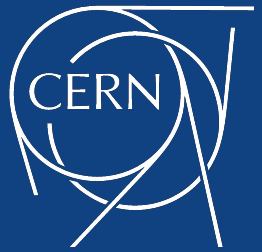
Open Data Principles in Publishing

A viewpoint from High-Energy Physics

Sünje Dallmeier-Tiessen
Artemis Lavasa

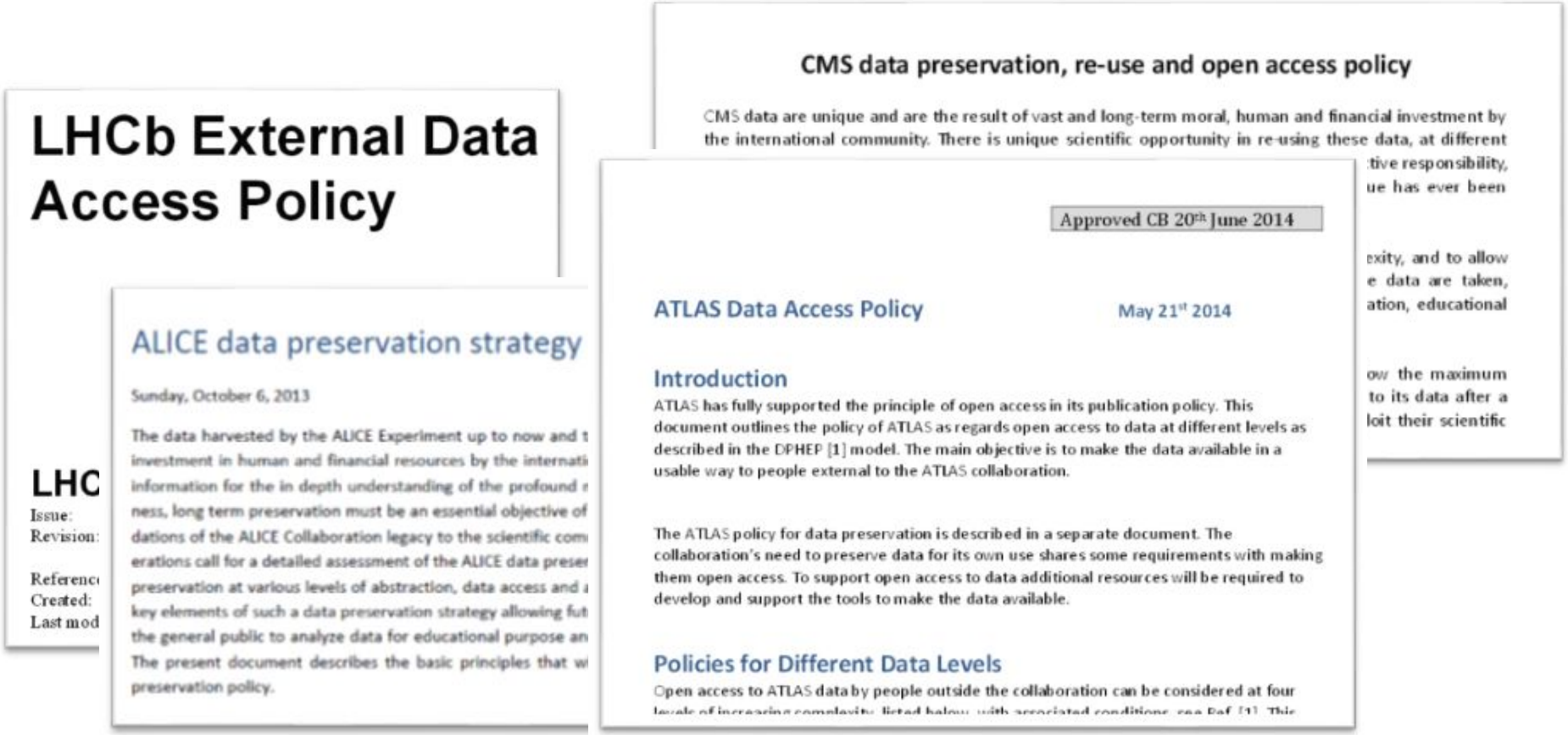
London, June 23rd 2017


CERN

CERN - Particle Physics

- Founded in 1954
- Intergovernmental research organisation
- 22 members states
- ~2500 employees
- 12,000 visiting scientists from over 70 countries and with 120 different nationalities
- A different dimension of “collaborative research”
- Central place for HEP: everyone either comes through here or uses CERN data

Data Policies in HEP



LHCb External Data Access Policy

LHC
Issue:
Revision:
Reference
Created:
Last mod

ALICE data preservation strategy

Sunday, October 6, 2013

The data harvested by the ALICE Experiment up to now and t
investment in human and financial resources by the internati
information for the in depth understanding of the profound
ness, long term preservation must be an essential objective of
dations of the ALICE Collaboration legacy to the scientific com
erations call for a detailed assessment of the ALICE data prese
preservation at various levels of abstraction, data access and
key elements of such a data preservation strategy allowing fut
the general public to analyze data for educational purpose an
The present document describes the basic principles that w
preservation policy.

CMS data preservation, re-use and open access policy

CMS data are unique and are the result of vast and long-term moral, human and financial investment by the international community. There is unique scientific opportunity in re-using these data, at different
tive responsibility,
ue has ever been

exity, and to allow
e data are taken,
ation, educational

ow the maximum
to its data after a
loit their scientific

Approved CB 20th June 2014

ATLAS Data Access Policy

May 21st 2014

Introduction

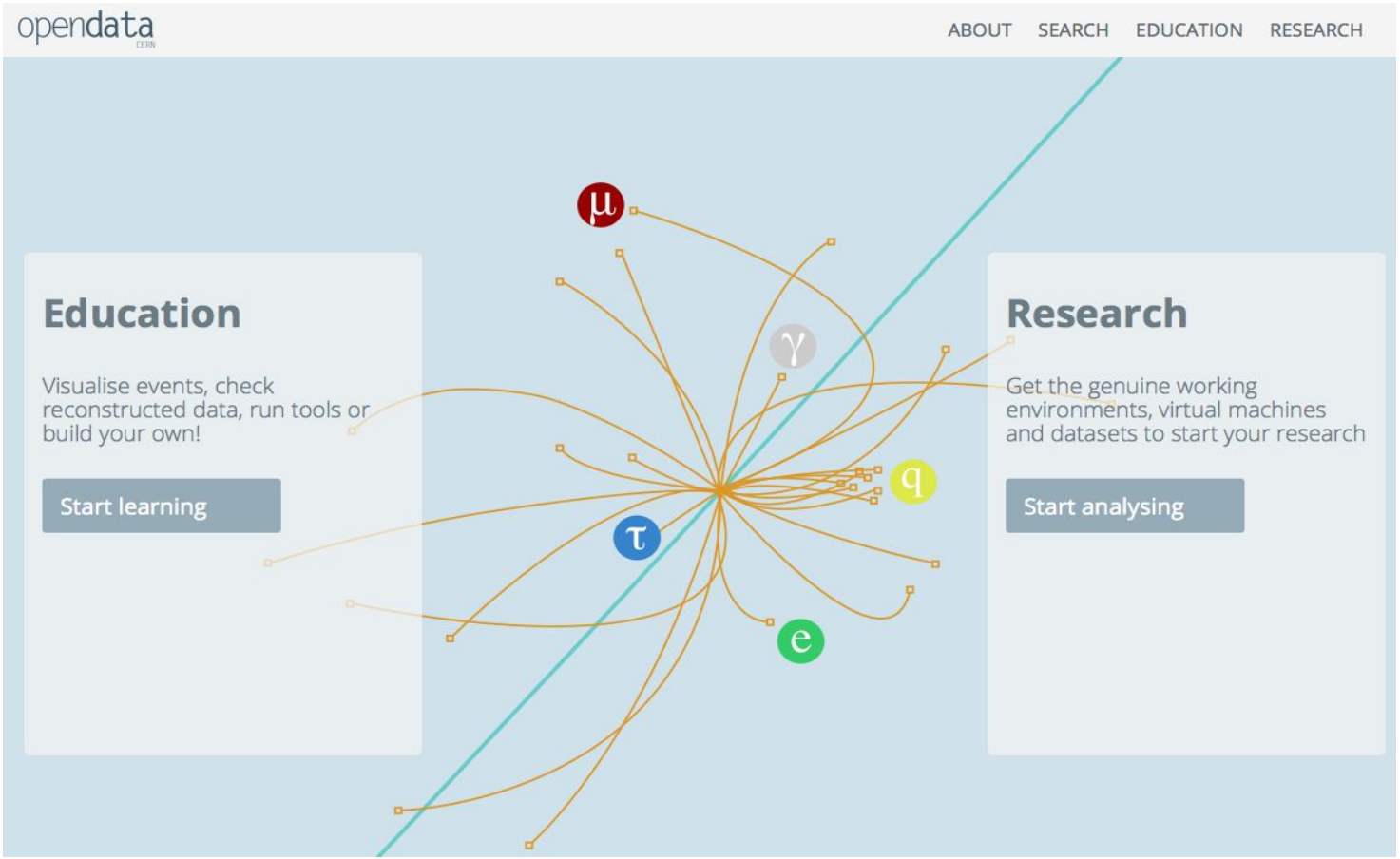
ATLAS has fully supported the principle of open access in its publication policy. This document outlines the policy of ATLAS as regards open access to data at different levels as described in the DPHEP [1] model. The main objective is to make the data available in a usable way to people external to the ATLAS collaboration.

The ATLAS policy for data preservation is described in a separate document. The collaboration's need to preserve data for its own use shares some requirements with making them open access. To support open access to data additional resources will be required to develop and support the tools to make the data available.

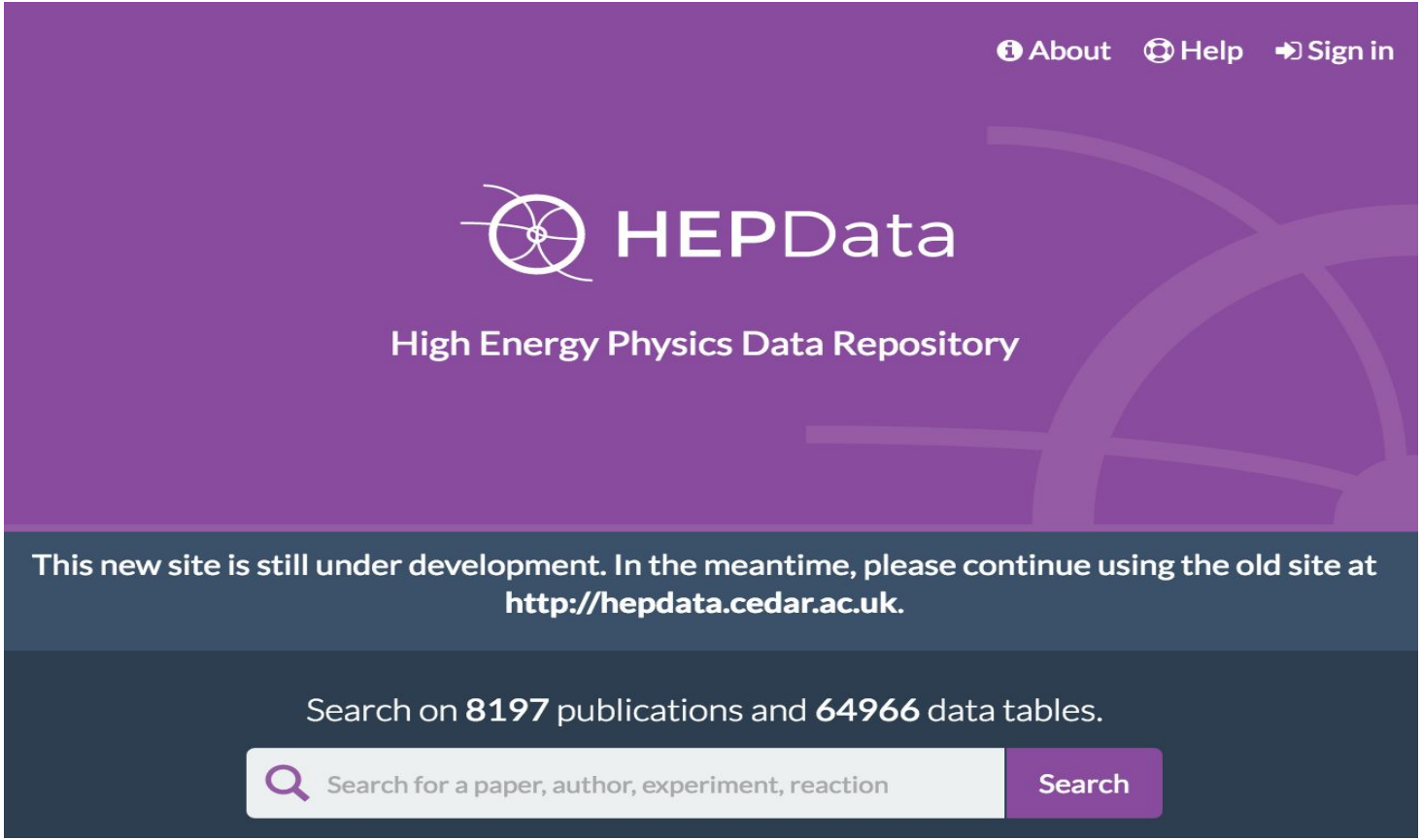
Policies for Different Data Levels

Open access to ATLAS data by people outside the collaboration can be considered at four

CERN Open Data (opendata.cern.ch)



HEPData (hepdata.net)



Gettings “things done”: Community workshop

Information Summit from/for Astronomy, Astrophysics and High-Energy Physics

- Workshop series taking place every 1.5 years with all stakeholders from publishing, repositories, indexing services, researchers

Last one: May 2017, focused on data, software, ORCID etc.

- What are the challenges specific to this community?
- How can we address them?
- What are the next steps?

Discussion and results from workshop

Data and Software

- Focus on data availability statements and data citation
- Springer Nature example showed flexibility in data policies: opportunity for community to tailor to specific needs
- Pressure vs. feasibility, mandates vs. voluntary measures

Challenges and solutions:

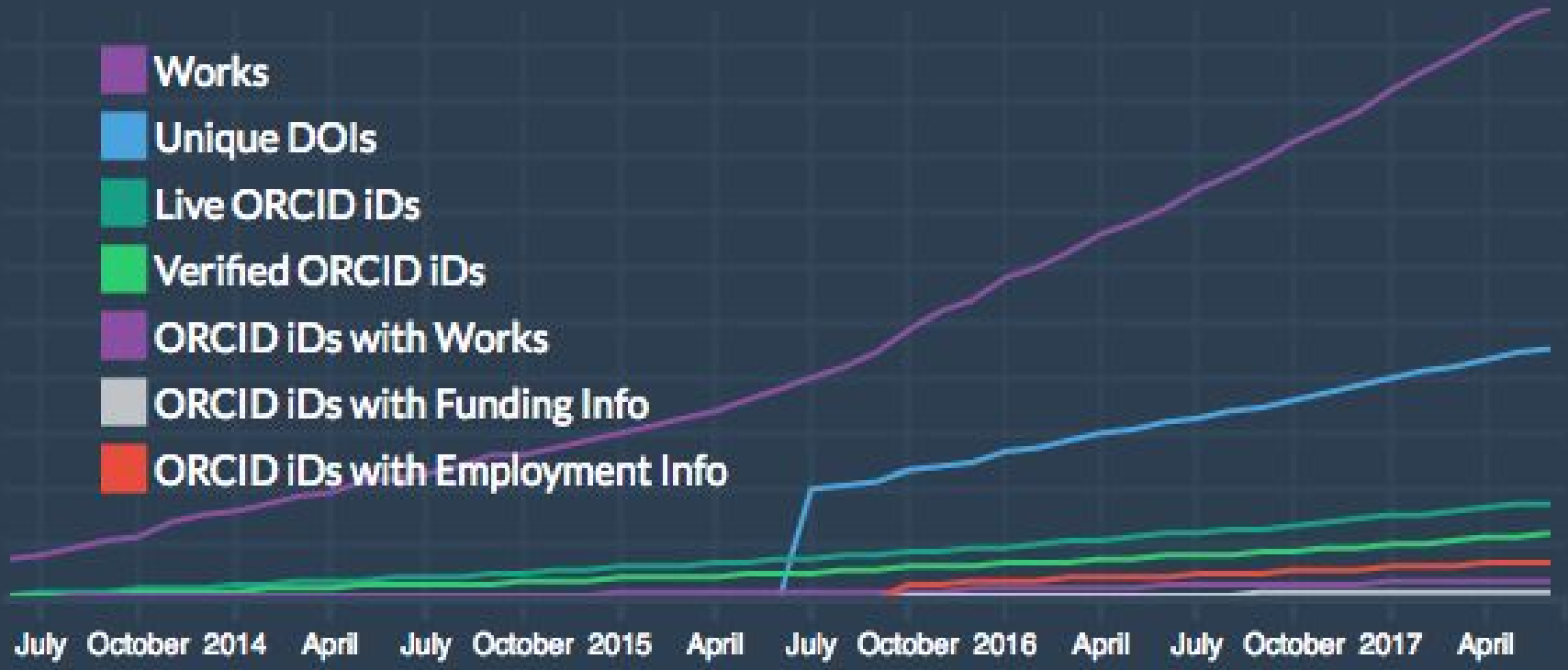
- Community-relevant publishers chair discussion on potential policies and implementations - take the driver's seat
- Indexing services: give data more visibility
- Improve collaboration and communication, e.g. lessons learnt
- Need for community awareness: organise joint sessions at conferences

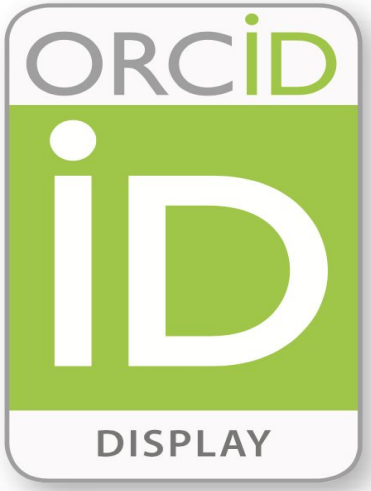
ORCID: What is it?

- An identifier for researchers
- A registry
- A set of standard procedures for connecting researchers to their affiliations and activities
- A committed community building connectors
- An international-scale open research effort
- If implemented well, could make “submissions” (repositories, journals) and updates easier

Core challenge in HEP: interoperability and adoption

overview


Works
Unique DOIs
Live ORCID iDs
Verified ORCID iDs
ORCID iDs with Works
ORCID iDs with Funding Info
ORCID iDs with Employment Info
July
October
2014
April
July
October
2015
April
July
October
2016
April
July
October
2017
April

ORCID
COLLECT

ORCID
iD
DISPLAY

ORCID
CONNECT

ORCID
SYNCHRONIZE

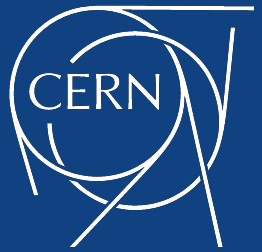
Technical and Human infrastructure for Open Research

project-thor.eu

Our goal is to ensure that every researcher, at any phase of their career, or at any institution, will have seamless access to Persistent Identifiers (PIDs) for their research artefacts and their work will be uniquely attributed to them

Lessons Learnt

- Work with each individual community and take into account the challenges specific to them
- Take hesitation, expectations and different positions into consideration
- Discuss openly advantages and disadvantages of possible approaches
- One-size-fits-all approaches can be challenging; providing adjustable solutions may work better
- Learn from experience (e.g. Springer Nature)



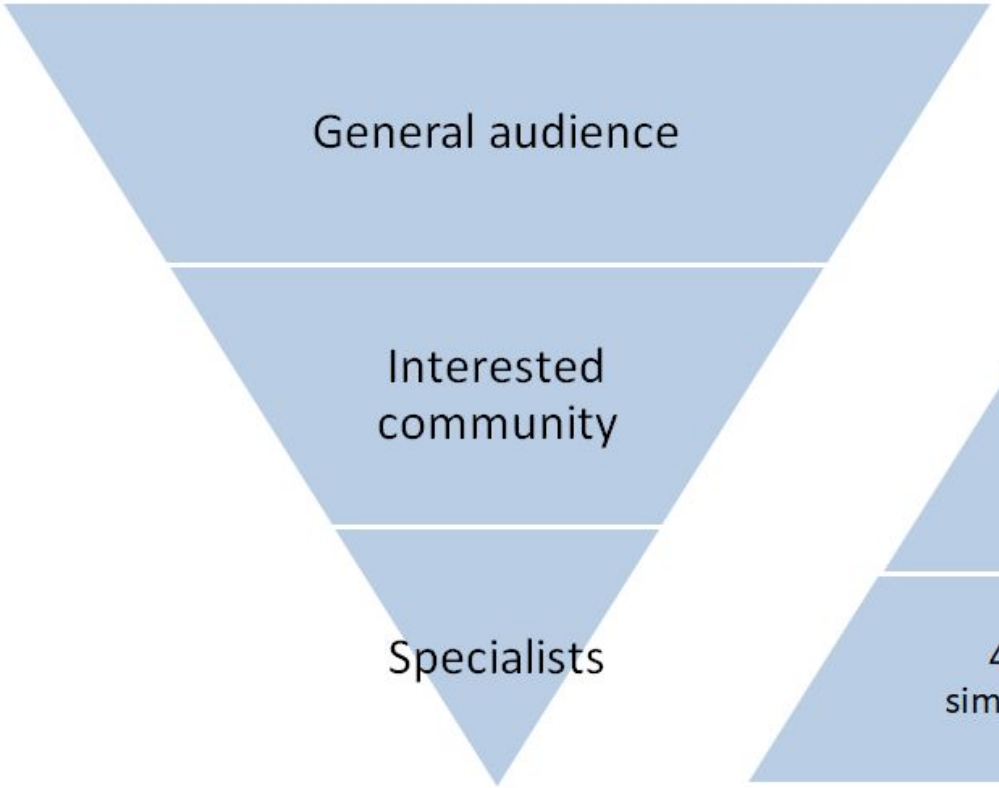
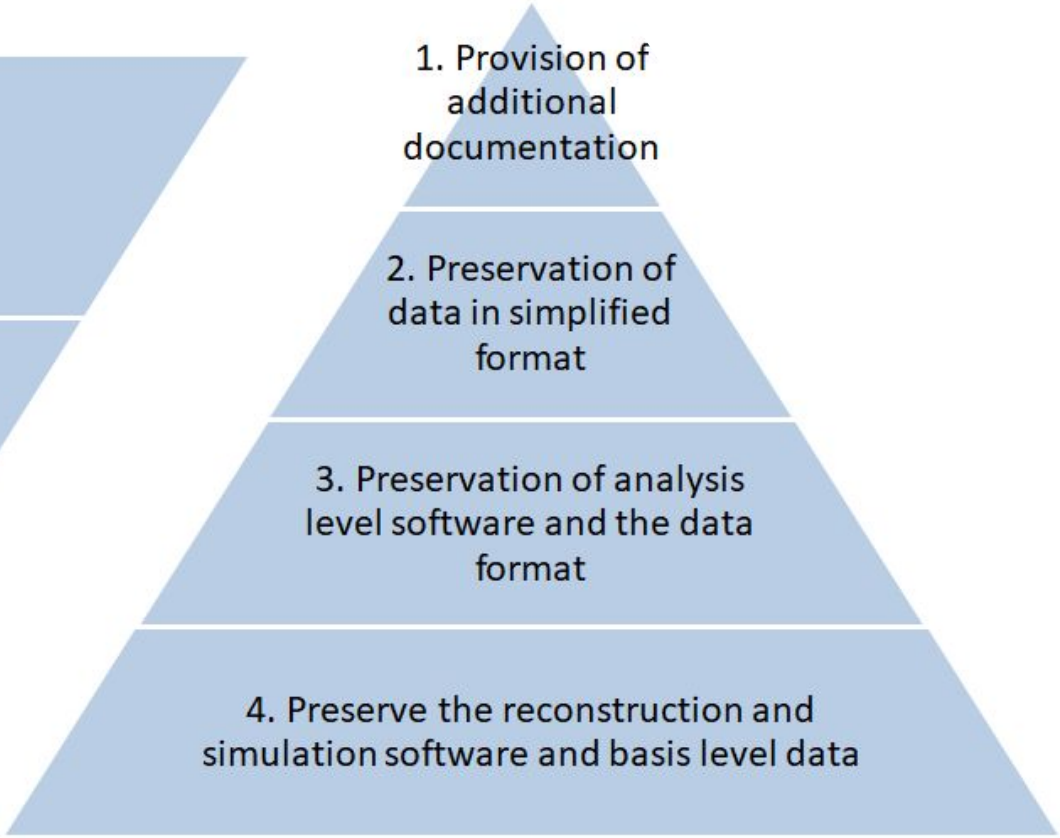
sunje.dallmeier-tiessen@cern.ch
artemis.lavasa@cern.ch

Backup Slides

Metrics

- Traditional metrics:
  - important role in community,
  - special role of indexing services
- Alternative metrics:
  - low interest in HEP community so far
  - Important to consider a wide range of sources
- Metrics for new scholarly materials, e.g. data and software
  - There is “little” to count still, more of a social challenge
  - Important to start “counting” the publications first, prepare to show reuse ⇒ already implemented

LHC Policies

- Open data for three levels
- Immediate release for level 1 (with publication)
- Embargo periods for level 3
- Exploitation within the collaboration
- CC0 – public domain dedication
- Data citation

Data in HEP - Use Cases