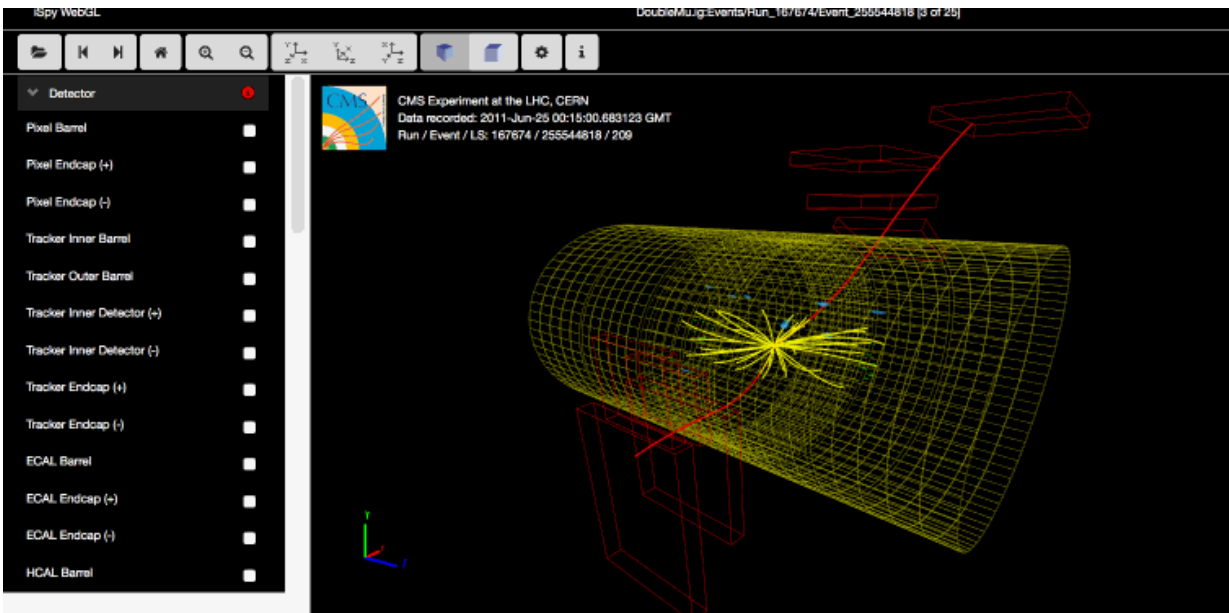


CERN CMS releases 300 terabytes of research data from LHC

April 25 2016, by Achintya Rao



A CMS collision event as seen in the built-in event display on the CERN Open Data Portal

Today, the CMS Collaboration at CERN has released more than [300 terabytes \(TB\) of high-quality open data](#). These include over 100 TB, or 2.5 [inverse femtobarns](#) (fb^{-1}), of data from proton collisions at 7 TeV, making up half the data collected at the LHC by the CMS detector in 2011. This follows a [previous release from November 2014](#), which made available around 27 TB of research data collected in 2010.

Available on the CERN Open Data Portal—which is built in collaboration with members of CERN's IT Department and Scientific Information Service—the collision data are released into the public domain under the CC0 waiver and come in types: The so-called "primary datasets" are in the same format used by the CMS Collaboration to perform research. The "derived datasets" on the other hand require a lot less computing power and can be readily analysed by university or high-school students, and CMS has provided a limited number of datasets in this format.

Notably, CMS is also providing the [simulated data](#) generated with the same software version that should be used to analyse the primary datasets. Simulations play a crucial role in particle-physics research and CMS is also making available the protocols for generating the simulations that are provided. The data release is accompanied by [analysis tools and code examples](#) tailored to the datasets. A virtual-machine image based on [CernVM](#), which comes preloaded with the software environment needed to analyse the CMS data, can also be downloaded from the portal.

These data are being made public in accordance with CMS's commitment to long-term data preservation and as part of the collaboration's [open-data](#) policy. "Members of the CMS Collaboration put in lots of effort and thousands of person-hours each of service work in order to operate the CMS detector and collect these [research data](#) for our analysis," explains Kati Lassila-Perini, a CMS physicist who leads these data-preservation efforts. "However, once we've exhausted our exploration of the data, we see no reason not to make them available publicly. The benefits are numerous, from inspiring high-school students to the training of the particle physicists of tomorrow. And personally, as CMS's data-preservation co-ordinator, this is a crucial part of ensuring the long-term availability of our research data."

The scope of open LHC data has already been demonstrated with the previous release of research data. A group of theorists at MIT wanted to study the substructure of jets—showers of hadron clusters recorded in the CMS detector. Since CMS had not performed this particular research, the theorists got in touch with the CMS scientists for advice on how to proceed. This blossomed into a fruitful collaboration between the theorists and CMS revolving around CMS open data. "As scientists, we should take the release of data from publicly funded research very seriously," says Salvatore Rappoccio, a CMS physicist who worked with the MIT theorists. "In addition to showing good stewardship of the funding we have received, it also provides a scientific benefit to our field as a whole. While it is a difficult and daunting task with much left to do, the release of CMS data is a giant step in the right direction."

Further, a CMS physicist in Germany tasked two undergraduates with validating the CMS Open Data by re-producing key plots from some highly cited CMS papers that used data collected in 2010. Using openly available documentation about CMS's analysis software and with some guidance from the physicist, the students were able to re-create plots that look nearly identical to those from CMS, showing what can be achieved with these data. "I was pleasantly surprised by how easy it was for the students to get started working with the CMS Open Data and how well the exercise worked," says Achim Geiser, the physicist behind this project. Simplified example code from one of these analyses is [available on the CERN Open Data Portal](#) and more is on its way.

Prior to the launch of the CERN Open Data Portal with the first batch of research-quality data from CMS, the Collaboration had provided certain curated datasets for use in high-school workshops. These "masterclasses", developed by QuarkNet and conducted under the aegis of the International Particle Physics Outreach Group, bring particle-physics data to thousands of high-school students each year. These educational datasets are also available on the CERN Open Data Portal,

along with an "[event display](#)" for visualising the particle-collision events.

"We are very pleased that we can make all these data publicly available," adds Kati. "We look forward to how they are utilised outside our collaboration, for research as well as for building educational tools."

CMS is one of two general-purpose experiments at the LHC that have been built to search for new physics. It is designed to detect a wide range of particles and phenomena produced in the LHC's high-energy proton-proton and heavy-ion collisions and will help to answer questions such as: "What is the Universe really made of and what forces act within it?" It will also measure the properties of well-known particles with unprecedented precision and be on the lookout for completely new, unpredicted phenomena. Such research not only increases our understanding of the way the Universe works, but may eventually spark new technologies that change the world in which we live as has often been true in the past.

The conceptual design of the CMS experiment dates back to 1992. The construction of the gigantic detector (15 m diameter by nearly 29 m long with a weight of 14000 tonnes) took 16 years of effort from one of the largest international scientific collaborations ever assembled: CMS currently has around 2900 scientists (including nearly 1000 graduate students) plus over 1000 engineers and technicians, from 182 institutions and research laboratories distributed in 42 countries all over the world.

More information: More information: cern.ch/cms

Provided by CERN

Citation: CERN CMS releases 300 terabytes of research data from LHC (2016, April 25)

retrieved 12 May 2024 from <https://phys.org/news/2016-04-cms-terabytes-lhc.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.