



Open Science at CERN: Infrastructure, policy and practice



Antonia Winkler
Contact: antonia.winkler@cern.ch
June 2024

Overview

- Open Science at CERN
- Policy
 - CERN Open Science Policy
- Infrastructure
 - From Invenio over Zenodo to InvenioRDM
 - White Rabbit
- Practice
 - CMS open data and software
- Open research practice vs. open infrastructure

CERN Convention

Founding principles of the Organization include that ... *the results of its experimental and theoretical work shall be published or otherwise made generally available.*

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CONVENTION

FOR THE ESTABLISHMENT OF A EUROPEAN ORGANIZATION
FOR NUCLEAR RESEARCH

PARIS, 1st JULY, 1953

As amended

CONVENTION

POUR L'ÉTABLISSEMENT D'UNE ORGANISATION EUROPÉENNE
POUR LA RECHERCHE NUCLÉAIRE

PARIS, le 1^{er} JUILLET 1953

Telle qu'elle a été modifiée

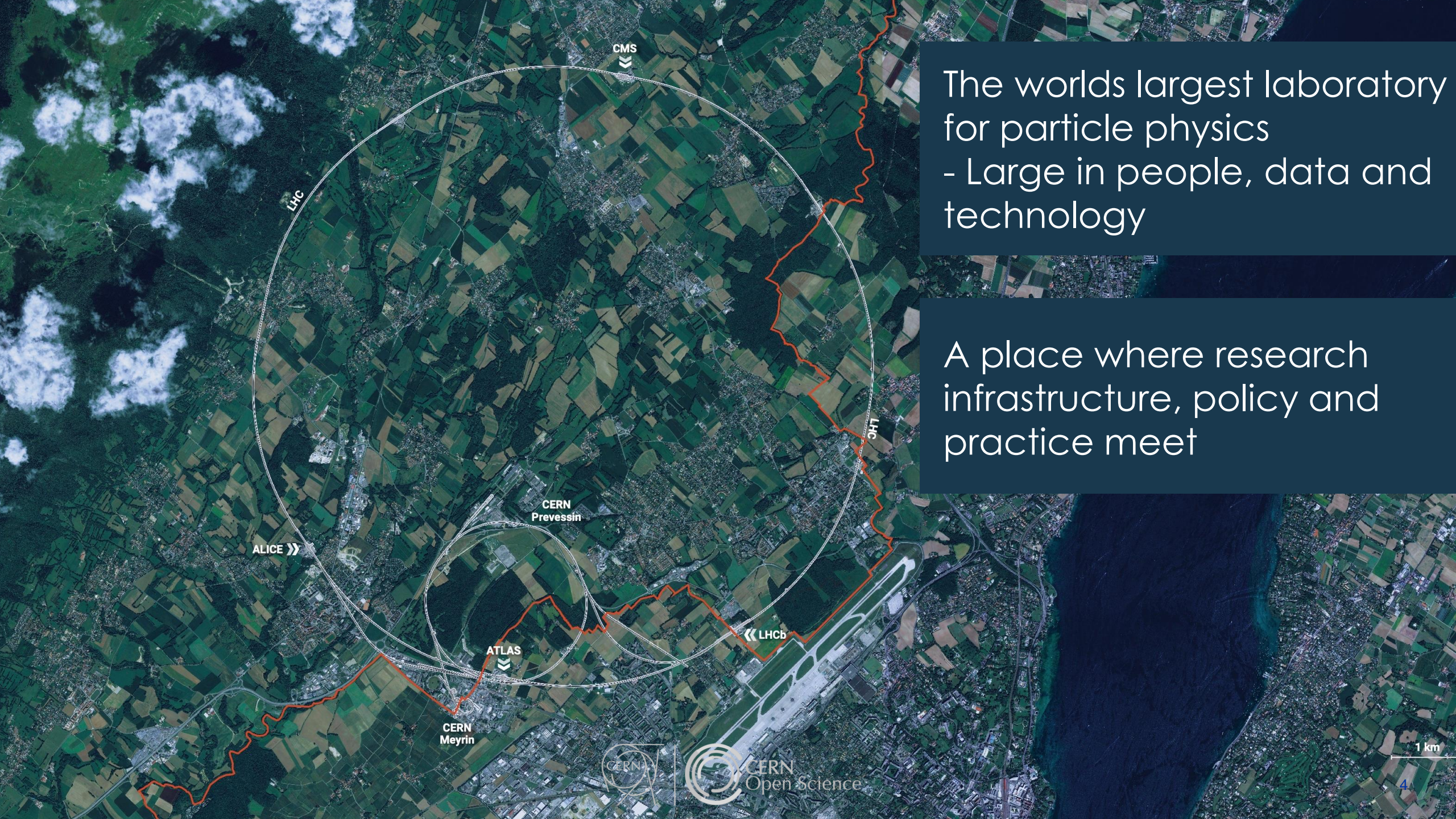
ÜBEREINKOMMEN

ZUR ERRICHTUNG EINER EUROPÄISCHEN ORGANISATION
FÜR KERNFORSCHUNG

PARIS, 1. JULI 1953

Revidierte Fassung





The worlds largest laboratory for particle physics
- Large in people, data and technology

A place where research infrastructure, policy and practice meet

Open Science milestones at CERN

Policy

Infrastructure

Practice

WWW



1989



1991

Article/Data linking



2011

OA Policy



2014

SCOAP3:
90% OA



2018

Open Data
Policy



2020

OSPO



2023

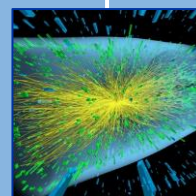


open hardware

Open
Hardware
License
V1.0

zenodo

Zenodo
Launch



LHC Open
Data

CERN
Analysis
Preservation

oreana

Reuse and
Reproducibility
Framework



Open Science
Policy

POLICY



CERN Open Science Policy

- Captures current practice and states progressive vision across multiple Open Science domains:
 - Open Access to Publications
 - Open Research Data
 - Open Software
 - Open Hardware
 - Infrastructure for Open Science
 - Research Integrity, Reuse & Reproducibility
 - Research Assessment & Evaluation
 - Education, Training & Outreach
 - Citizen Science
- Policy and its implementation plan are developed and governed by the community.
- V1.0, formally adopted by CERN Council, in force since Oct 2022:
<https://cds.cern.ch/record/2835057>

Open Source software

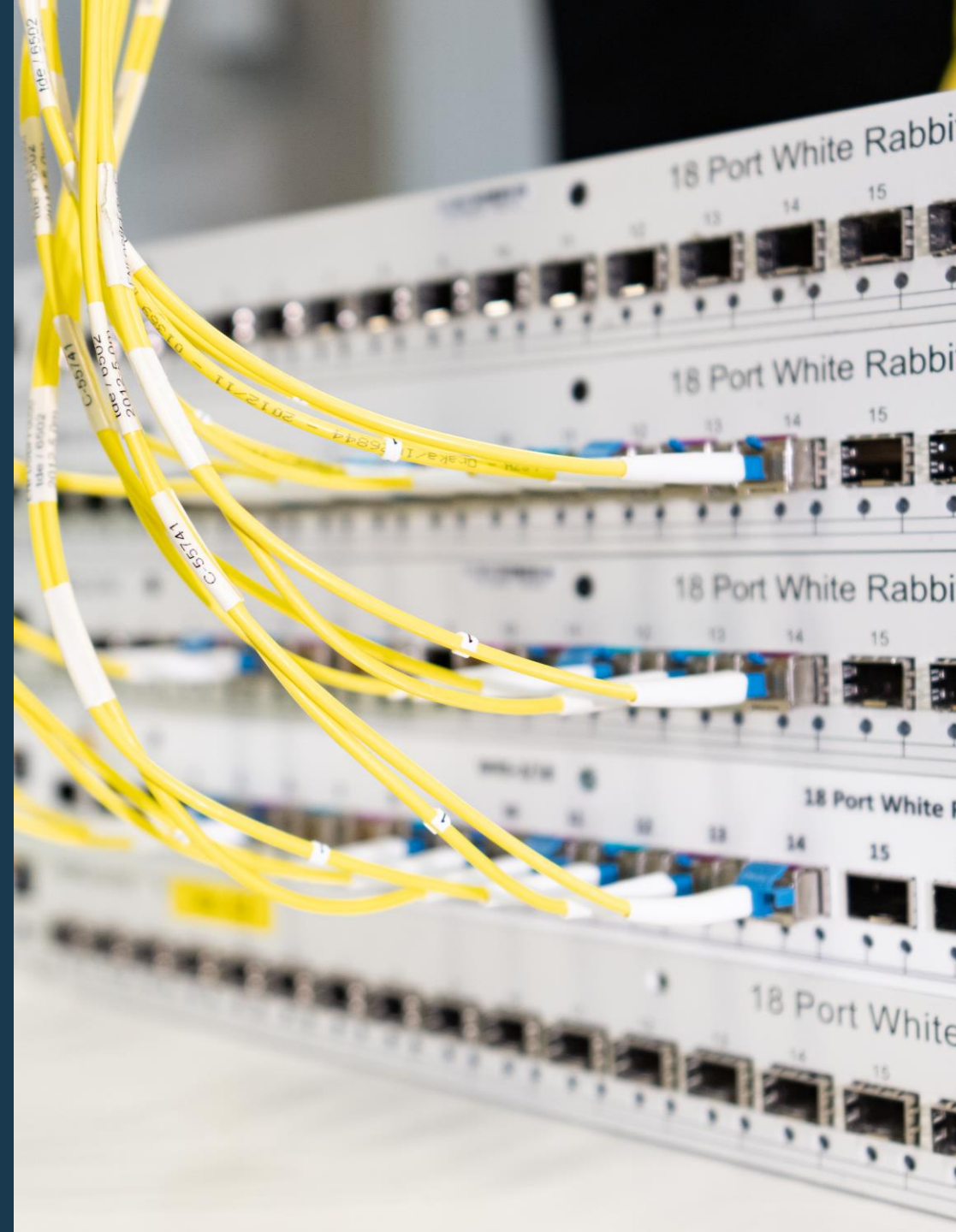
“CERN software is made available as open source wherever possible, applying a licence approved by the Open Source Initiative (OSI). CERN handles its research-related software as an integral part of its research products. Analysis of the CERN experiments’ physics data must be possible with open source software.” (CERN Open Science Policy, 2022)

```
if (electrons_sel.size() > 0)
if (muons_sel.size() > 0)
if (muons_sel.size() > 1)

// double leading_cut = (leptons[0]->PT > leading_cut)
// double subleading_cut = (leptons[1]->PT > subleading_cut)
// if ((leptons[0]->PT > leading_cut) && (leptons[1]->PT > subleading_cut))
if (tight_lepton)
{
countCutflowEvent("CutFlow_2L_3_Tight_1_0");
if (leptons[0]->Charge * leptons[1]->Charge > 0)
{
countCutflowEvent("CutFlow_2L_3_000_000");
bool CA_3th_1 = false;
if (leptons.size() > 1)
{
CA_3th_1 = leptons[1]->PT > ((leptons[0]->PT > 20) || (leptons[0]->PT > 10));
}
if (taus.size() > 0)
{
CA_3th_1 = taus[0]->PT > 20;
}
if (!CA_3th_1)
{
countCutflowEvent("CutFlow_2L_3_3th_1_000");
}
bool CA_low_mass = false;
```


Open Hardware

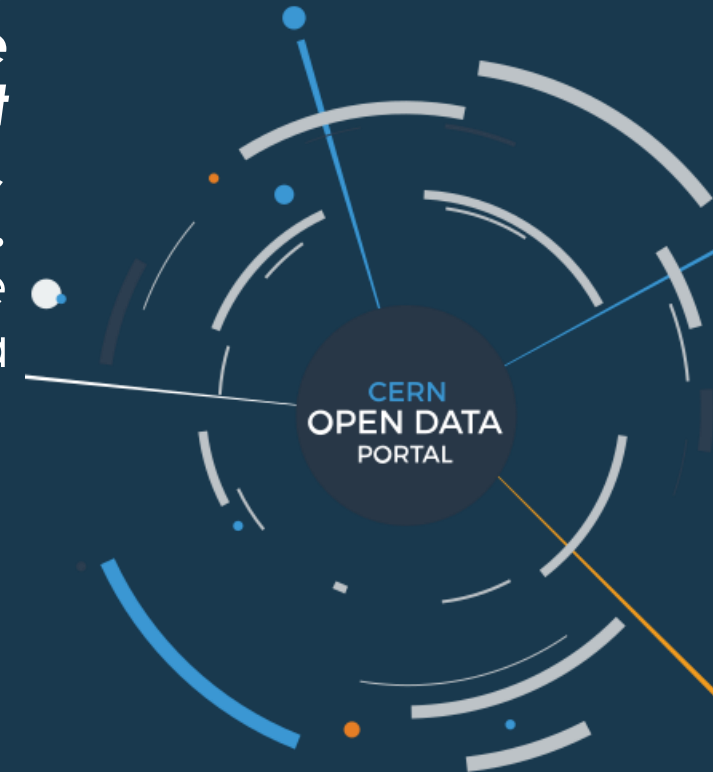
“CERN makes its technologies broadly available to society and has introduced open hardware licensing as a key mechanism to achieve this goal. [...] In cases where extensive documentation [...] like software for interfacing and testing are required for projects, these should be licensed under appropriate open source documentation and software licences respectively.” (CERN Open Science Policy, 2022)



Open Data

“CERN experimental collaborations are committed to making their research data publicly available.” (CERN open science policy, 2022)

“The LHC experiments will release calibrated reconstructed data with the level of detail useful for algorithmic, performance and physics studies. **The release of these data will be accompanied by provenance metadata, and by a concurrent release of appropriate simulated data samples, software, reproducible example analysis workflows, and documentation.** Virtual computing environments that are compatible with the data and software will be made available.” (CERN Open Data Policy, 2020)



Infrastructure provision for open science

“CERN is committed to ensuring the integrity of research. In order to facilitate the reuse of its research products, CERN provides infrastructures to accommodate the scale and complexity of its research outputs.” (CERN Open Science Policy, 2022)

***What counts as open science
Infrastructure and what counts as open
science research output?***

INFRASTRUCTURE



Community as make or break point of a project!

We are trying to do two things: build the best product out there in terms of usability, [...] but also we are trying to have the best community out there. So we were giving as much weight to the community aspect of the project [as to] the quality and the simplicity of the product.

```
ble sublead  
((leptons[0]  
ght_lepton)  
tCutflowEver  
leptons[0]->  
untCutflowEv  
01 CA_3th_1  
(leptons.s  
CA_3th_1 =  
(taus.size  
CA 3th 1 =
```

White Rabbit

- Ethernet based network that enables synchronous timing across large installations
- Released under the CERN Open Hardware Licence
- (Re)use by research institutions and companies
- Launch of the **White Rabbit collaboration** in 2024:
 - To attain membership, members pay up to 50,000 Swiss Francs per year
 - They receive:
 - A stake in the development of the technology
 - Guaranteed reply in fora and a number of support hours
 - Access to testing facilities



PRACTICE



CMS Open Data and Software

- Disseminating research-grade event-level particle physics data since 2014
- In **2023**, the content grew to over **4 petabytes**
- Goal: comprehensive publication of the collaboration's resources
- Release of:
 - Detector data
 - Simulated data
 - Analysis software
 - Various types of metadata
- Focus on documentation of material

Simulated dataset [ZZ_TuneCP5_13TeV-pythia8](#) in NANOADSIM format for 2016 collision data

[/ZZ_TuneCP5_13TeV-pythia8/RunIIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOADSIM](#), CMS Collaboration

Dataset Simulated Standard Model Physics ElectroWeak CMS 13TeV pp CERN-LHC

Description

Simulated dataset [ZZ_TuneCP5_13TeV-pythia8](#) in NANOADSIM format for 2016 collision data. See the description of the simulated dataset names in: [About CMS simulated dataset names](#).

These simulated datasets correspond to the collision data collected by the CMS experiment in 2016.

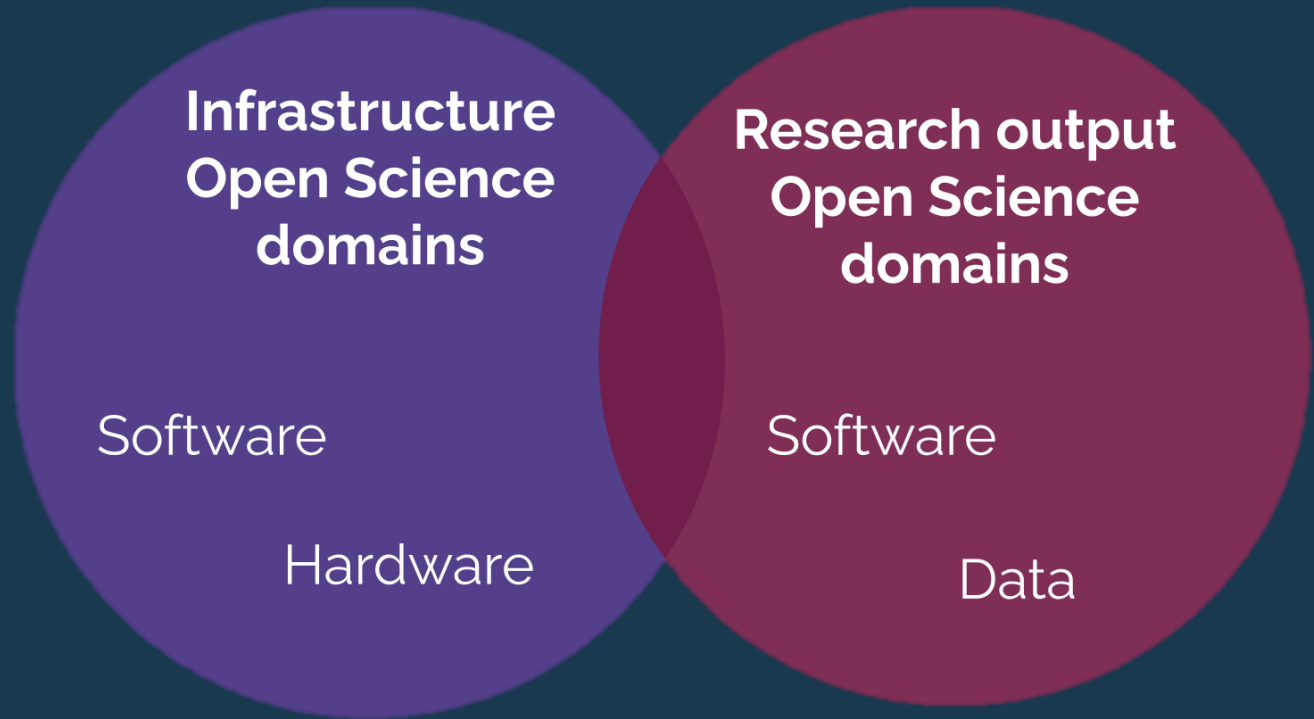
Related datasets

The corresponding MINIAADSIM dataset: [/ZZ_TuneCP5_13TeV-pythia8/RunIIISummer20UL16MiniAODv2-106X_mcRun2_asymptotic_v17-v1/MINIAADSIM](#)

Policy



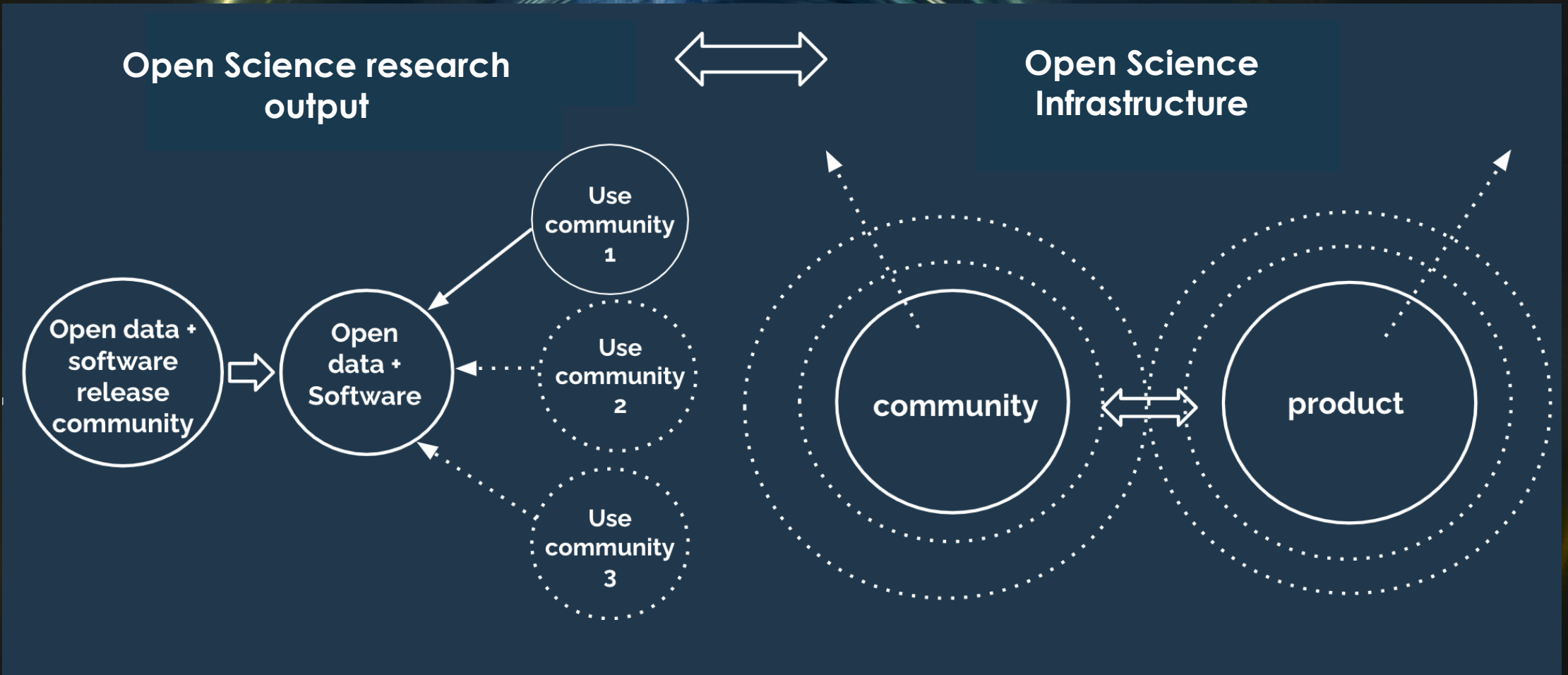
Practice



What are different approaches in
Infrastructure generating and
knowledge producing open
science domains?
How could they inform each
other?

INFRASTRUCTURE VS. PRACTICE

Temporality



Community curation vs. documentation

Open Science Infrastructure

Community curation strategies:

- Governance structure
- Granting a say in decision making processes
- Retaining authority as core development group
- Cultivating of a shared vision
- Eliciting financial investment
- Granting administrative/service capacities
- Granting access to testing facilities

Open Science research output

Documentation strategies:

- Metadata application
 - Context metadata
 - Content metadata
 - Provenance metadata
- DOI application
- Searchability

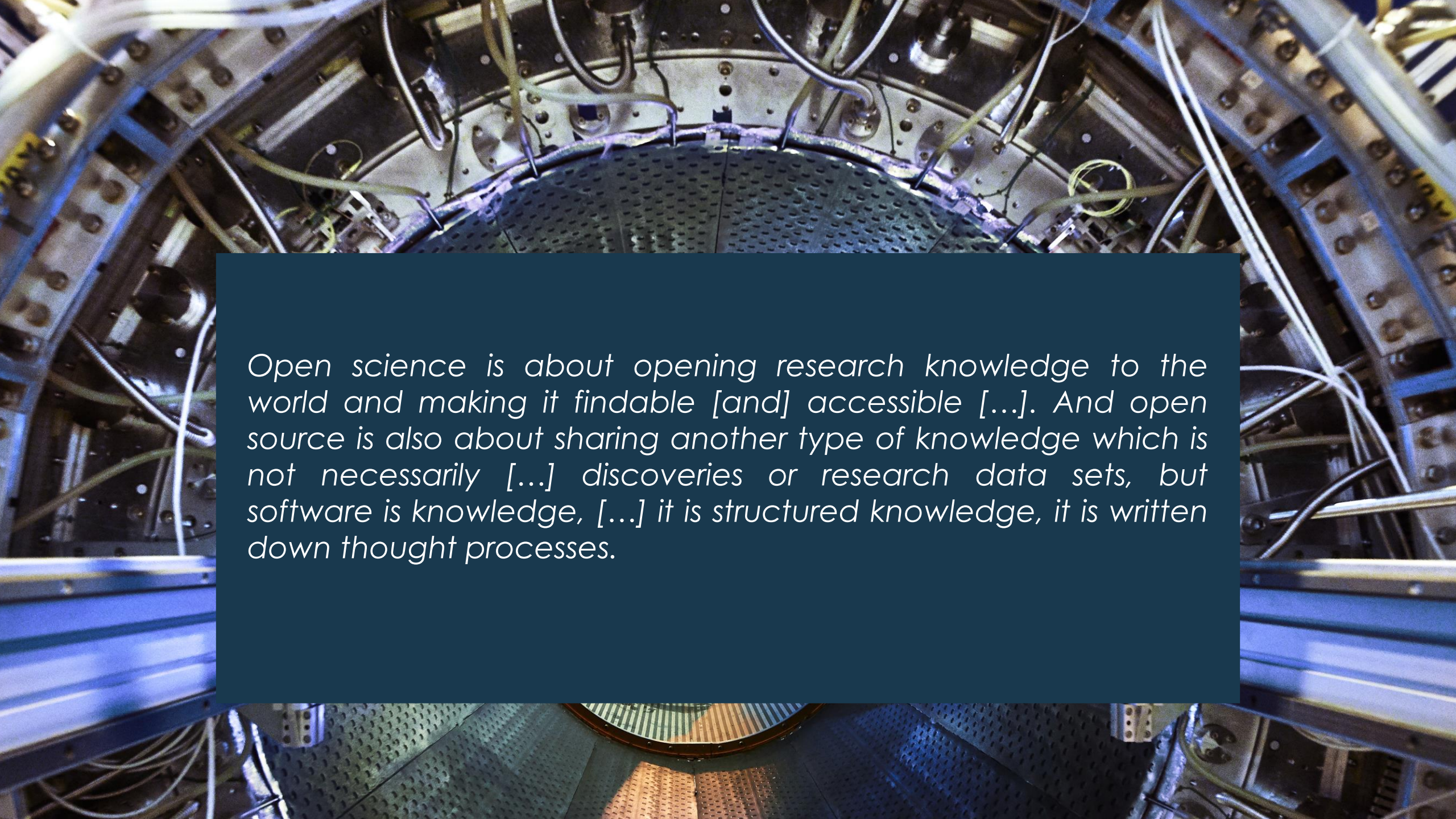
(Re)use openness vs. epistemic openness

“We want [our software] to be used by as many people as possible and that's [...] the premise of open source and free software”.



CMS open data and software efforts are carried out to ensure that a “full exploitation of [CMS resources] takes place”.

IMPLICATIONS

The background image shows a complex scientific instrument, possibly a particle detector or a large-scale experiment. It features a central circular component with a perforated surface, surrounded by a dense network of cables, pipes, and structural elements. The lighting is dramatic, with blue and white tones, highlighting the intricate details of the machinery.

Open science is about opening research knowledge to the world and making it findable [and] accessible [...]. And open source is also about sharing another type of knowledge which is not necessarily [...] discoveries or research data sets, but software is knowledge, [...] it is structured knowledge, it is written down thought processes.



Implications

Open science research outputs:

- Community focus?
 - Integration of (re)use communities in the release process

Open Science Infrastructure:

- Ask: What types of knowledge are made possible through our project?



The CERN Open Source Program Office (OSPO)

A bridge between open science infrastructure and research practice?

“The OSPO is an entry point to CERN’s expertise in Open Source. The OSPO is an open and inclusive service, working internally across departments with and for the entire CERN community” (OSPO mandate)





Questions:

Do you also perceive this distinction in your daily working practice?

What about Root?



Thank you!

Contact:
antonia.winkler@cern.ch

Backup slides

```
double subleading_cut = 0.001;
// if ((leptons[0]->OPT > leading
if (tight_lepton)
{
countCutflowEvent("CutFlow_2A_1");
if (leptons[0]->Charge * leptons[1]->Charge > 0)
{
countCutflowEvent("CutFlow_2A_2");
bool CA_3th_1 = false;
if (leptons.size() > 2)
{
CA_3th_1 = leptons[2]->OPT > 0;
}
if (taus.size() > 0)
{
CA_3th_1 = taus[0]->OPT > 0;
}
```

CERN Open Source Program Office



Internal Mandate

- Consult, advise, train on Open Source best practices, tools, licenses, etc.
- Advise on open-sourcing CERN software and hardware.
- Catalogue of Open Source software and hardware.
- Identify dependencies and compatibility for critical services.
- Advise CERN on Open Source matters.

External Mandate

- Showcase CERN contributions to e.g. member states' Open Source ecosystems.
- Facilitate partnerships with external entities, e.g. companies.
- Promote CERN as an Open Source lab.

Contact: Open.Source@cern.ch
<https://opensource.cern/>
Mandate:
<http://cds.cern.ch/record/2879995>

Open Science Policy Implementation Plan

- Policy accompanied by implementation document outlining measures for all aspects of the policy: <https://cds.cern.ch/record/2856044/> [V1]
- Each chapter had oversight by editors, but everyone within the old WG could contribute to the development of each part
- Different maturity of Open Science “elements” evident

CERN Open Science Policy: Implementation Plan

V1.0

Authors and contributors: Members of the Open Science Strategy Working Group,
April 2023 at CERN

Contact: open-science@cern.ch

Table of Contents

Preamble - Governance, Communication and Monitoring	2
1. Open Access to Publications [Editors: Anne Gentil-Beccot, Alex Kohls, Kamran Naim]	3
2. Open Data and Reuse [Editors: Sunje Dallmeier-Tiessen, Sebastian Neubert]	5
3. Open Source Software [Editors: Clemens Lange, Zach Marshall, Axel Naumann]	6
4. Open Hardware [Editors: Myriam Ayass, Javier Serrano]	8
5. Research Integrity, Reuse and Reproducibility [Editors: Sunje Dallmeier-Tiessen, Clemens Lange]	9
6. Infrastructure Provision for Open Science [Editors: Sunje Dallmeier-Tiessen, Jose Benito Gonzalez, Lukas Heinrich, Clemens Lange]	11

Policy and implementation plan are starting points

- Road ahead is ambitious and requires coordination and (new) collaboration on almost all topics of the policy
 - A challenge and an opportunity to find new synergies, ideas, projects
- Working with the community
 - Identifying best practices, share, learn together, build new projects together
 - to identify KPIs to monitor policy implementation
 - bi-annual CERN Open Science Report
 - a regular Open Source report etc.
- Overarching Goal: Better accessibility of CERN's various outputs