Open Data @CERN

Sünje Dallmeier-Tiessen
for many others in GS-SIS and IT-CIS
CERN Open Data Portal

Released Nov 2014

http://opendata.cern.ch/

Github repo

https://github.com/cernopendata/

Based on Invenio 2.0
CERN launches Open Data Portal to make public the data of LHC experiments cern.ch/go/tN15T #cernopendata
At the public release:
- serving ~15 GB per hour [usage ~50 times higher than it is now]
- The next day or two after the public release was about ~4 GB per hour
A few days ago, CERN launched an Open Data Portal to publicly share data from the Large Hadron Collider. We are some of the scientists behind this project, working to make science more open globally. Ask Us (Almost) Anything about open data, open access, data preservation, big data and open science!

Hi reddit!

We unveiled the CERN Open Data Portal to the world recently, releasing samples for education from all the main LHC experiments and around 27 TB of high-level and analysable LHC data from the CMS Experiment.

Following CERN’s last AMA, we’re thrilled to be here today to talk to you not only about open science but also our Open Data Portal, #cernopendata and the tools you can build on top of our data. We are:

- From CERN Information Technology:
  - Tim Smith, Head of Collaboration and Information Services (tjs)
  - Jamie Shiels, Project leader, Data and Knowledge Preservation in High-Energy Physics (js)
  - Tibor Simko, Technology Lead for the Open Data Portal (ts)

- From CERN Scientific Information Service:
  - Salvatore Mele, Head of Open Access (sm)
  - Sünje Dallmeier-Tiessen, Open Science Research Fellow (sdt)

- From the CMS Experiment:
  - Kati Lassila-Perini, Physicist and Co-ordinator of the CMS Data Preservation and Open Data project (klp)
  - Tom McCauley, Physicist and Developer of CMS education/outreach tools (tm)

We’ll sign our posts with our initials (see above) so you know who said what. Just to be clear, we are speaking with you in our personal capacities and CERN does not necessarily support the views expressed during the AMA. Joining us are a few of our friends from CERN:

- Kate Kahle (/u/kate_kahle), CERN social-media manager
- Achintya Rao (/u/RaoOfPhysics), CMS science communicator and Science Communication doctoral student
- Patricia Herterich (/u/Phriterich), Data librarian and Open Science doctoral student

We’ll answer your questions from 16:00 CET until 17:30 CET (UTC+01).
If the government funds scientific research why isn’t that science published openly and freely? Why is so much scientific articles hidden behind paywalls that make it impossible to research something without an institution supporting you? How can we change the system for the better?

Here at CERN we believe in Open Access, and have published openly and freely all articles from the LHC experiments in peer-reviewed journals. The (c) stays with the authors, and the articles are available under a Creative Common license for everyone to read, re-post and re-use.

We agree with you that we can change the system for the better, and together with partners in 40 countries we have been organizing for most of the results in particle physics to be now published Open Access, without paywalls, through the SCOAP3 initiative.

This is a great step. Thank you. Information is so much more powerful when it's available. You lead the way, and I hope that many more people follow. Thanks for your time.

Can you recommend anything for a small school physics course learning about electromagnetism and Lorentz force?

Have a look at the tracks of charged particles in the magnetic field inside the CMS experiment. Load an event in the event display, turn it to the x-y plane and observe the track curvature. (klp)

Wow. Thanks!

This is going on our physics forum :)

Also take a look at the educational resources on the portal, they're aimed mostly for high-school students: http://opendata.cern.ch/resources
Usage

Typical month now:
- ~1000 visitors with real file access/handling [EOS]
- ~200 visitors ‘lurking’ around
- resulting in various amounts GB being served

- Data reuse for physics and outreach
- Re-use of primary datasets for machine learning
- Adaption of code examples for new analysis
CERN Analysis Preservation

• “closed counterpart” to CERN Open Data that captures the complexity of
  – The data
  – The processing steps
  – Code involved
  – Documentation, Physics information
  – Peer review, QA
i.e. all the information contributing to the research claim/presentation/publication to enable future reuse
Preserve an analysis
Creating the knowledge base
Thanks to

**CERN IT** J. Cowton, J. Delgado, J. Kunčar, M. Neumann, T. Smith, T. Šimko

**CERN SIS** A. Dani, P. Fokianos, P. Herterich, L. Rueda

**ALICE** M. Gheata, C. Grigoras

**ATLAS** K. Cranmer, L. Heinrich, D. Rousseau, F. Socher

**CMS** A. Calderon, A. Huffman, K. Lassila-Perini, T. McCauley, A. Rao, A. Rodriguez Marrero

**LHCb** S. Amerio, M. Bettler, B. Couturier, T. Head, A. Trisovic, A. Ustyuzhanin

**CERN CernVM** J. Blomer

**CERN EOS** L. Mascetti

**DASPOS** M. Hildreth, C. Vardeman, G. Watts

**DPHEP** F. Berghaus, J. Shiers