

The world's largest archive for climate simulation data

DKRZ, supported by IBM, took in operation its new HPSS-based hierarchical storage management (HSM) system. The system manages all archived simulation data at DKRZ and serves as the central input and output interface of the tape library.

The data archive at DKRZ

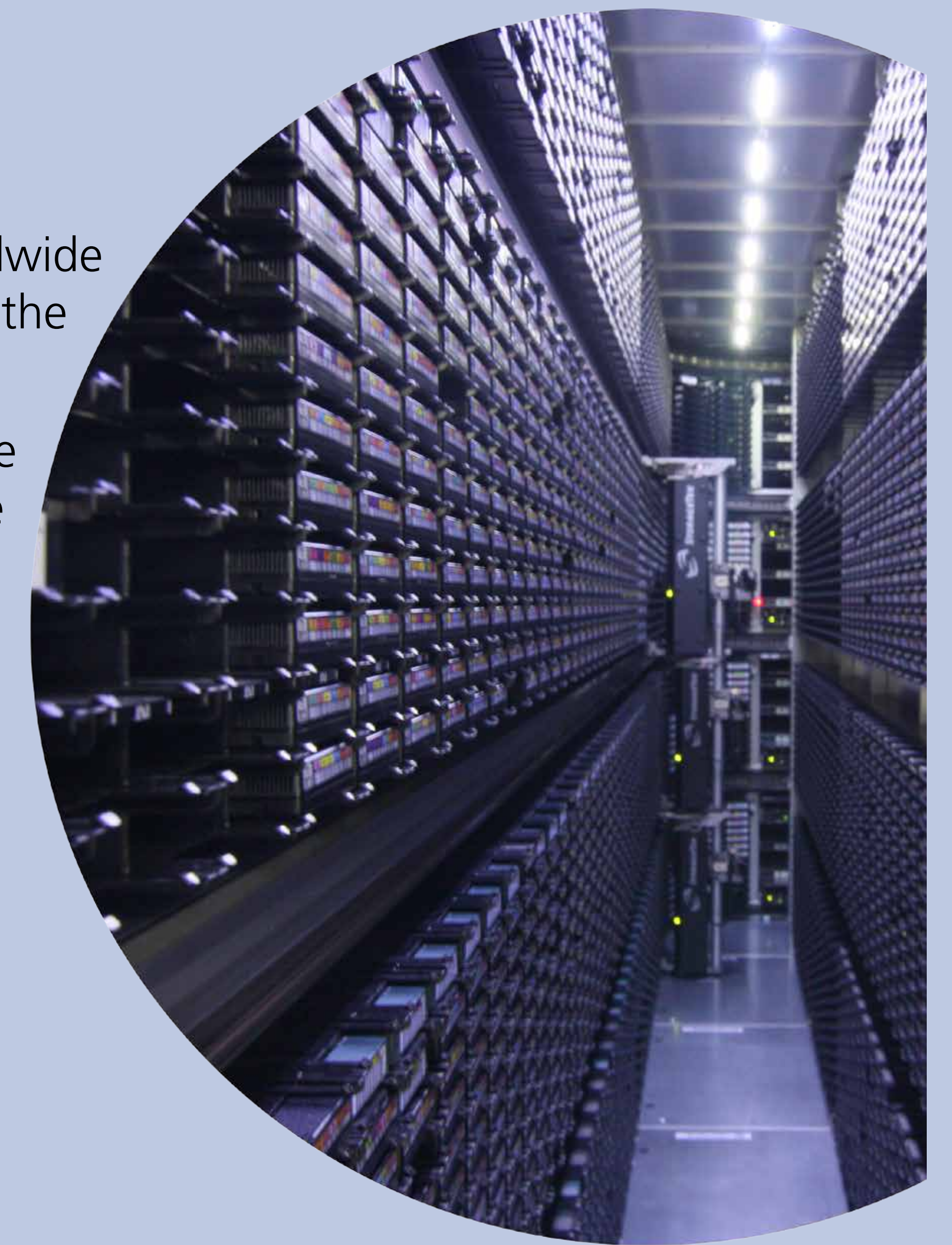
Climate models produce extremely large amounts of data. DKRZ is one of the few computing centers worldwide that has the technical facilities to store such data volumes and the scientific knowledge to manage them. At the moment the tape library at DKRZ contains about 20 million files with a total data volume of 40 petabytes.

In order to adapt DKRZ's archive to the expected data production rate of the new HPC system Mistral, the archive system was upgraded by IBM with the latest HPSS software, IBM x86 servers as well as a new five petabyte disk cache, which serves as a temporary buffer. This size corresponds to the storage capacity of 5,000 well equipped notebook computers.

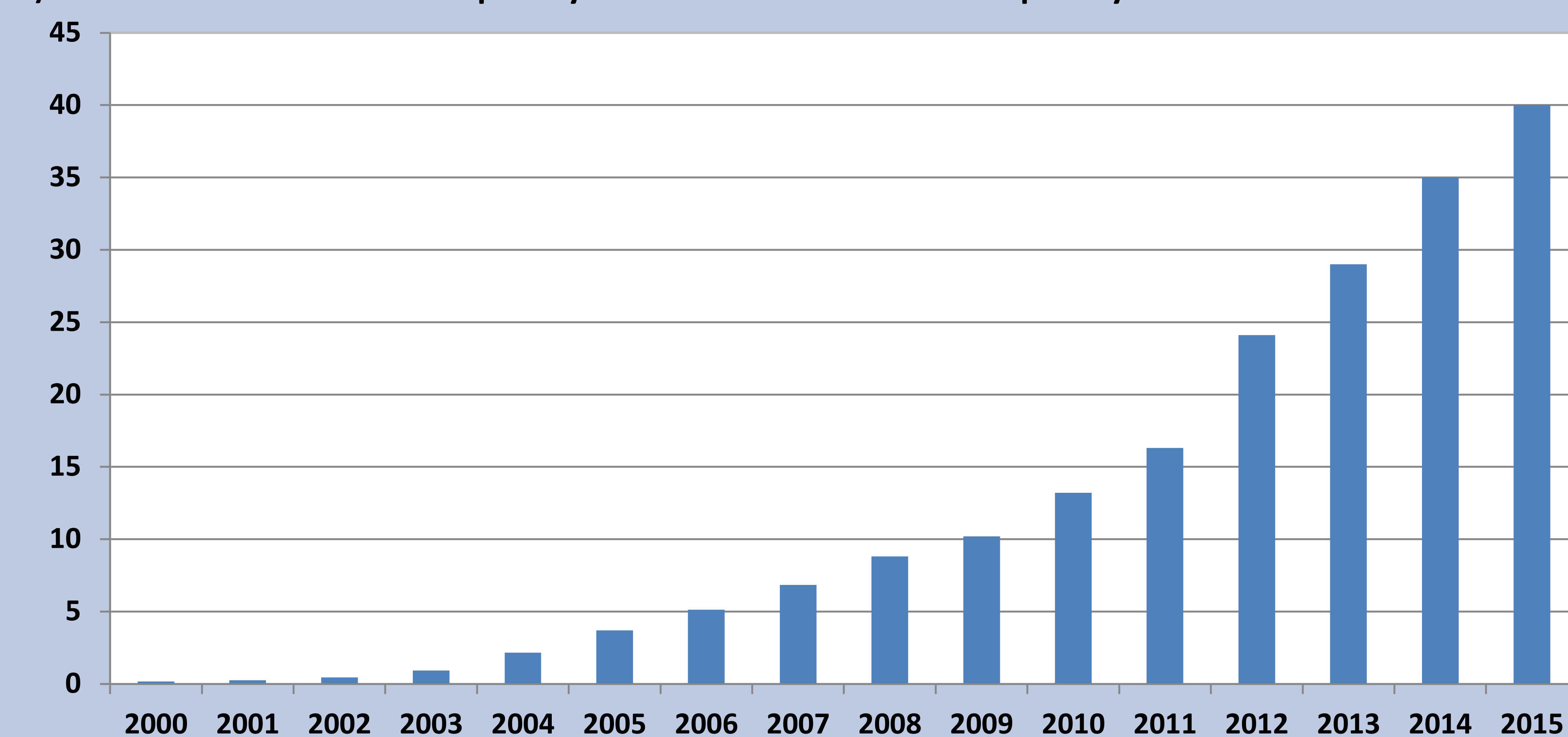


The DKRZ HSM system and tape library at a glance:

- 5 petabyte HPSS disk cache
- 8 automatic Sun StorageTek SL8500 tape libraries
- 77.000 tape cartridges
- 8 robots per library
- 75 tape drives
- total capacity of up to 500 petabytes
- projected data growth of 75 petabytes per year
- bidirectional bandwidth up to 18 gigabytes per second



The amount of data in the DKRZ archive skyrockets from 1 terabyte in the year 1993 to 1 petabyte in 2003 and more than 40 petabytes in 2015



Data life cycle services at DKRZ

For enabling sustainable use of the large amount of climate model data produced, DKRZ operates data lifecycle services. These services support the whole data life cycle: A seamless end-to-end workflow from data production over data storage, data processing, data dissemination to data storage is supported to optimize the data management in climate science. With the ICSU World Data Center Climate (WDCC), DKRZ runs a fully documented long-term data archive with a size of currently 4 petabytes.

