Accelerating Open Science

February 2024

Dr. Kamran Naim
Head of Open Science
European Organization for Nuclear Research (CERN)

CERN Open Science

Overview

CERN Introduction

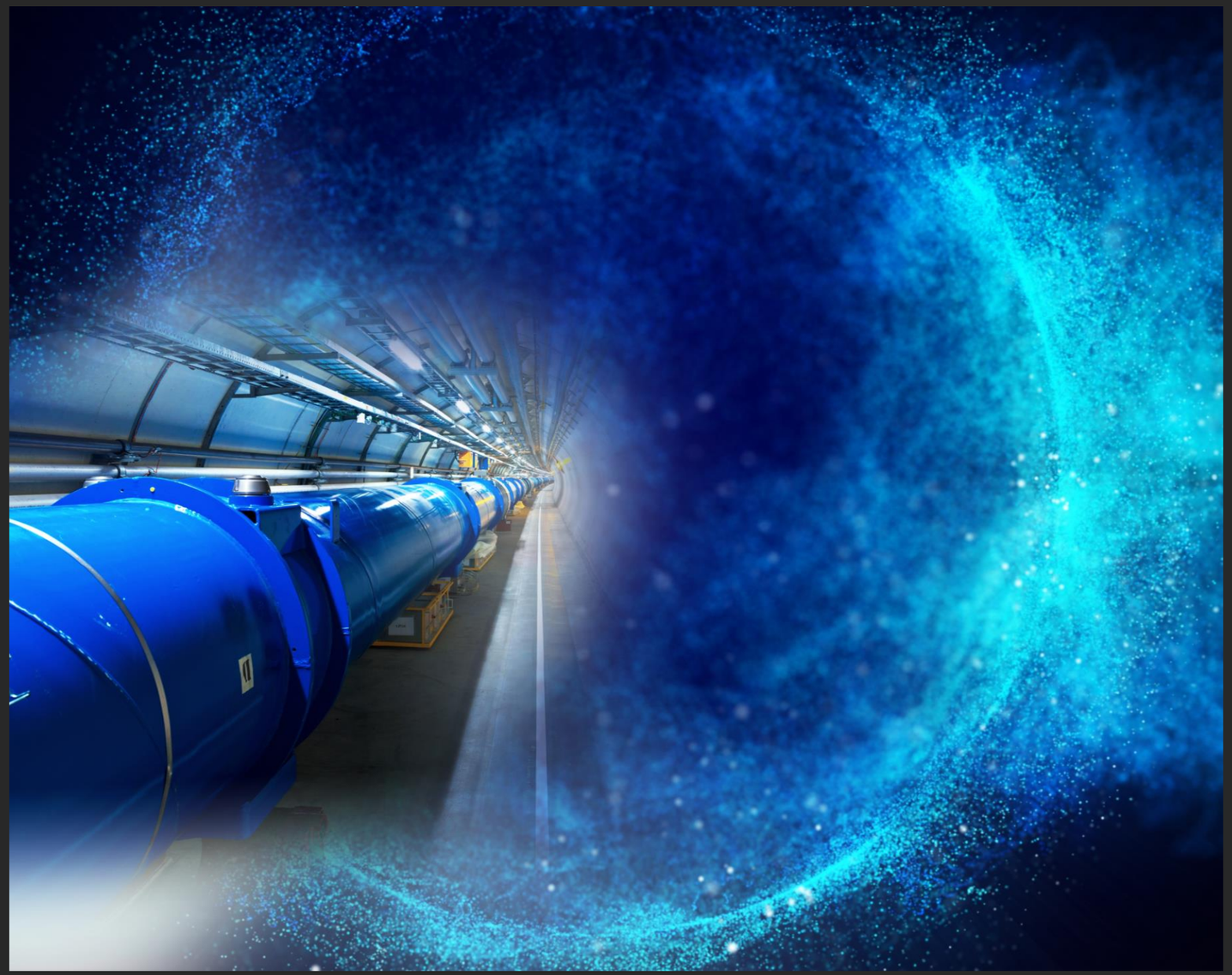
Holistic perspective on Open Science

Case studies: SCOAP3 and Zenodo

Aligning Policies and Practice

Conclusions

CERN Mission



- Perform world-class research in fundamental physics;
- Provide particle accelerator facilities in an environmentally responsible and sustainable way;
- Unite people from all over the world to push the frontiers of science and technology;
- Train new generations of physicists, engineers and technicians;
- Engage all citizens in research and in the values of science.

CERN
Organisation and Statistics

- 24 Member States
- 2 Pre-stage Membership States
- 8 Associate Member States
- 4 Observers

- 1 400 000 000 CHF Budget

- 2,600 Personnel
- 2,000 Contractors
- 13,000+ Scientists and Researchers
- 110 Nationalities

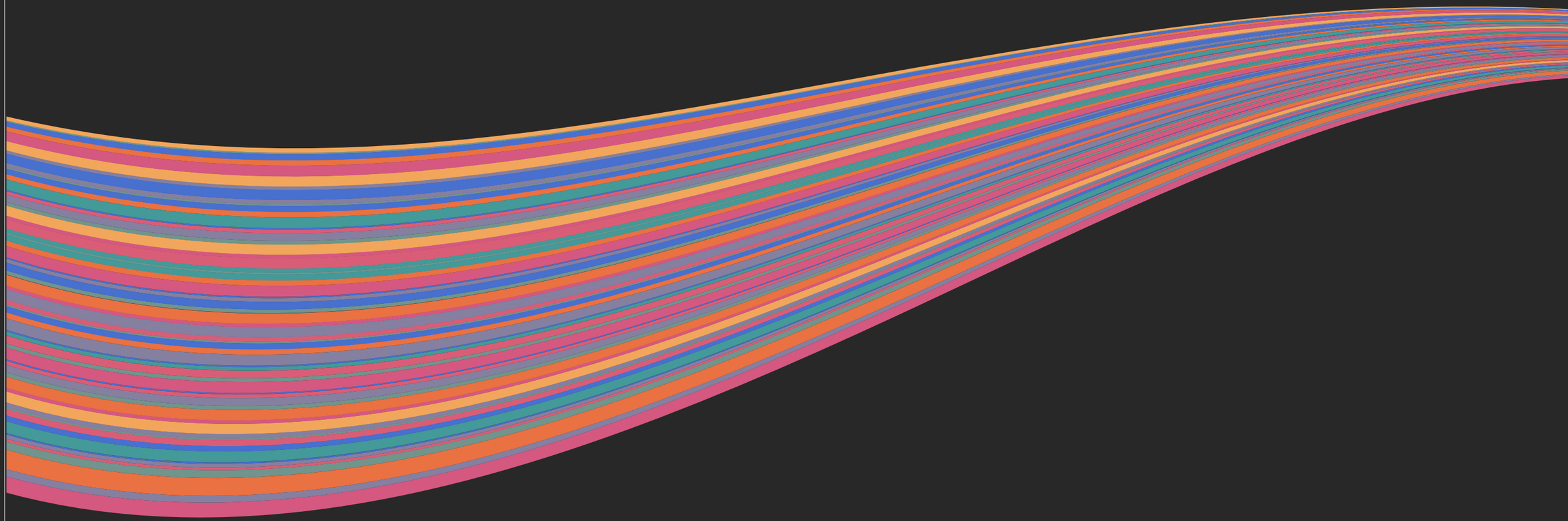
The Large Hadron Collider

CERN's flagship

- 100m underground
- 27km circumference
- Coldest known place in the universe (-271°C)
- Hottest place in the solar system (100000x the sun)
- 1 billion particle collisions per second
- Detectors collect 1PB per second of data

“… and the results of its experimental and theoretical work shall be published or otherwise made generally available.”

CERN CONVENTION, 1954



Open Access
1991–2024

1991
arXiv.org

2014
CERN Open Access Policy

2024
SCOAP^3 Introduces Open Science Elements

2014
Launch of SCOAP^3

2018
APS joins SCOAP^3, 90% of OA in HEP

CERN Open Science

Open Access
Mechanisms

SCOAP³

Sponsoring consortium for Open Access publishing in Particle Physics.

COLLECTIVE MODELS

CERN supports other OA Model transparent for the author (S2O, sponsorship).

OA AGREEMENTS

CERN has negotiated Open Access Agreements with 10+ publishers for 4000+ journals.

INDIVIDUAL APC

For other articles fees can be centrally covered under Certain conditions.

Open Data

CERN
Open Science

Open Data
2012–2024

2012
CMS Open Data Policy

2012
HEPdata connected with INSPIRE

2014
CERN Open Data Portal

2020
LHC Open Data Policy

2024
More than 5000TB of data in CERN Open Data Portal

CERN Open Science

Open Data Experimental Programme

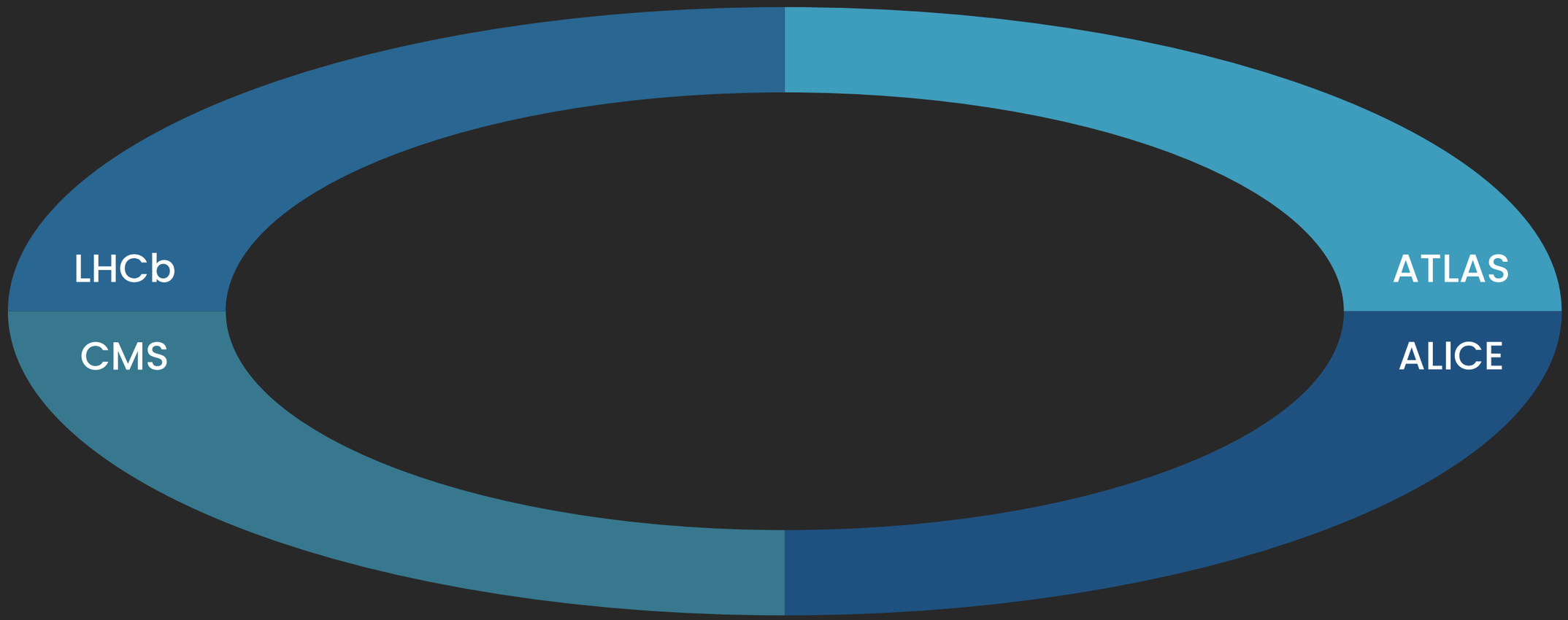
CERN´s flagship experiment is the

Large Hadron Collider

CERN Open Science

Open Data:
Experimental Programme

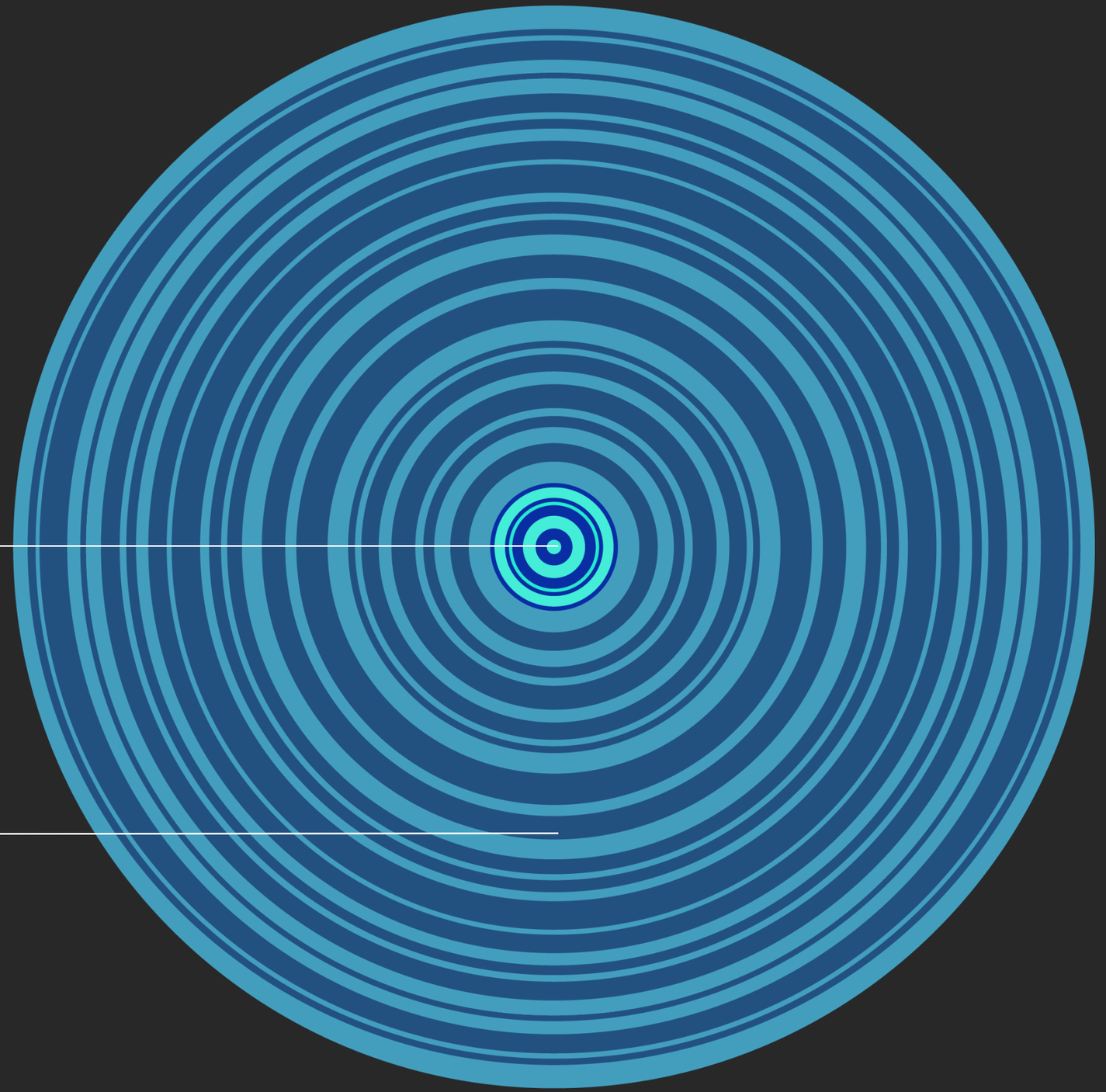
LHC Major Experiments



CERN Open Science

Open Data

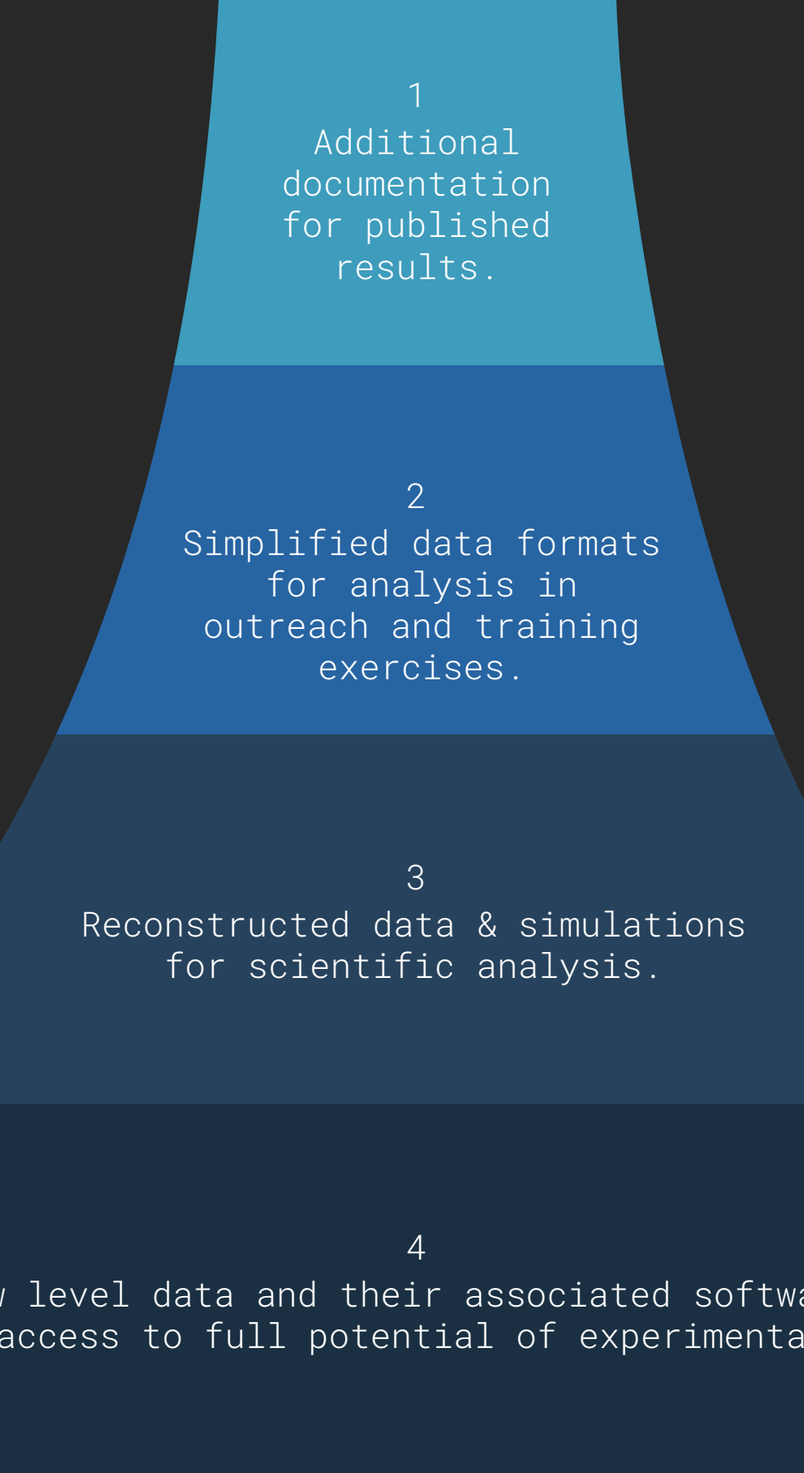
LHC & High-Lumi Data production



LHC: 50-60PB/Year

High-Lumi: 600PB/Year

Open Data
Data Levels



Akopov et al., Status report of the DPHEP Study Group: Towards a global effort for sustainable data preservation in high energy physics. arXiv preprint arXiv:1205.4667 (2012).

Open Data
Policy for LHC Experiments

Full Policy Text

- Published in 2020
- All experiments engage into opening datasets to the world.
- 4 levels of data: commitment is for levels 1, 2, 3, due to level 4 volumes and complexity.
- All data are released with persistent identifiers.
- Data and associated data services apply open and FAIR principles.
- CC-0 waivers are applied as standard.
- Researchers and experiments are expected to develop data management plans for their research activities.
- Level 3 Releases: calibrated reconstructed data with the level of detail useful for algorithmic, performance and physics studies.
- Release of these data will be accompanied by:
  - provenance metadata;
  - simulated data samples;
  - Analysis software;
  - Reproducible example analysis workflows; and
  - Virtual computing environments.

Dr. Kamran Naim

CERN Open Science

Open Data Portal

Portal Access

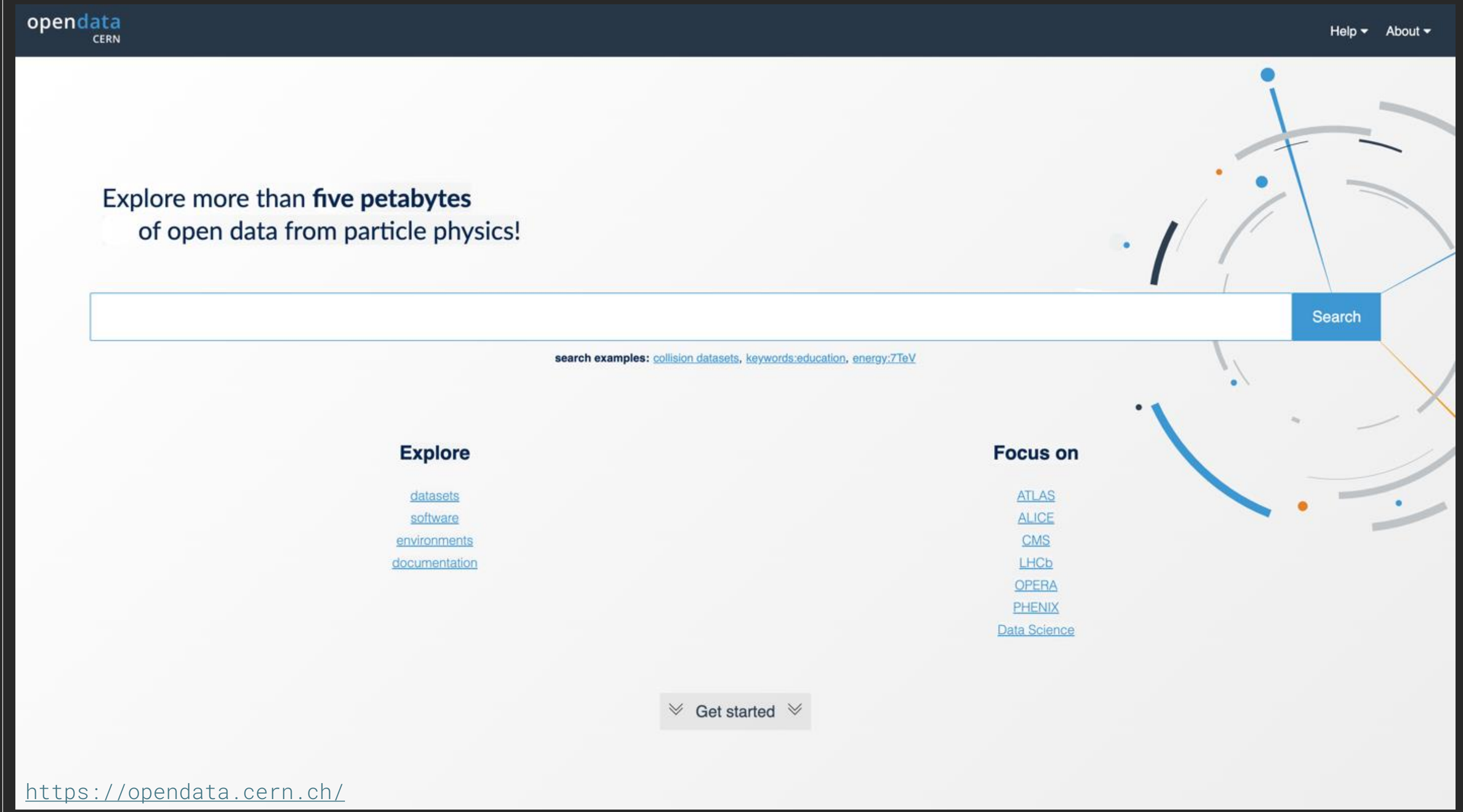
Portal to Access data form the LHC Experiments and beyond.

BUILD WITH THE EXPERIMENTS FOR THE PHYSICS WORLD

- Digital repository for event-level particle physics open data.
- Rich content:
  - Collision and simulated datasets for research.
  - Derived datasets for education.
  - Virtual machines and container images.
  - Software tools and analysis examples.
- Launched in November 2014, publishing about 30TB.
- Total size in June 2024: about 5000TB.

USES CASES

- Education-oriented & Research-oriented.



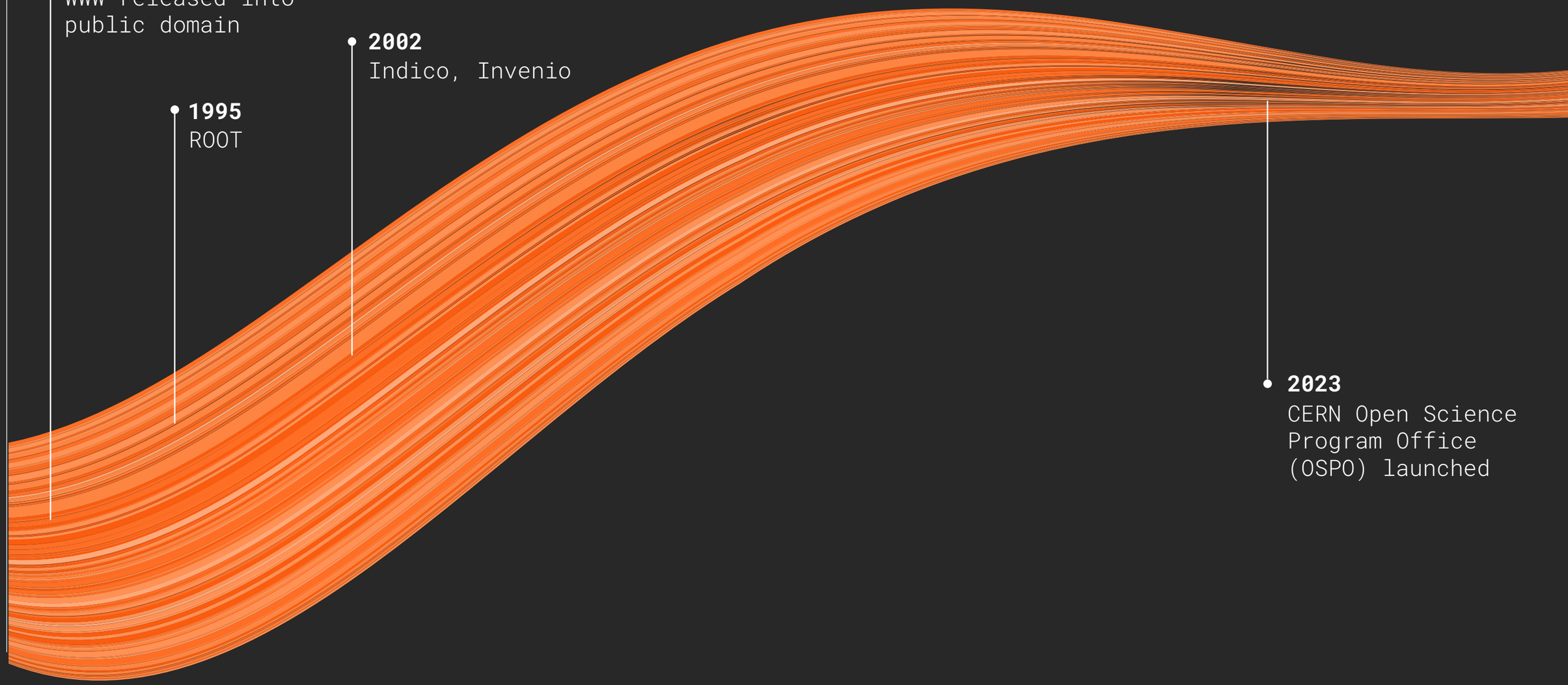
https://opendata.cern.ch/

Open Software
1993–2023

1993
WWW released into public domain

1995
ROOT

2002
Indico, Invenio



2023
CERN Open Science Program Office (OSPO) launched

CERN Open Science

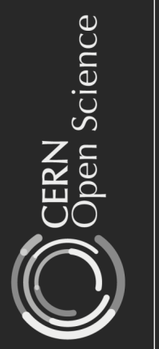
Open Source Library Technologies

Invenio Software

Open source software framework for large-scale digital repositories

Originally developed to support CERN Document Server

Dr. Kamran Naim



Open Source Library Technologies

Built on Invenio Software

Used to host the Zenodo Service- the world's largest multidisciplinary research repository

Developed in collaboration with OpenAIRE to support implementation of their Open Ddta policy in 2013

Zenodo service now available to support the Digital Repository needs of global research communities worldwide

- 400,000 daily users
- 9,000 organizations
- 161 countries
- 28 million visits per year
- 5 million research outputs / over 1 PB of data

Dr. Kamran Naim

CERN Open Science

CERN Open Science

Open Source Library Technologies

INVENIO RDM

Turn-key research data management (RDM) repository platform based on Invenio Framework and Zenodo.

Designed to be safe, scalable, RESTful, and open.

Key features of Invenio:

- Security: Invenio is designed with security and long-term preservation in mind.
- Scalability: Invenio can manage petabytes of files and over 100 million records.
- RESTful: Invenio is JSON-native and provides RESTful APIs that allow apps to be built on top of it.
- Open: Invenio is open source and licensed under MIT.

Built by CERN in partnership with community of institutions and companies, including Brookhaven National Laboratory, Caltech Library, and Northwestern University,

Open Source

Reach beyond High-Energy Physics

SOFTWARE

- Conference management (Indico).
- Document management (InvenioRDM).
- Storage (CTA, CERNbox, CernVM-FS).
- Data analysis/computing (ROOT, BioDynaMo, EMP2,
  REANA, SWAN, Rucio).

HARDWARE

- High-precision timing (White Rabbit).
- EDA tools (kiCAD).

Open Hardware
2008–2024

2008
White Rabbit

2009
Open Hardware Repository

2011
CERN Open Hardware License V1

2011
CERN joins KiCad development

2020
CERN OHL V2

2023
CERN OSPO

2023
KiCad supported through commercial company

2024
White Rabbit Collaboration established

Open Source
at CERN

PRODUCER

CERN regularly creates or updates numerous open-source products, such as
- analysis or experiment software (e.g. root),
- general software projects (e.g. Indico),
- hardware design (e.g. 3D-printed face mask).

CONTRIBUTOR

CERN extensively contributes to a wide range of worldwide open-source initiatives.

CONSUMER

More than 70% of open-source software used at CERN is based on external projects; most CERN services rely on components such as Python, Kubernetes and the Linux Kernel.

GOVERNANCE DRIVER

CERN has been a driver of several open source infrastructure governing attempts, e.g. by creating the CERN Open Hardware licenses.

Open Source Program Office Mandate

INTERNAL MANDATE

- Consult, advice, train on Open source best practices, tools licenses, etc.
- Advices on open-sourcing CERN software and hardware.
- Identify dependencies and compatibility for critical services.
- Advices CERN on Open Source matters.

EXTERNAL MANDATE

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

Contact: Open.Source@cern.ch
opensource.cern
Mandate: cds.cern.ch/record/2B79995

CERN Open Source Program Office

STRATEGIC

OS Policy
OSPO Mandate

OSPO Outreach
OS Partnership

INTERNAL

EXTERNAL

Licensing
Tips & Tools

OSPO Catalogue
CLA Management

OPERATIONAL

Incentives
2022–2024

2022
CERN signs CoARA Agreement

2024
Preliminary CoARA Action Plan prepared

Research Asesstment

CoARA

1. Recognize the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research.

2. Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators.

3. Abandon inappropriate uses in research assessment of journal and publication-based metrics, in particular inappropriate uses of Journal Impact Factor (JIF) and h-index.

4. Avoid the use of rankings of research organizations in research assessment.

Research Assessment

CoARA at CERN

The CoARA principles are well aligned with CERN Values and established particles physics practices.

To further enhance this alignment, the following actions will be taken in the next five years:

- Promote open science, teaching outreach activities.
- Acknowledge diverse career trajectories within the institution.
- Acknowledge the societal impact of research taking place at CERN.

Policies
2014-2025

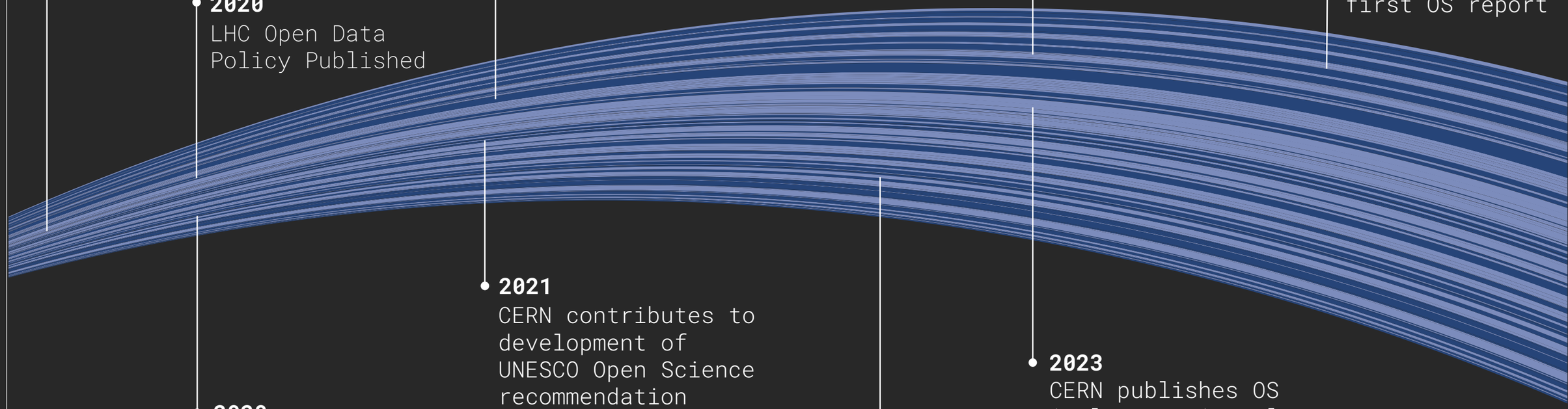
2014
CERN Open Access Policy Published

2020
LHC Open Data Policy Published

2021
CERN Open Science Strategy Working group to devise organisational OS policy

2023
OS Governance Framework
- implement of OS Policy
- align with CERN strategy

2025
CERN to publish first OS report



2020
ESSP Update states:
"The particle physics community should work with the relevant authorities to help, shape the emerging consensus on Open Science to be adopted for publicly-funded research, and should then implement a policy of Open Science field."

2021
CERN contributes to development of UNESCO Open Science recommendation

2022
CERN publishes its first organizational OS Policy

2023
CERN publishes OS implementation plan

CERN Open Science

Open Science Implementation

- Policy accompanied by implementation document outlining roles, responsibilities, mechanisms & resources
- Includes actionable measures to support the policy’s implementation across the organization and in the experiments@CERN
- Implementation led by Communities of Practice across the various OS activities
- Identifies/surfaces both existing and required resources for policy implementation
- Published openly here.

Open Science
Governance Framework

Creating efficient and effective coordination of CERN´s Open Science efforts.

OPEN SCIENCE STEERING BOARD

- To oversee OS efforts at the strategic level.
- Reports to DRC.
- Inter-departmental body (currently 15 members).

OPEN SCIENCE PRACTITIONERS' FORUM

- De-facto continuation of the OSWG
- CERN-wide Exchange amongst different OS expert groups.
- OSPF elects a member to represent it at the OSSB.

OPEN SCIENCE OFFICE

- Organisationally set up in RCS-SIS-OS
- Support of different expert groups, OSPF and OSSB, coordinates communication efforts and creation of OS Report.
- Provides additional concrete services to CERN community (e.g. advice on creating Data Management Plans)

Open Science
New Governance Framework

Creating efficient and effective coordination of CERN´s Open Science efforts.

DRC

Open Science Steering Board

OS Practitioners Forum (OSPF)

- Open Data WG
- WG
- CoARA Implementation WG
- Interest Group

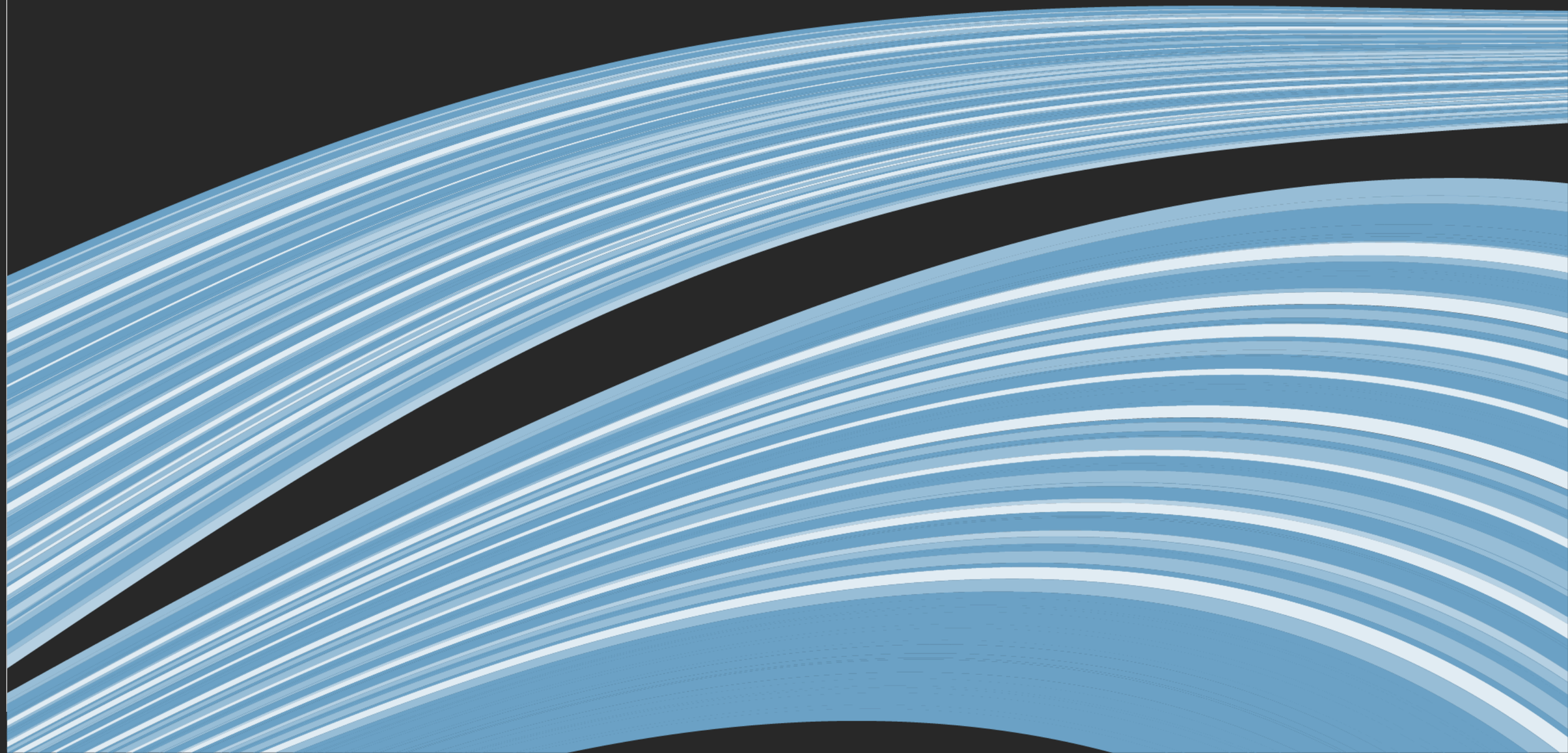
OS Office

- SIPB
- OSPO

CERN Open Science

CERN Infrastructure

A set of technological tools and services that support and enable open science practices.



CERN Open Science

Open Infrastructure

Services & Activities

CERN Open Data Portal	CERN Analysis Preservation	INSPIRE	REANA (Reproducible Analysis)
CERN Document Server	Indico	Zenodo	Open Hardware Repository

Organizational OS Infrastructure

Services & Activities

OS Governance	Central OS Support Office	OSPO	Technical Infrastructure
Decision Making Priority Alignment	Community Support, Outreach, Education	Internal & External Mandate	Mediates OS Practice

CERN Open Science

Open Science Implementation

- Grassroots efforts aligned with evolving policy frameworks, organizational resources, and technical developments to support implementation.
- Building sustainable solutions requires resources and cooperative approaches.
- Public goods well supported through collaborative approaches/governance models.
- Building community around technological solutions can support development and sustainability.
- Interface between research communities mediated by infrastructure: software, services.
- Collaboration is key!

Thank You

Dr. Kamran Naim
Head of Open Science
European Organization for Nuclear Research (CERN)
kamran.naim@cern.ch

CERN Open Science