky.app/profile/katilp.bsky.social)



1

Why am I here?
Why are you here?

Decade of CMS open data - with a small dedicated team

Type something

13 result(s) found

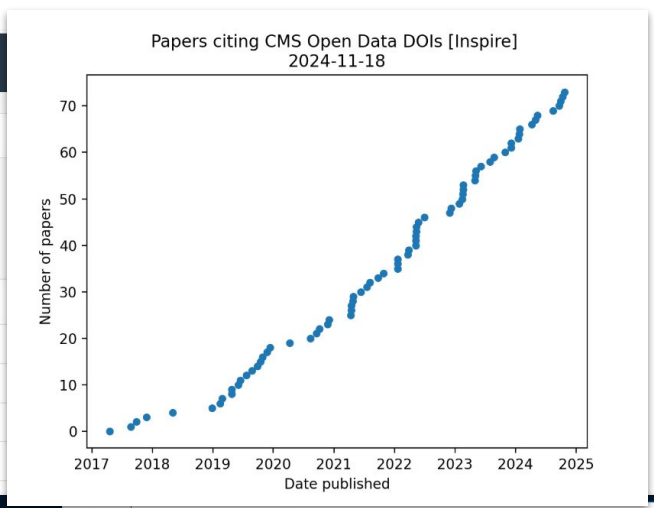
Sort by Most recent

Current parameters [Clear all](#)
News x CMS x

Availability
 Include on-demand datasets

- Type
- Dataset (42,583)
 - Collision (342)
 - Derived (253)
 - Simulated (41,988)
 - Documentation (54)
 - About (4)
 - Activities (13)
 - Authors (3)
 - Guide (27)
 - Help (2)

- [CMS releases 13 TeV proton collision data from 2016](#)
- [CMS completes Run-1 heavy ion open data collection](#)
- [CMS completes the release of its entire Run-1 proton-proton data](#)
- [First CMS open data from LHC Run 2 released](#)
- [CMS releases heavy-ion data from 2010 and 2011](#)
- [CERN Open Data Policy for the LHC Experiments](#)
- [CMS completes 2010-2011 proton-proton data release](#)
- [CMS releases open data for Machine Learning](#)
- [Observing the Higgs with over one petabyte of new CMS Open Data](#)
- [The Future of Particle Physics is "Open"](#)
- [Improving educational content with high-school teachers: A field report from CMS](#)
- [CMS releases new batch of research data from LHC](#)
- [CMS releases first batch of high-level LHC open data](#)



INSPIRE HEP

literature | references.reference.dois:10.7483/OPENDATA.CMS*

88 results | [cite all](#) Citation Summary Most Recent

New Angles on Energy Correlators #1
Samuel Alipour-fard (MIT, Cambridge, CTP), Anika Budhraj (NIKHEF, Amsterdam), Jesse Thaler (MIT, Cambridge, CTP), Wouter J. Waalewijn (NIKHEF, Amsterdam and Amsterdam U.) (Oct 21, 2024)
e-Print: 2410.16368 [hep-ph]
[pdf](#) [cite](#) [claim](#) [reference search](#) [0 citations](#)

Angular distribution study for high mass dimuon pairs in CMS open 2012 data and for Mono-Z' model #2
S. Elgammal (British U. in Egypt) (Oct 8, 2024)
e-Print: 2410.05755 [hep-ex]
[pdf](#) [cite](#) [claim](#) [reference search](#) [0 citations](#)

Measurement of the background in the CMS muon detector in pp -collisions at $\sqrt{s} = 13$ TeV #3
CMS Muon Collaboration • M. Tytgat (Vrije U. Brussels and Gent U.) et al. (Sep 27, 2024)
Published in: *Eur.Phys.J.C* 84 (2024) 9, 955
[pdf](#) [DOI](#) [cite](#) [claim](#) [reference search](#) [0 citations](#)

ν -point energy correlators with FastEEC: small- x physics from LHC jets #4
Anika Budhraj (NIKHEF, Amsterdam and Amsterdam U.) et al. (Sep 18, 2024)
e-Print: 2409.11811 [hep-ph]
[pdf](#) [DOI](#) [cite](#) [claim](#) [reference search](#) [1 citation](#)

Date of paper

Number of authors
 Single author 16
 10 authors or less 76
Exclude RPP
 Exclude Review of Particle Physics 88
Document Type
 article 56
 published 41
 conference paper 26
 thesis 7
 note 1



CMS open data in use

From CMS open data to open science

Tools:

- common software
- environments
- interfaces

... the tools that you will learn are essential for open data use.

I am here because...

... we struggled with this!

Data:

- collision data
- simulations
- additional data for analysis

Knowledge:

- instructions
- example code, actionable examples
- understanding of experimental data



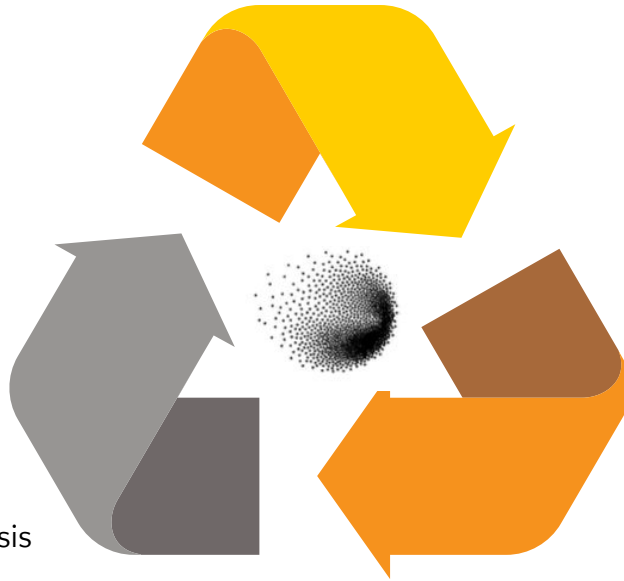


Tools:

- software
- environments
- interfaces

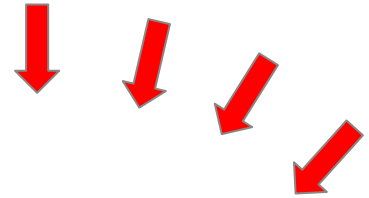
Data:

- collected data
- simulations
- additional data for analysis



Knowledge:

- instructions
- actionable examples
- understanding of experimental data



2

Open Science?

SNSF

Data and publications must be freely accessible.

UNESCO


Open science is a set of principles and practices that aim to make scientific research from all fields accessible to everyone for the benefits of scientists and society as a whole. Open science is about making sure not only that scientific knowledge is accessible but also that the production of that knowledge itself is inclusive, equitable and sustainable.

CERN


“ Supported by long term financial investments from its Member and Associate Member States, with significant contributions also from non-Member States, CERN is committed to the advancement of science and the wide dissemination of knowledge by embracing and promoting practices making scientific research more open, collaborative, and responsive to societal changes. ... CERN accordingly recognizes the holistic practice of open science as one of its guiding principles. ”

NASA adheres to the principle of Open Science which involves the practice of making research products and processes available to all while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity.

NASA



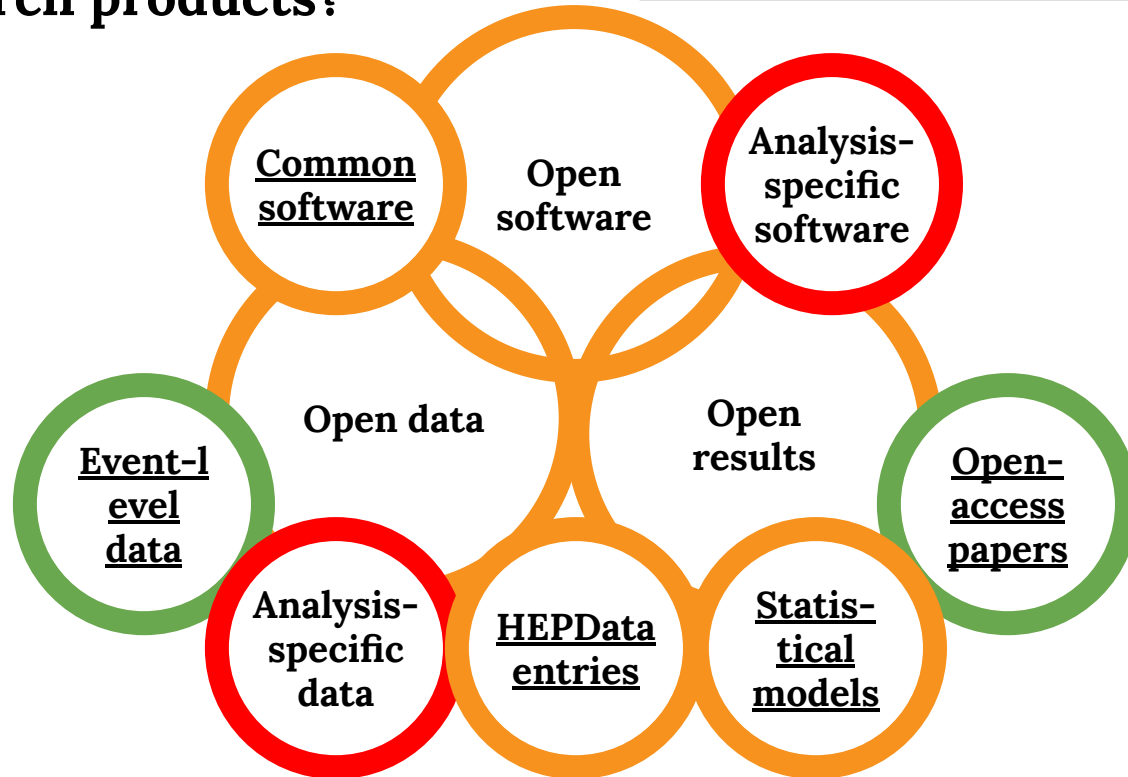
**Are some of your
research products open?**



**Have you used open
science “products”?**



Research products?





Winds of change? An example:

- NASA has a strong initiative, “Transform to Open Science (TOPS)”, for improving access to their research products, including explicitly software



- An updated plan
- Main changes →
- Learn more in [NASA Open Science 101](#)

February 21, 2023

Updated: November 15, 2024

Revisions

This section highlights the significant changes to this document since the original plan was released in 2014¹. To wit:

- There shall be no publication embargo period for peer-reviewed publications.
- Data that support peer-reviewed publications shall be made available in a public archive at the time of publication.
- Software used to generate research findings/results should be included as part of public access at the time of publication, subject to National Aeronautics and Space Administration (NASA) software release requirements.
- Other data products beyond peer-reviewed publications and software are encouraged to be considered as part of public access.
- Reasonable costs for making research activities publicly accessible shall be allowed in grant proposals. Review panels will be given guidance on what constitutes “reasonable” for compliance purposes which may depend on the particular program solicitation.

3

Using open software

What would **you** want to know before using software that you've found?

- good README
- examples
- requirements, dependencies
- user community
- test with an example output
- troubleshooting
- is it actively maintained
- licence
- where is has been used
- institute rules
- contact information
- installation instructions



FAIR? For data and software!

FINDABLE

Do you know where to look for them?

Can you find what you need?

F

A

ACCESSIBLE

Can you download them?

Can you install them?

Are they in some common format?

Can you run them?

I

R

Do you know how to use?

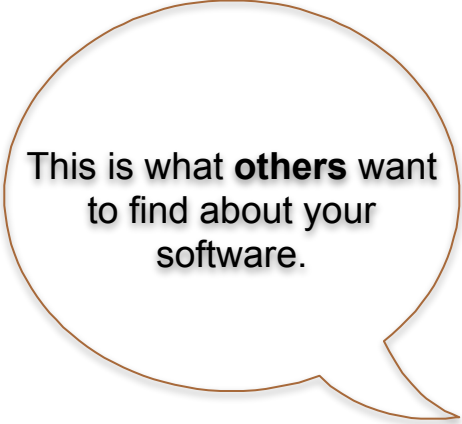
Can you make new research with them?

INTEROPERABLE

REUSABLE

4

Making open software



This is what **others** want to find about your software.

First, they need to find it! Then:

- A well-defined purpose in a README
- An example demonstrating how it is used.
- Clear instructions on how to compile and/or install.
- Dependencies, libraries etc well defined or packaged in a software container image.
- A method to verify that it is working correctly.
- Readable comments in the code.
- Versioning.
- Open questions / problems in an issue tracker.
- Licence.
- How to cite?
- How to contribute?









Start early!

Never seen enough human willpower to make it happen once results are under review or published.

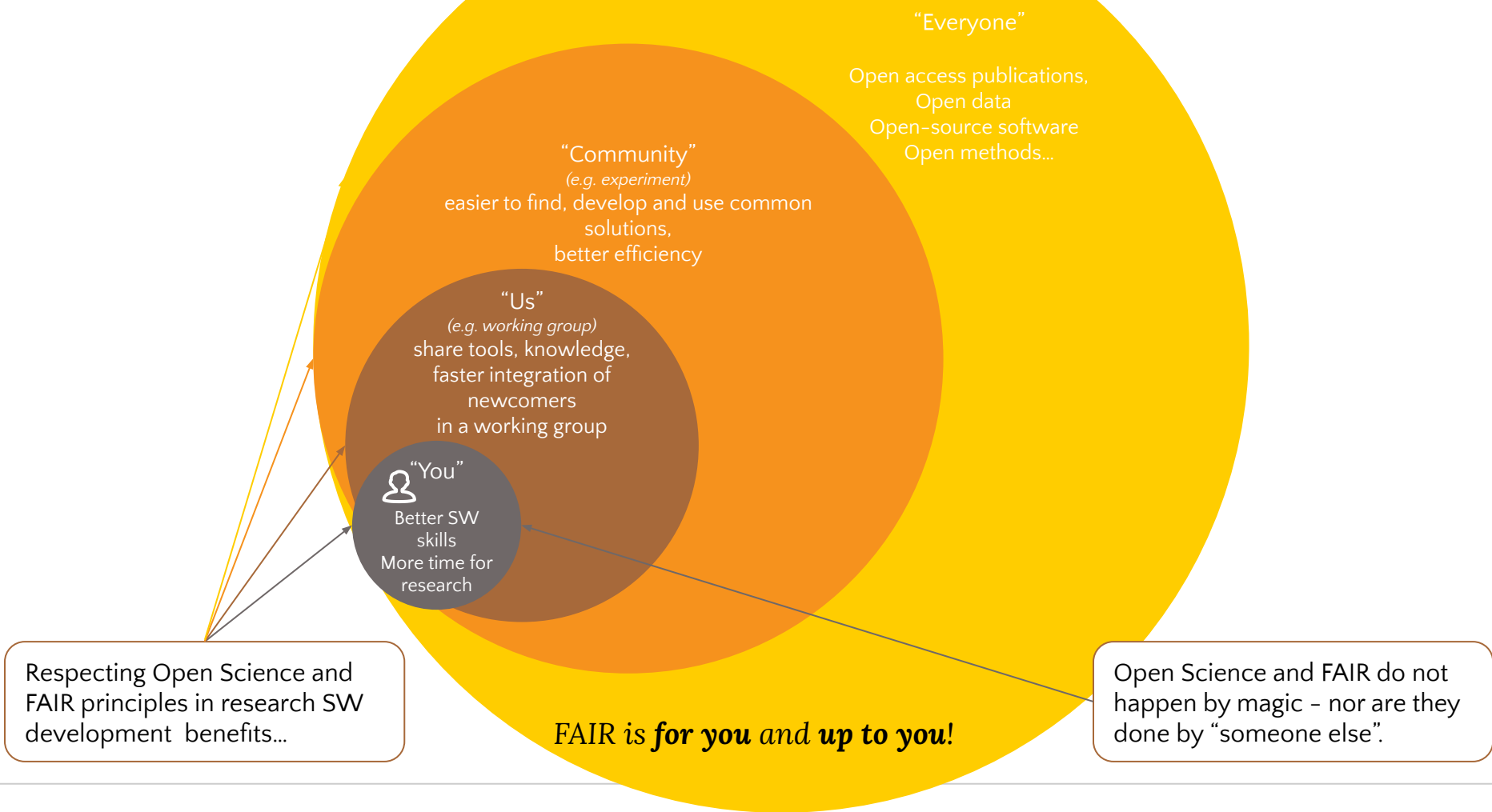


To make this easy:

- ◉ Make friends with git 
- ◉ Commit and push **often** to the repository 
 - ◉ if you work alone: commit and push **often!**
 - ◉ if you work with others: commit and push **often!**
- ◉ Agree on how to contribute and write it down 
 - ◉ use of branches, commit messages, reviewing, merging
- ◉ Use the repository's issue tracker 
 - ◉ Describe planned work in small enough entities.
 - ◉ Close them with merging commits.
- ◉ Use software containers 
 - ◉ for a well-defined sharable and reusable environment.
- ◉ Automate the functionality tests using CI/CD pipelines 

5

Benefits





Thank you!

Questions ?

And thanks to [SlidesCarnival](#) for this free presentation template