

Breaking data records bit by bit

December 14 2017



Magnetic tapes, retrieved by robotic arms, are used for long-term storage. Credit: Julian Ordan/CERN

This year CERN's data centre broke its own record, when it collected more data than ever before.

During October 2017, the data centre stored the colossal amount of 12.3

petabytes of data. To put this in context, one petabyte is equivalent to the [storage](#) capacity of around 15,000 64GB smartphones. Most of this data come from the Large Hadron Collider's experiments, so this record is a direct result of the outstanding LHC performance, the rest is made up of data from other experiments and backups.

"For the last ten years, the data volume stored on tape at CERN has been growing at an almost exponential rate. By the end of June we had already passed a data storage milestone, with a total of 200 petabytes of data permanently archived on tape," explains German Cancio, who leads the tape, archive & backups storage section in CERN's IT department.

The CERN data centre is at the heart of the Organization's infrastructure. Here data from every experiment at CERN is collected, the first stage in reconstructing that data is performed, and copies of all the experiments' data are archived to long-term tape storage.

Most of the data collected at CERN will be stored forever, the physics data is so valuable that it will never be deleted and needs to be preserved for future generations of physicists.

"An important characteristic of the CERN data archive is its longevity," Cancio adds. "Even after an experiment ends all recorded data has to remain available for at least 20 years, but usually longer. Some of the archive files produced by previous CERN experiments have been migrated across different hardware, software and media generations for over 30 years. For archives like CERN's, that do not only preserve existing data but also continue to grow, our data preservation is particularly challenging."

While tapes may sound like an outdated mode of storage, they are actually the most reliable and cost-effective technology for large-scale archiving of data, and have always been used in this field. One copy of

data on a tape is considered much more reliable than the same copy on a disk.

CERN currently manages the largest scientific data [archive](#) in the High Energy Physics (HEP) domain and keeps innovating in [data](#) storage," concludes Cancio.

Provided by CERN

Citation: Breaking data records bit by bit (2017, December 14) retrieved 26 April 2024 from <https://phys.org/news/2017-12-bit.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.