Open Science at CERN

Dr Sünje Dallmeier-Tiessen, Chair of the OS Office at CERN

sunje.dallmeier-tiessen@cern.ch

January 2024

Agenda

1 Introduction to Open Science at CERN

2 CERN Open Science Policy

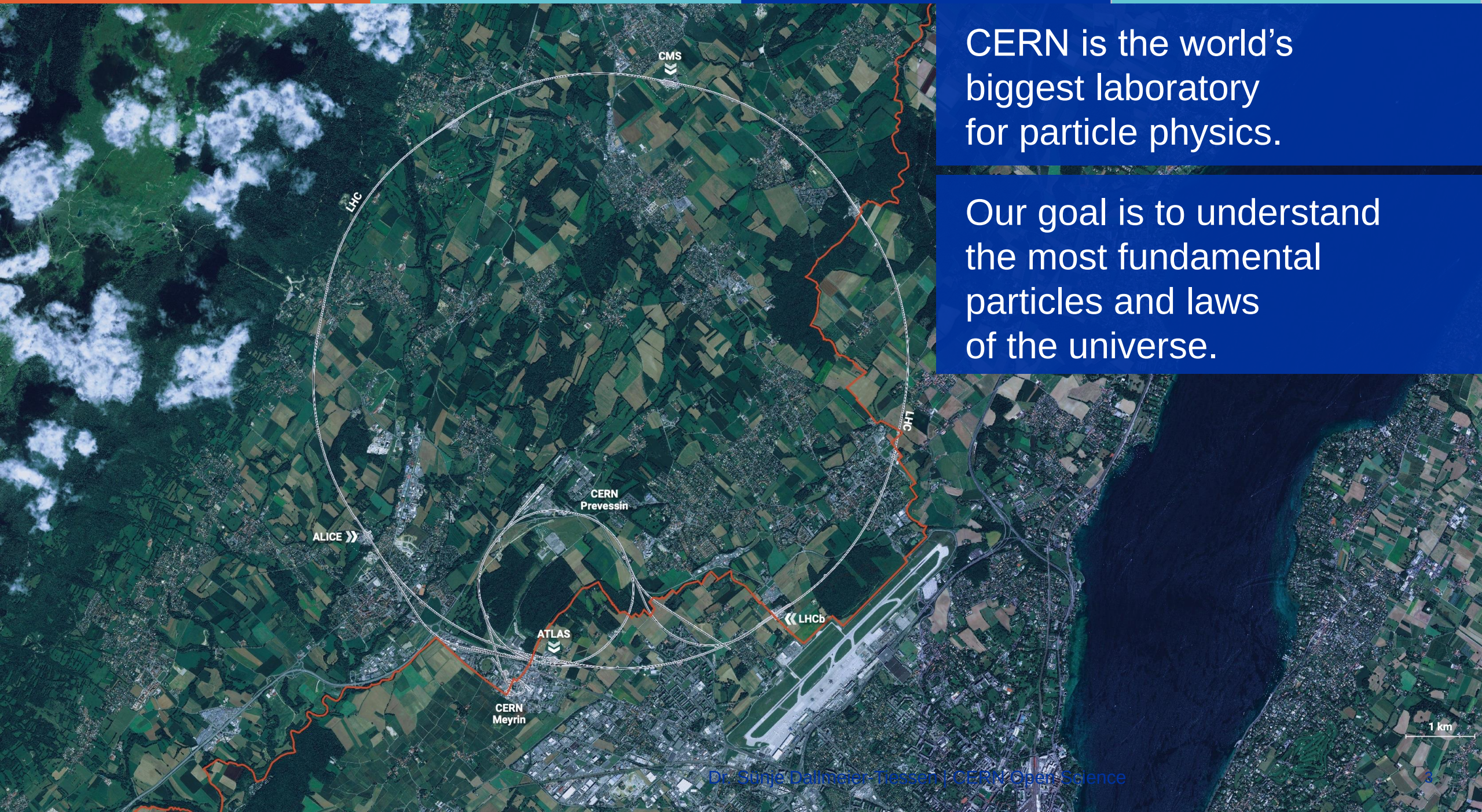
3 Open Science Implementation Plan

4 Selected projects und updates

5 Next steps

CERN is the world's biggest laboratory for particle physics.

Our goal is to understand the most fundamental particles and laws of the universe.



Four pillars underpin CERN's mission

What is Open Science? The UNESCO definition

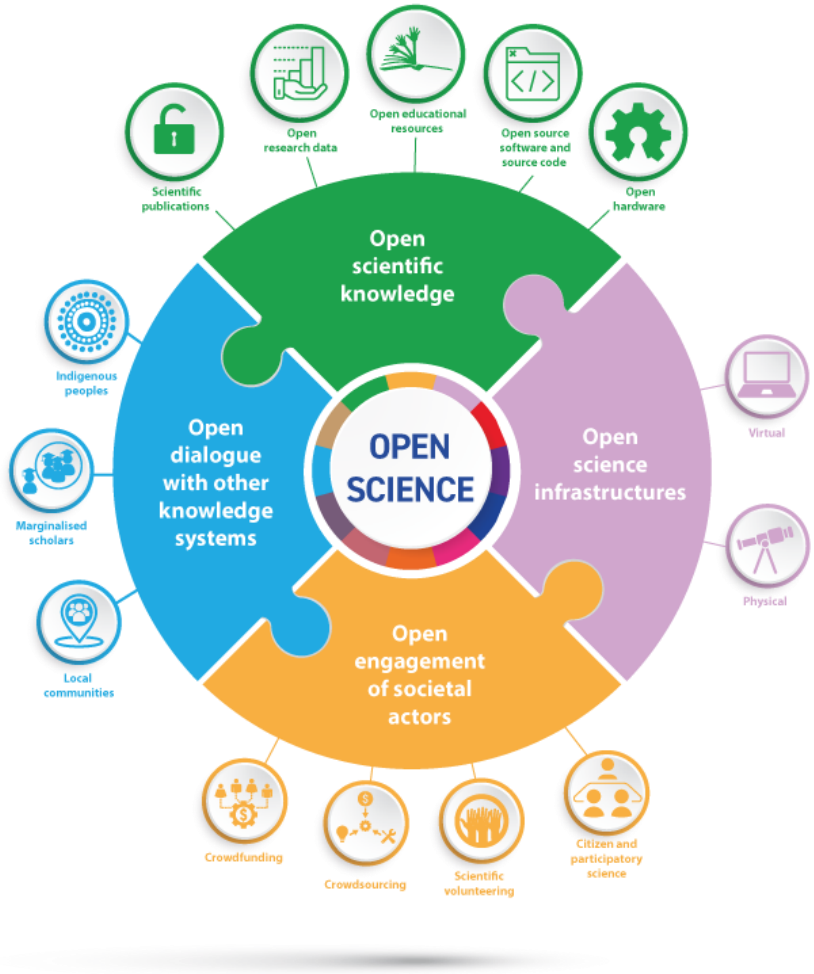
Open science is defined as an inclusive construct that combines various movements and practices aiming

to make multilingual scientific **knowledge openly available, accessible and reusable** for everyone,

to increase scientific **collaborations** and **sharing** of information for the benefits of science and society, and

to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community.

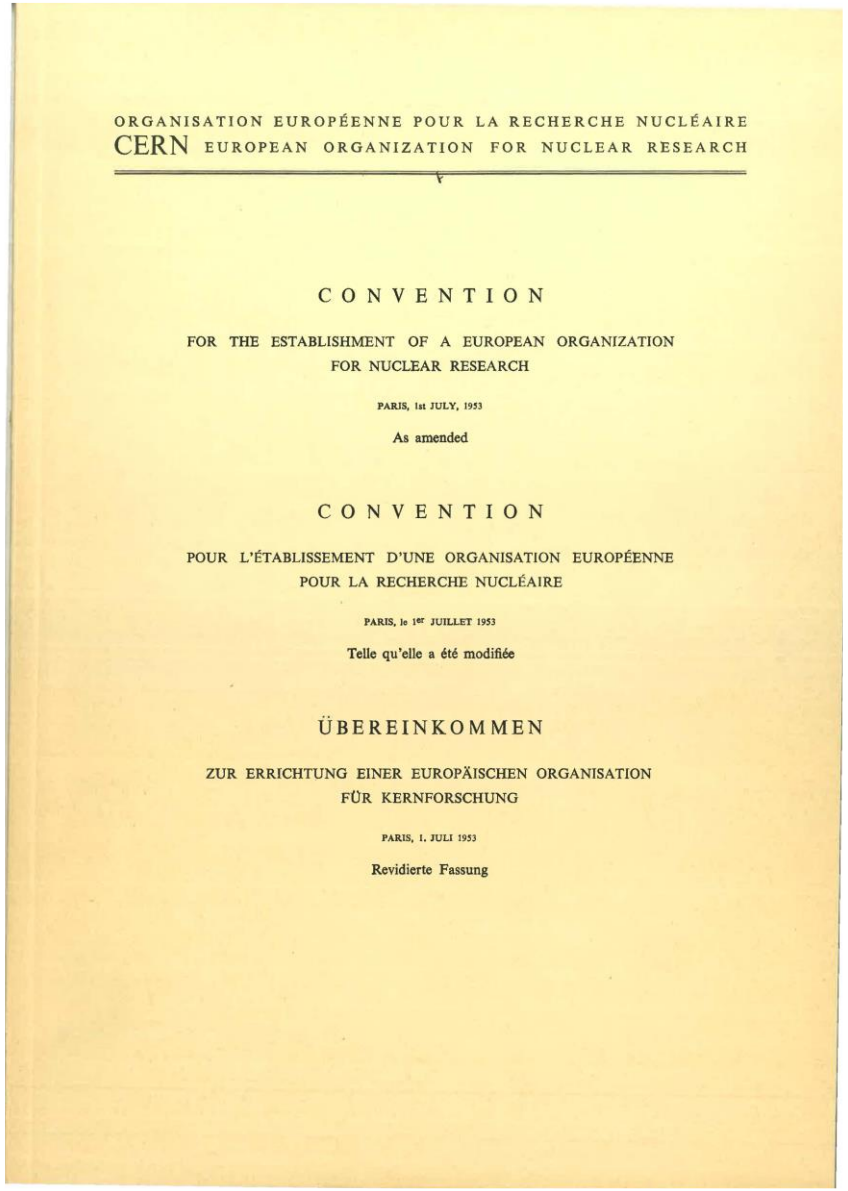
[…] it builds on the following key pillars: open scientific knowledge, open science infrastructures, science communication, open engagement of societal actors and open dialogue with other knowledge systems.



"UNESCO Recommendation on Open Science" (2021) https://doi.org/10.54677/MNMH8546

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 1er JUILLET 1953

Telle qu'elle a été modifiée

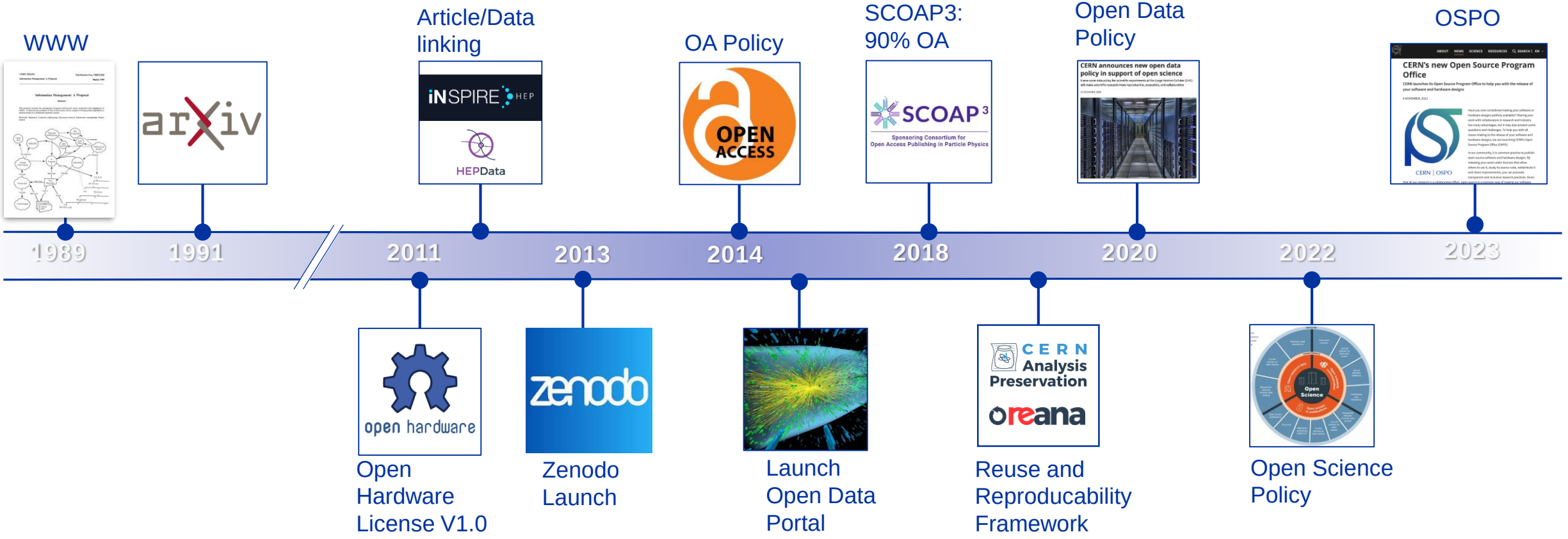
ÜBEREINKOMMEN

ZUR ERRICHTUNG EINER EUROPÄISCHEN ORGANISATION
FÜR KERNFORSCHUNG

PARIS, 1. JULI 1953

Revidierte Fassung

Driving Open Science Globally



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
  - Research Integrity, Reuse & Reproducibility
  - Infrastructure for Open Science
  - Research Assessment & Evaluation
  - Education, Training & Outreach
  - Citizen Science
- Policy to be regularly updated to reflect changes in landscape, practices, funder requirements & community demands
- 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 Access

1. Open access to publications

All CERN scientific publications are to be made immediately publicly available and reusable. The Open Access Policy for CERN Publications (2014, updated 2017 and 2021) requires that all original research publications by CERN authors are published open access, centrally supported by the CERN Open Access fund. CERN users and visiting scientists are also encouraged to publish their work under similar terms, according to the CERN General Conditions applicable to the Execution of Experiments.

CERN scientific publications, including submissions to trusted repositories (such as arXiv), should be released under an open licence, with CC-BY as the default standard. Publication-related metadata are made available for reuse under the CC0 waiver in line with FAIR principles (findability, accessibility, interoperability, reusability). Open access publishing support is also provided for monographs related to CERN experiments or accelerators, applied research processes or technologies, and other areas of relevance.

https://cds.cern.ch/record/2835057

Open Data

2. Open data

CERN experimental collaborations are committed to making their research data publicly available. The CERN Open Data Policy for the LHC Experiments (2020) aims to support CERN experiments' consistent approach towards the openness and preservation of experimental data to maximise their long-term value. All data are released with persistent identifiers. Data and associated data services apply open and FAIR principles. For experimental data releases, CC0 waivers are applied as standard. Researchers and experiments are expected to develop data management plans for their research activities.

https://cds.cern.ch/record/2835057

Open Source software

3. 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. External communities should be invited to use and contribute to the evolution of CERN's software projects. CERN's software expertise should be shared with other science disciplines. Software development processes are expected to follow best practices[4]. CERN contributes to open source software relevant to its mission through code contributions, participation in the evolution of software, and standardisation.

https://cds.cern.ch/record/2835057

Open Hardware

4. Open hardware

CERN makes its technologies broadly available to society and has introduced open hardware licensing as a key mechanism to achieve this goal. Open hardware designs are made available through the Open Hardware Repository. The legal basis for the sharing of open hardware is enabled through variants of the CERN Open Hardware Licence. Hardware design releases will consider opportunities for collaboration with other research communities and industry. In cases where extensive documentation and ancillary components like software for interfacing and testing are required for projects, these should be licensed under appropriate open source documentation and software licences respectively.

https://cds.cern.ch/record/2835057

Research Integrity, Reuse and Reproducibility

5. Research integrity, reuse and reproducibility

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. Reuse and reproducibility are facilitated by practising comprehensive analysis preservation to capture relevant research objects, such as research data releases with supporting metadata, auxiliary data, linked software, reproducible analysis workflows, documentation, etc.

https://cds.cern.ch/record/2835057

Who wrote the policy?

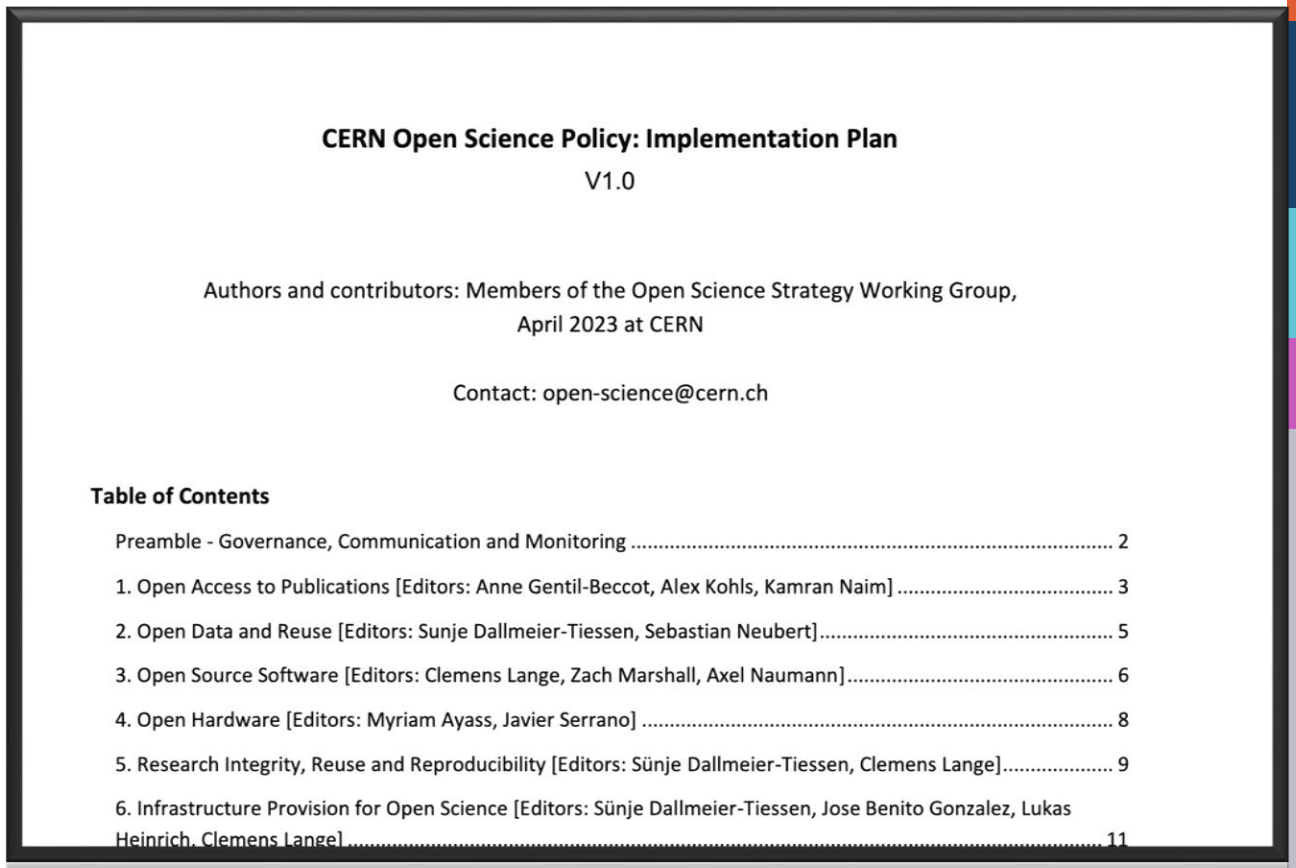
- CERN Open Science Strategy Working Group was established in summer 2021 with two principal objectives:
  - to create a framework for a regular and proactive platform for all active stakeholders in Open Science at CERN; and
  - to develop an organizational Open Science Policy for CERN
- Working Group consisted of representatives from across departments and experiments
- Diverse perspectives on Open Science, challenging consensus building process

Open Science Implementation Plan

Building on a vast experience across the organization

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

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

Examples from the implementation plan

Software

[…] Software projects are recommended to consult a future CERN Open S… e (OSPO) and the CERN Knowledge Transfer (KT) group for topics such as licenc…

The KT group at CERN is tasked to make assessments o… pplications, and advises on licensing matters for software with di… ergy physics. The OSPO is in the planning phase. […]

Hardware

[…] Whenever a t… design which could be generally useful beyond the laboratory… nowledge Transfer (KT) group. KT will be able to recommend the best li… n, including through open-sourcing that design under one of the three varia…]

OUTDATED!

Selected Open Science Highlights

Launch of CERN‘s Open Source Program Office (OSPO)

https://indico.cern.ch/event/1327562/

CERN Open Source

- **Open Source is fully embedded in CERN culture**
  - WWW
- **CERN Software available via Open Source Initiative licenses**
  - Software to analyse experimental data, such as ROOT
- **CERN hosts Open Source projects**
  - Zenodo - platform for sharing research output
- **CERN contributes to other Open Source initiatives**
  - Software for IT infrastructures
- **CERN established an Open Hardware License**
  - White Rabbit to distribute precise timing across distributed systems

CERN Open Source Program Office: Mandate

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.

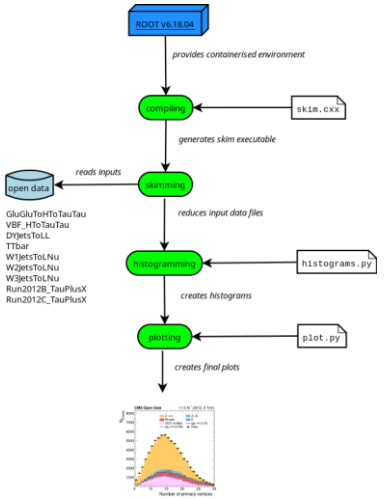
Contact: Open.Source@cern.ch
https://opensource.cern/
Mandate: http://cds.cern.ch/record/2879995

CERN Open Data grew to over 4 petabytes

- Disseminating research-grade event-level particle physics data since 2014
- In **2023**, the content grew to over **4 petabytes**
- **CMS completed Run-1 heavy-ion data release**
  - About 560 TB of 2013 and 2015 proton-proton reference collision data, proton-lead collision data, and simulations (announcement)
- **LHCb completed Run-1 proton-proton data release**
  - About 600 TB of additional data streams (announcement)

Fostering reuse and reinterpretation

"The adaptable software examples [are] the most efficient way to pass on the knowledge needed for research-level studies on these data" — CMS



opendata CERN ⇌ reana

Data are being released with accompanying software and **analysis examples**

New "**Continuous Reuse**" system using REANA to ensure the validity of data usage patterns (paper)

Zenodo - update

- **Celebrated Zenodo's 10th anniversary**
- **Launched Zenodo on new technical platform (InvenioRDM) - major milestone for:**
  - Making collaboration easy
  - FAIR-enabling features (ORCID, ROR, PIDs, …)
  - Enabling collaboration with external partners on development
- **HORIZON-ZEN project progress:**
  - Grant from European Comission to CERN
  - Zenodo will host the EC's official digital repository service – i.e. similar to the EC's Open Research Europe platform for publishing.
  - Pilot launching end-March, fully operational September/October.

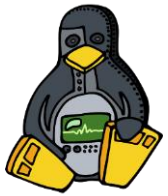
"Why do I like Zenodo? Because Zenodo is FAIR. Fair in sense of lowercase, and FAIR in the sense of uppercase meaning, […]"

Open Access: 10th Anniversary of SCOAP3

- The Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP3)
- A global collective action of libraries, research institutions and funding agencies from 44 countries with a mission to deliver barrier-free open access publishing in HEP
- Since 2014, SCOAP3 has funded the OA publishing of over 62,000 articles across 11 leading journals in the discipline (including JHEP, EPJC, PLB, NPB, PRD, PRC, PRL and others)
- Average investment per article of around EUR 1250 (a fraction of APCs charged by similar journals)
- Phase 4 to commence in 2025 will include incentivization of Open Science elements to enable improved interoperability of systems and quality of service

Open Science “products” at CERN

This is only a selection of projects!



opendata CERN

open hardware

zenodo

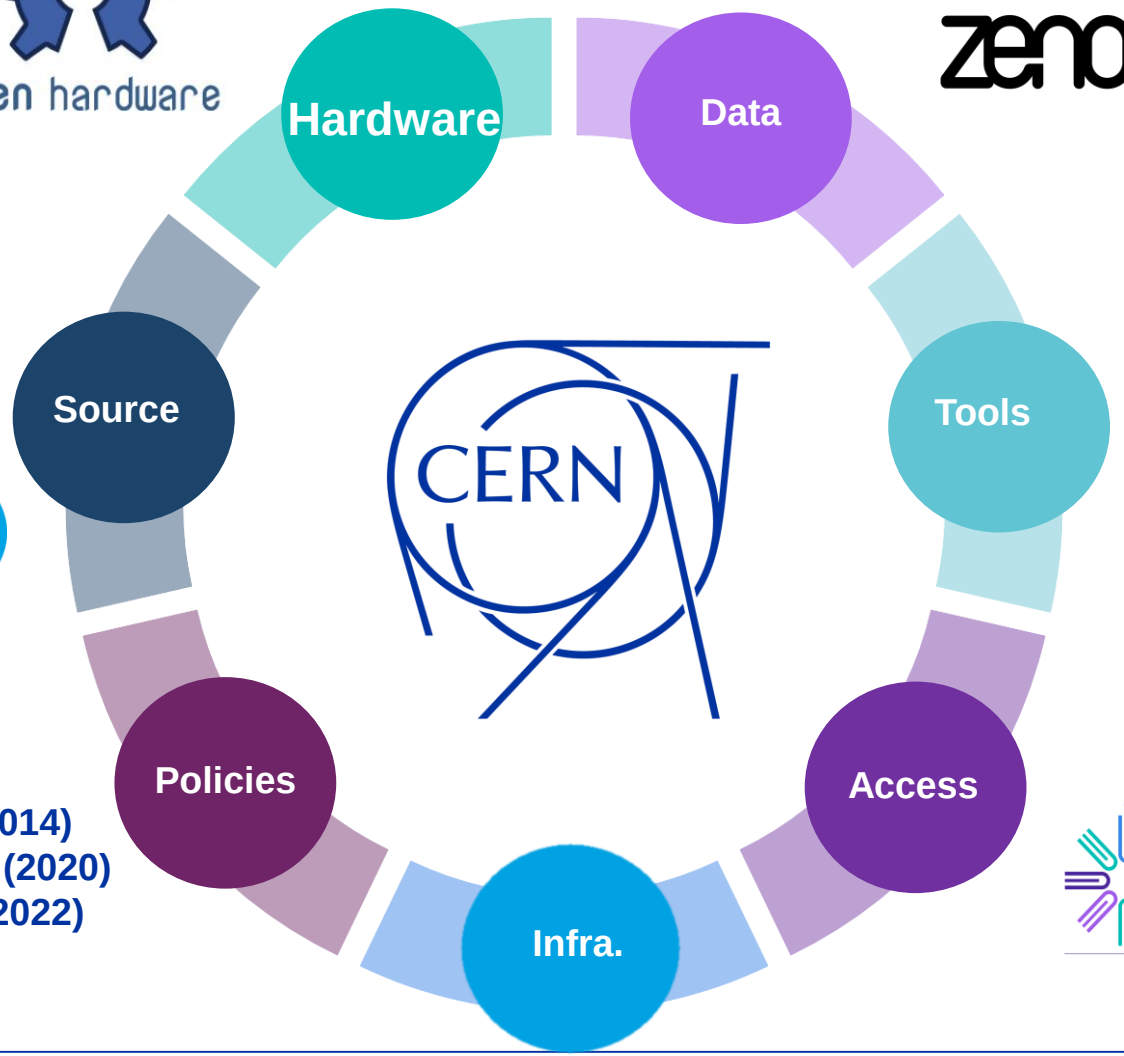
ROOT Data Analysis Framework

INVENIO

CERN Analysis Preservation

indico

reana



SCOAP3

Open Access (2014)
LHC Open Data (2020)
Open Science (2022)

Summary

- More and more expectations on Open Science from our (associate) member states and society at large.
- Our community with a long tradition of Open Science – for certain Open Science elements.
- Highlight Open Access: 10 years of SCOAP3
- Highlight Open Data: Petabyte release of Open Data
- Highlight Open Source: Launch of the new Open Source Program Office (OSPO)
- Highlight 10 Years of Zenodo service
- Community approach: learn from each other, support each other and build collaboratively.

2024 Next steps

- Leveraging the Open Science Practitioners Forum: more community work!
- Advancing the monitoring framework, including discussion of KPIs and qualitative indicators
- Preparation of Open Science Report during Q3/4 2024
- ML/AI and Open Science project
- Support OSPO operations
- Start conception of Open Science training

Thank you!

Contact:
sunje.dallmeier-tiessen@cern.ch