Big Data@FNAL

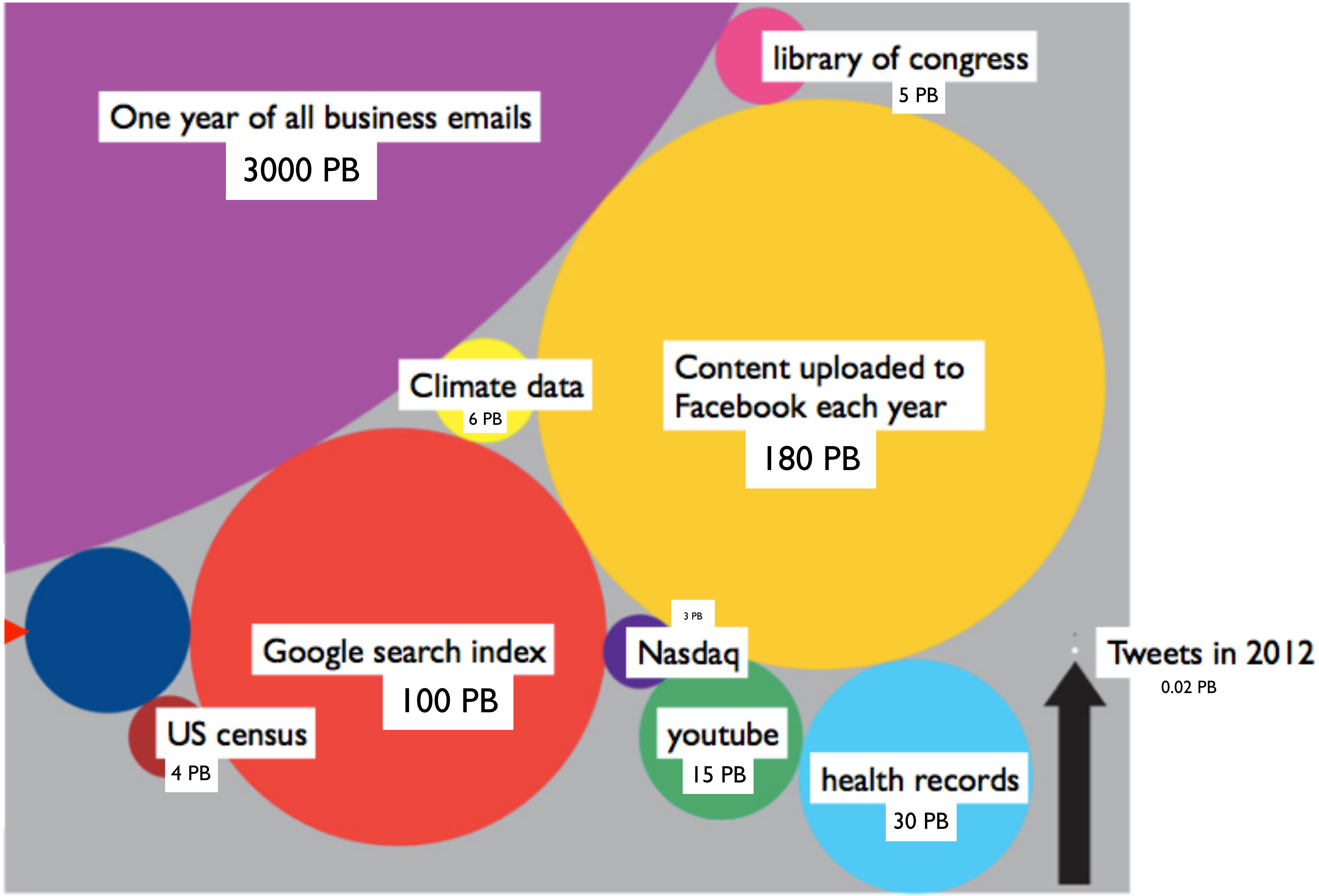
Oliver Gutsche

8th INFIERI Workshop

20. October 2016

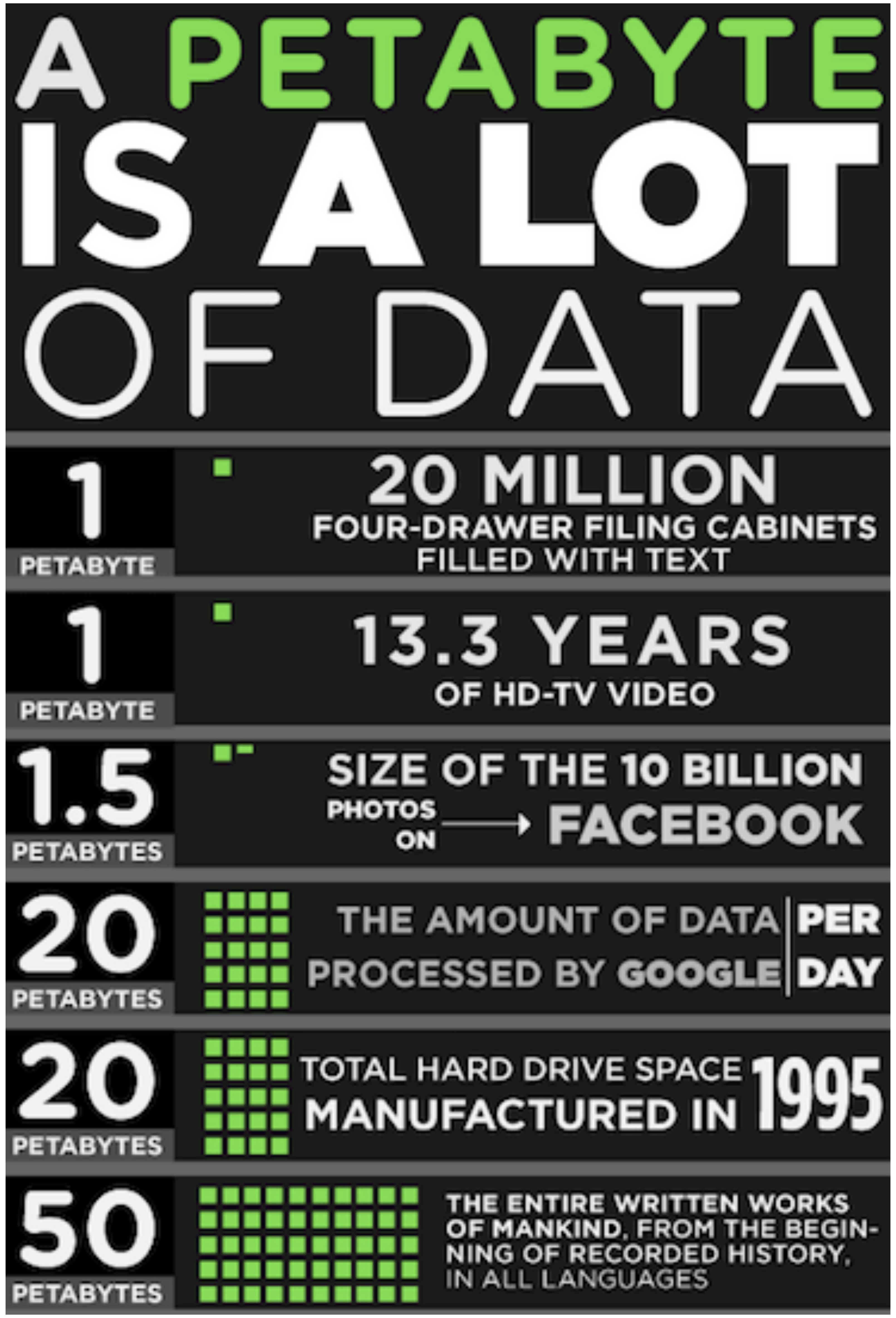
Big Data

What is Big Data?



Adapted from Wired: http://www.wired.com/magazine/2013/04/bigdata/

How Big is Data?

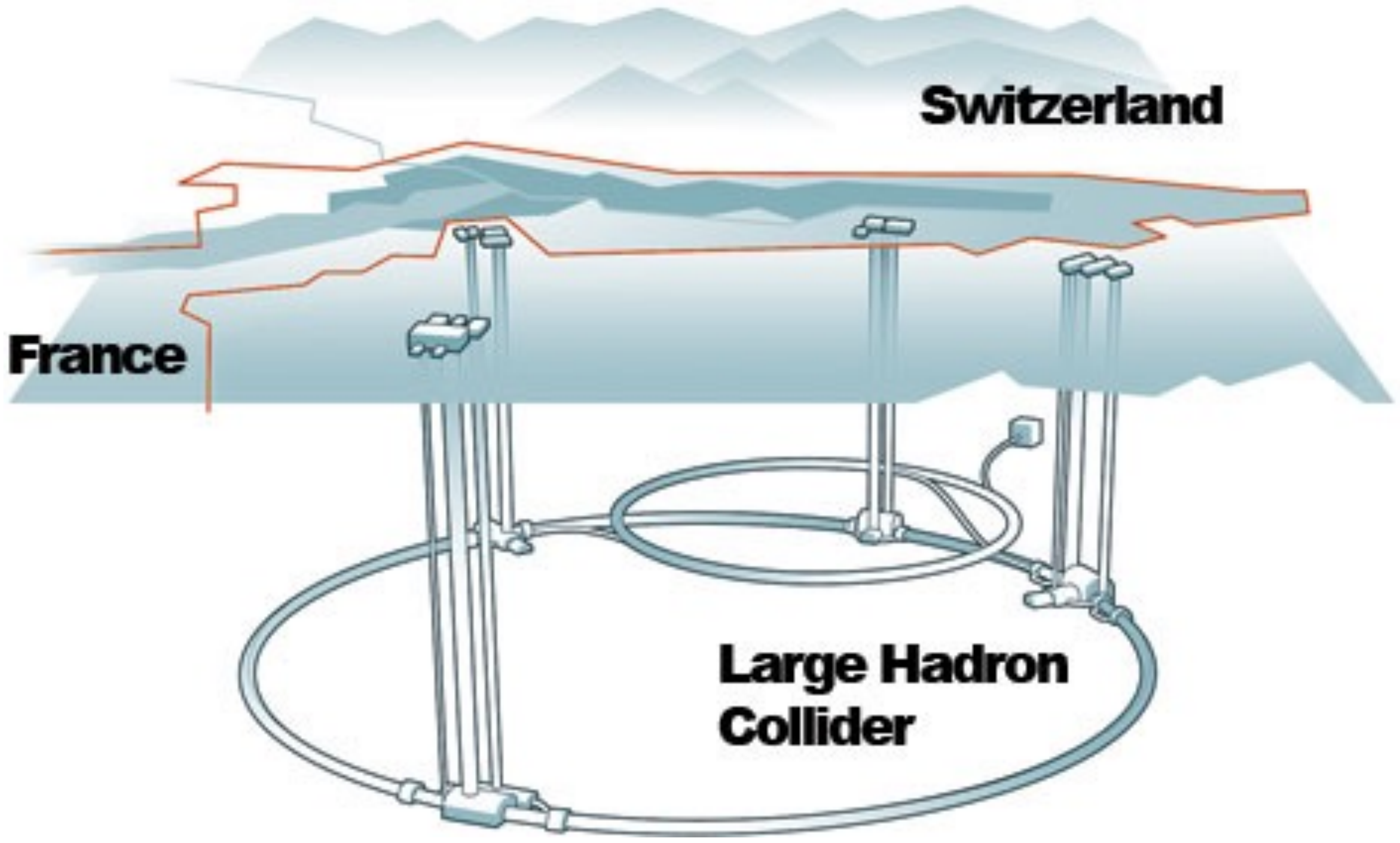


Who is producing Big Data?

Science

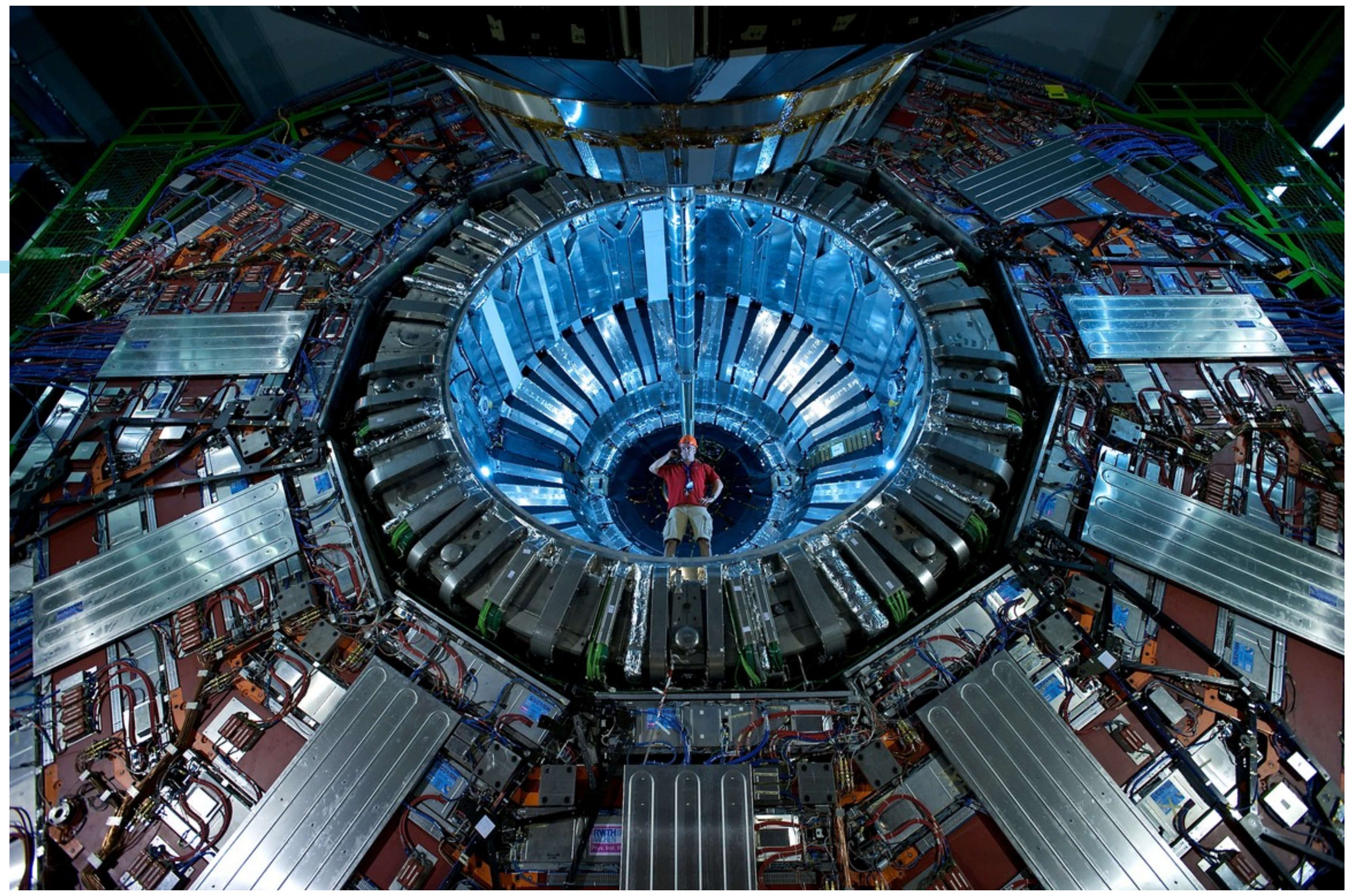
Large Hadron Collider (LHC)

- Circumference: almost 17 Miles
- 2 proton beams circulating at 99.9999991% of the speed of light
- A particle beam consists of bunches of protons (100 Billion protons per bunch)
- Beams cross and are brought to collision at 4 points
  - 20 Million collisions per second per crossing point
- Energy stored in one LHC beam is equivalent to a 40t truck crashing into a concrete wall at 90 Mph

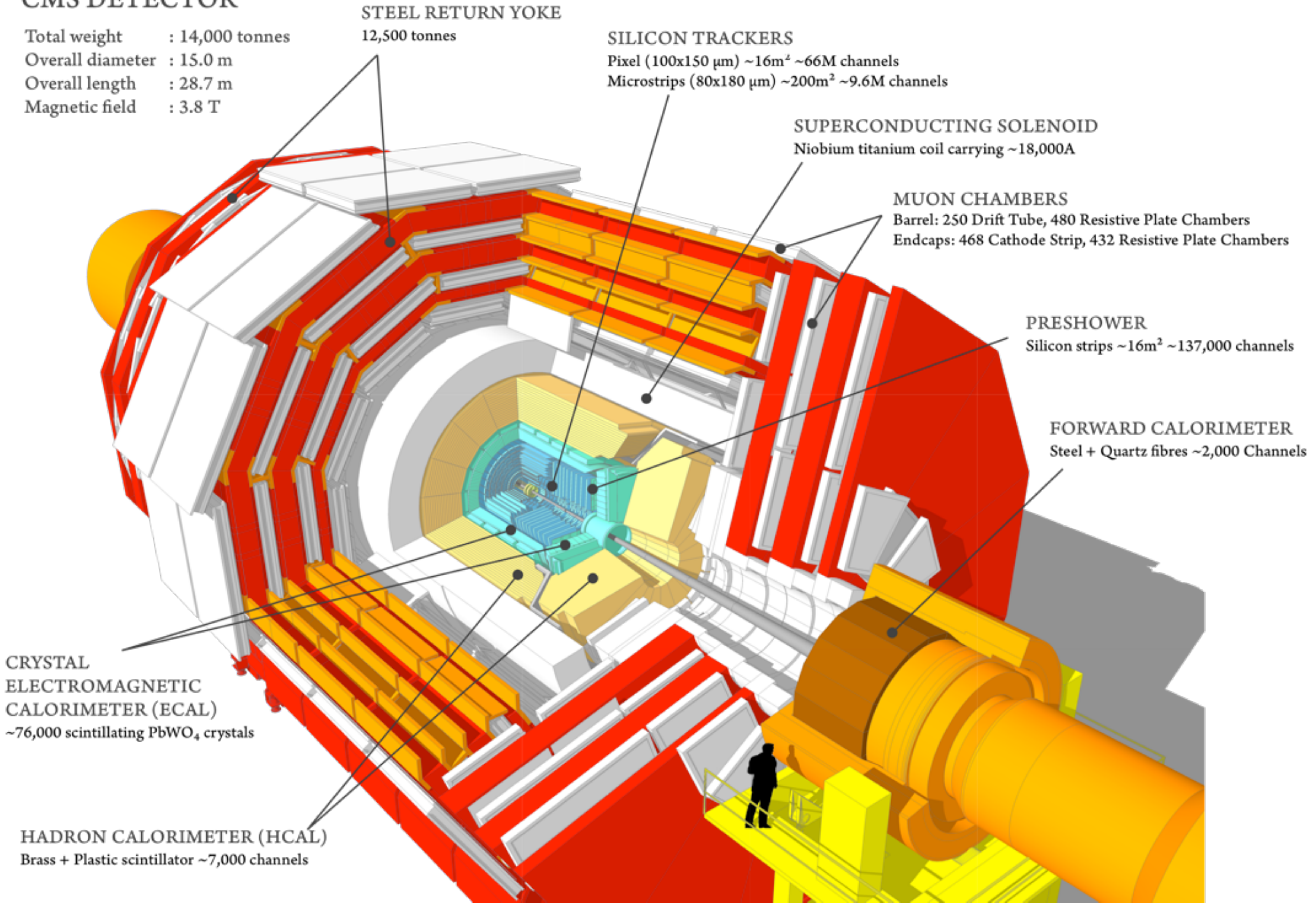


LHC guide: http://cds.cern.ch/record/1165534/files/CERN-Brochure-2009-003-Eng.pdf

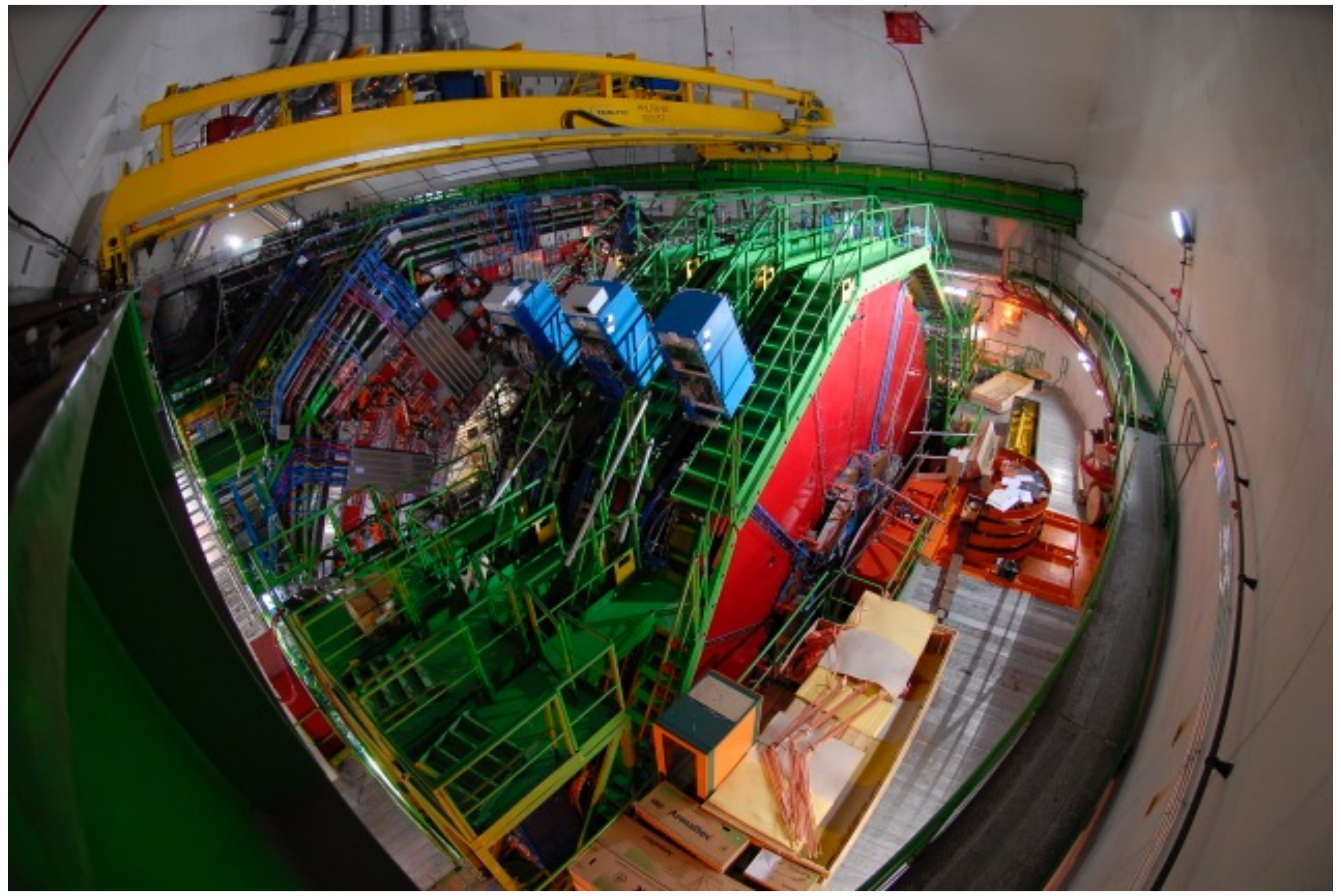
Compact Muon Solenoid (CMS)

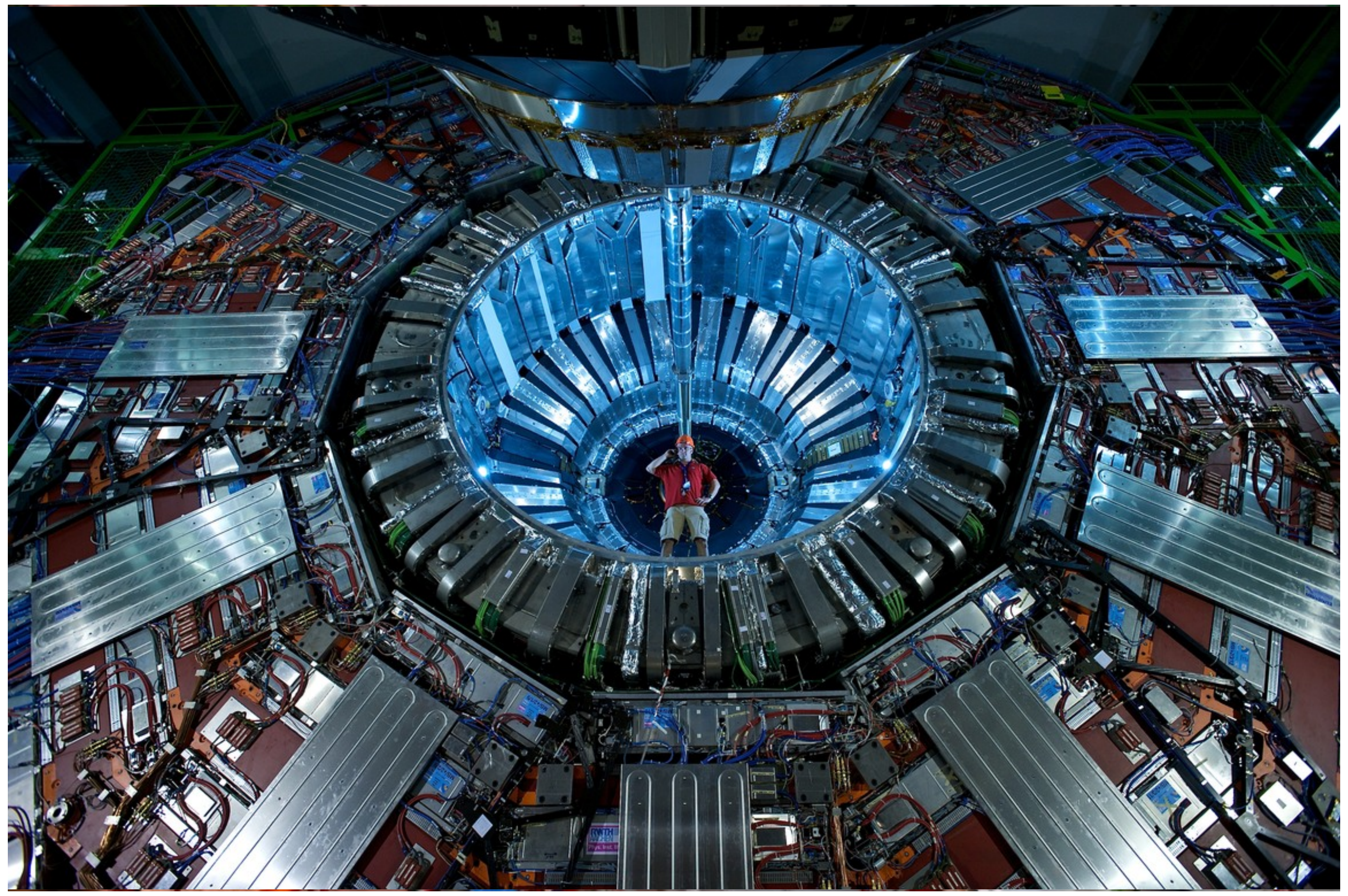


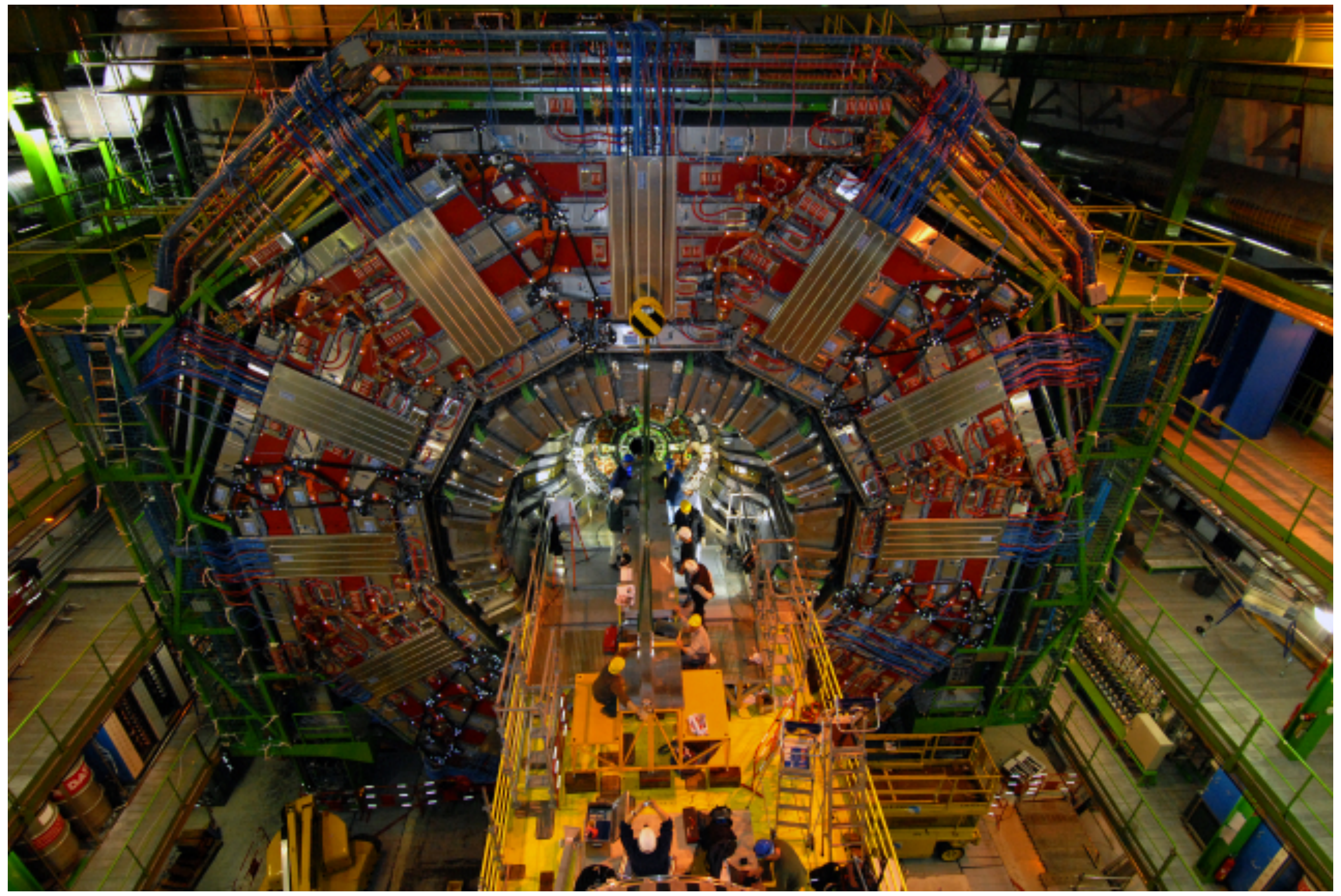
- Detector built around collision point
- Records flight path and energy of all particles produced in a collision
- 100 Million individual measurements (channels)
- All measurements of a collision together are called: **event**



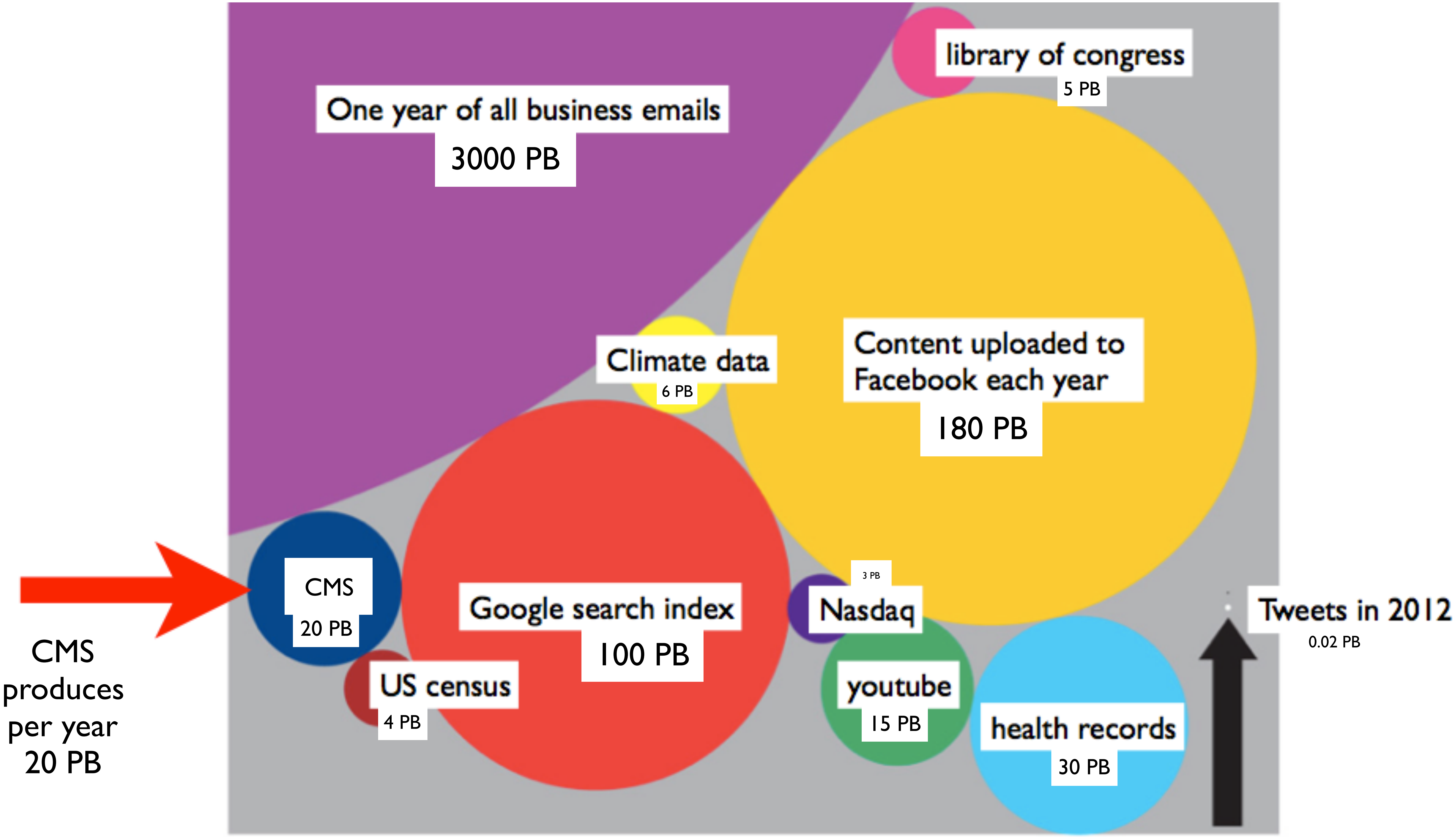
Compact Muon Solenoid (CMS)





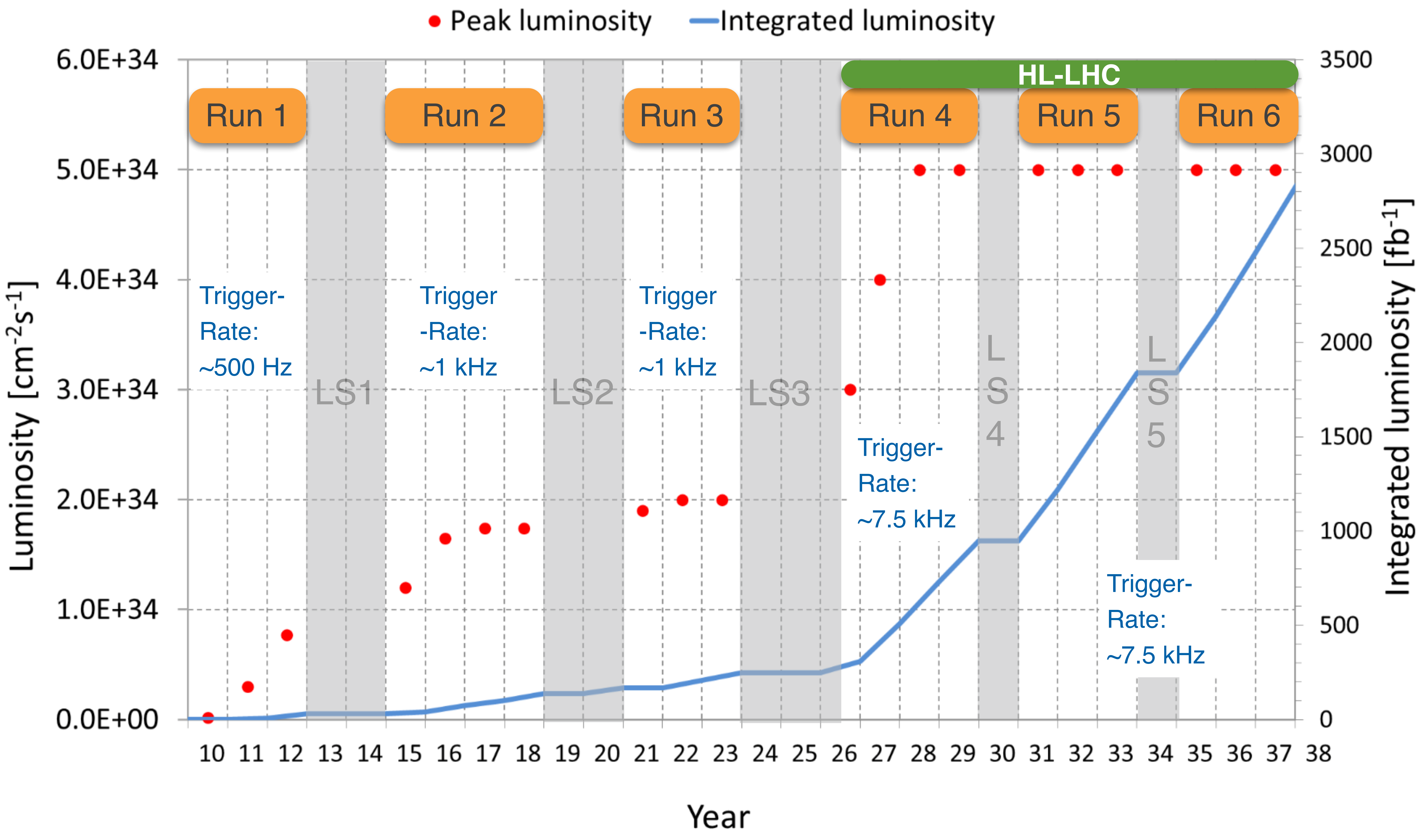


CMS is producing a lot of data

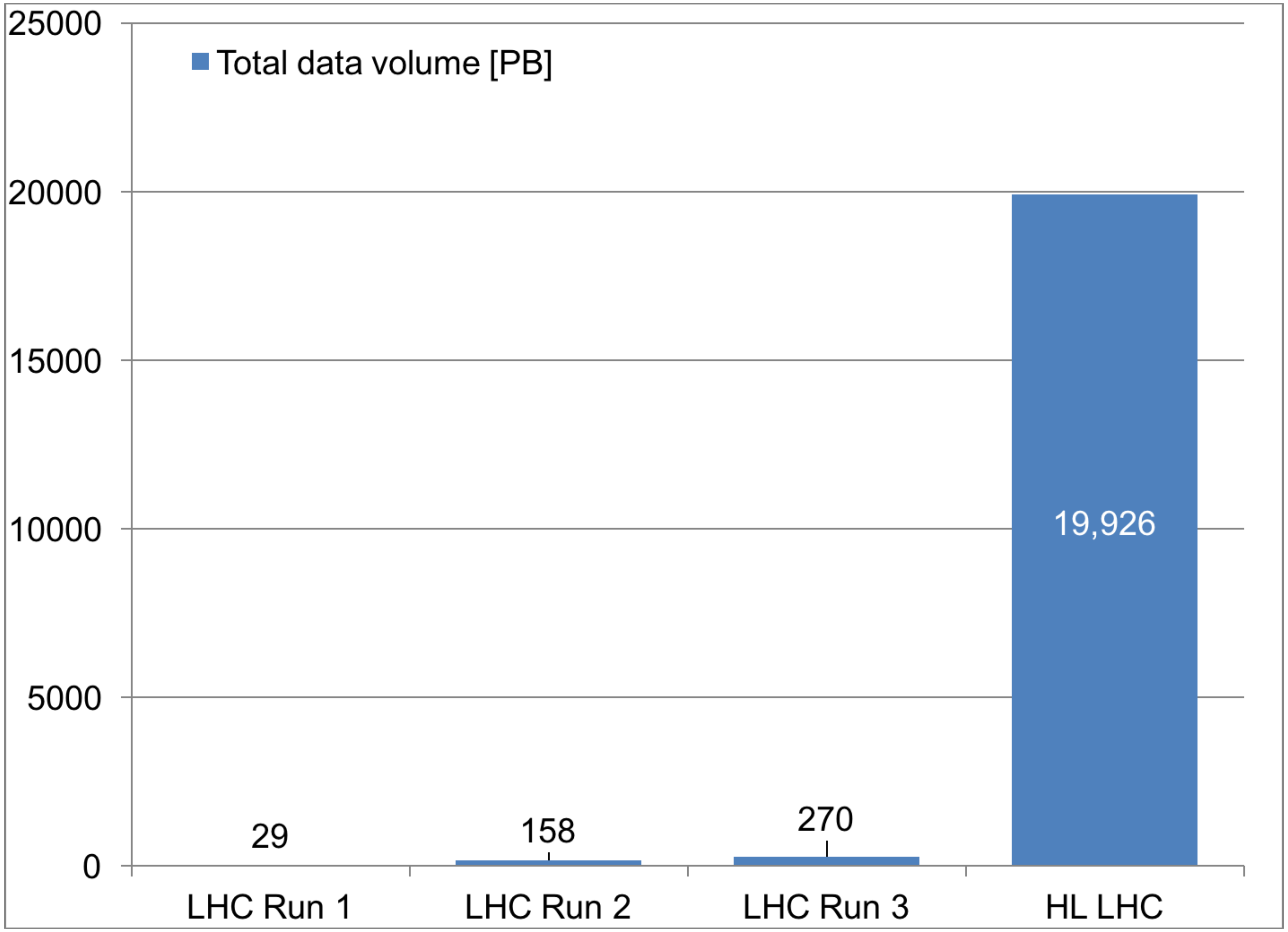


CMS produces per year 20 PB

The Future - HL-LHC



LHC expectation data volumes



2010 - 2012

2015 - 2018

2021 - 2023

2026 - 2037

EXABYTES!

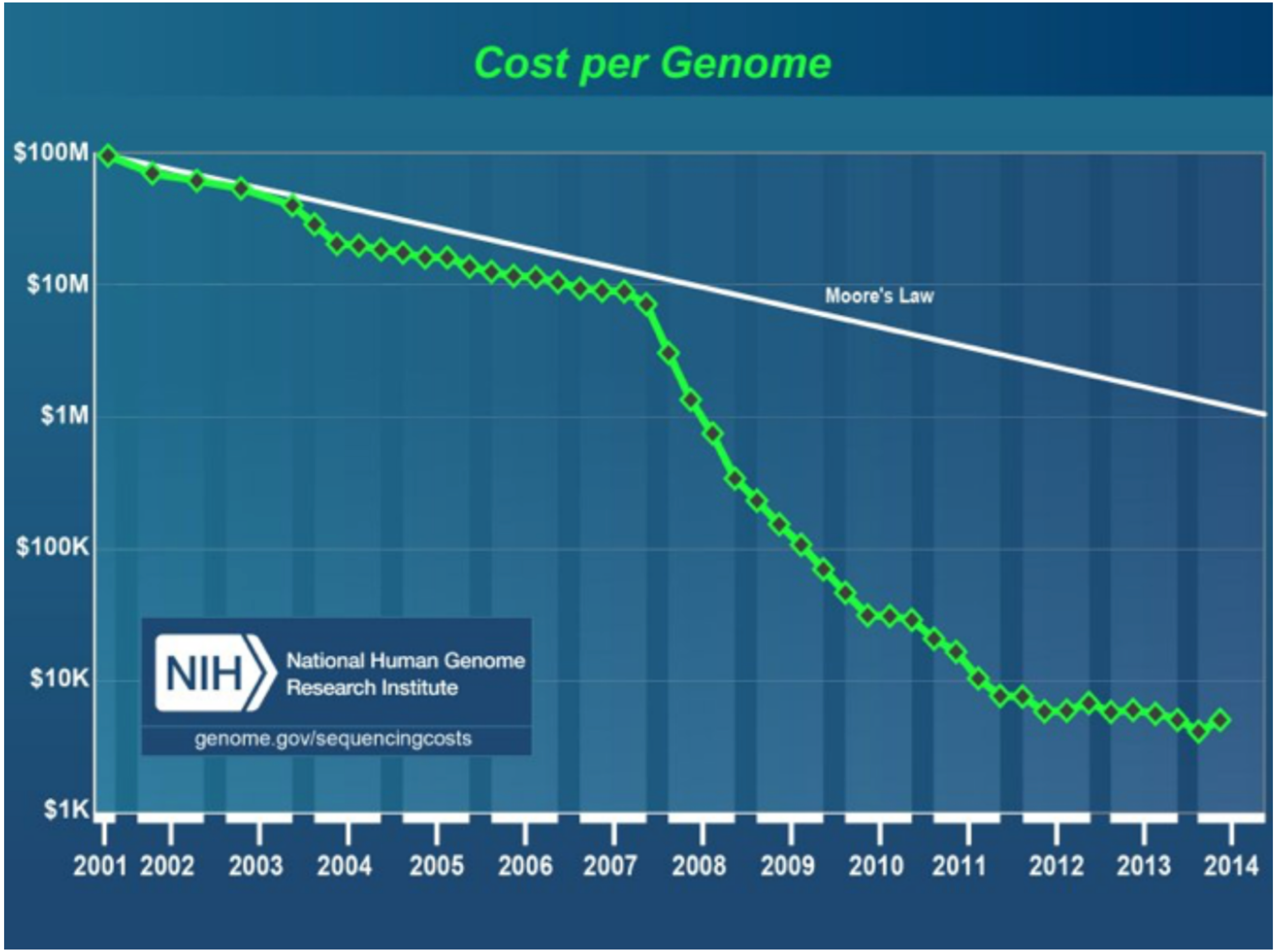
We are not alone!

- Many examples of other science disciplines producing and analyzing vast amounts of data

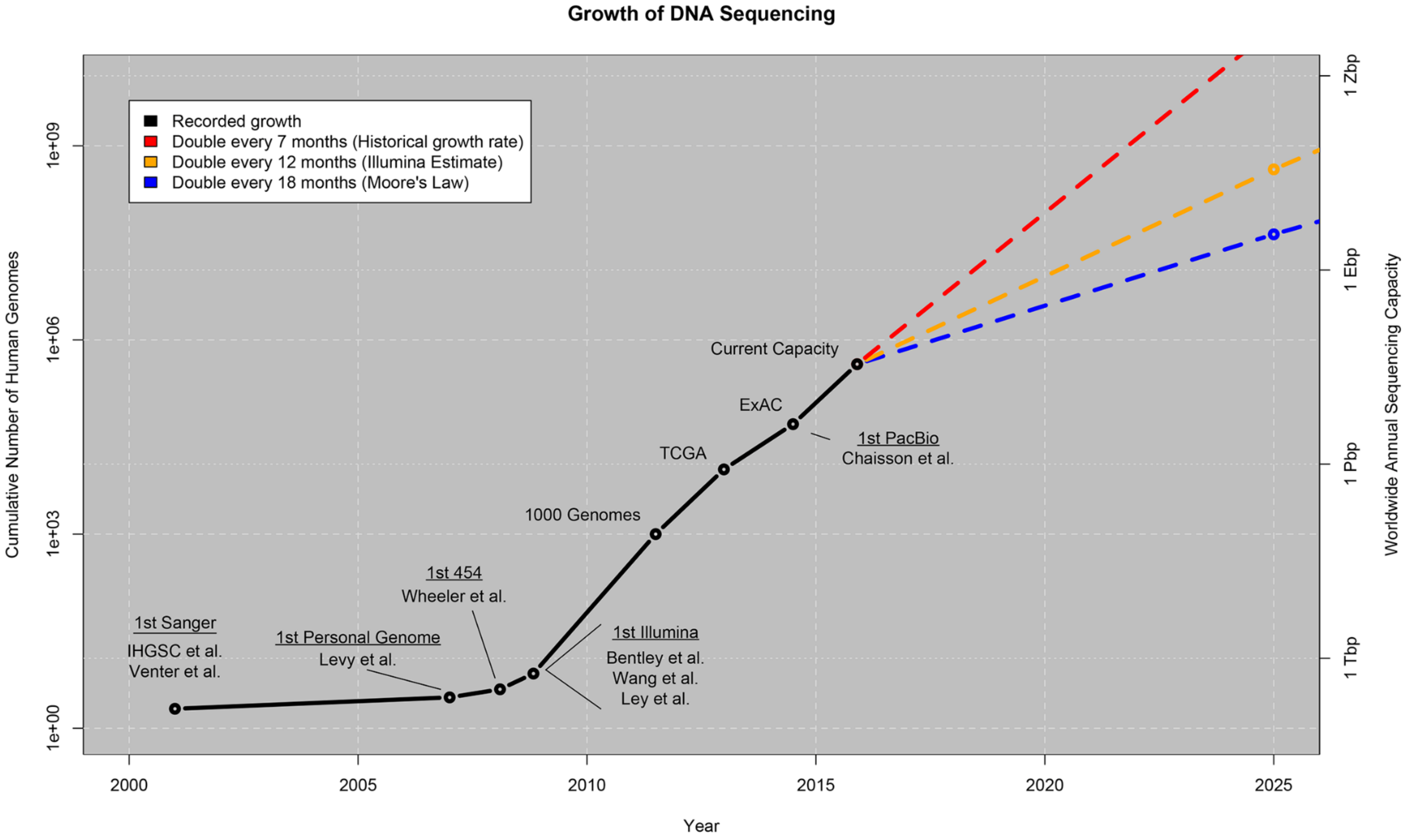
- Example: Genomic Research
- Genetic sequencing cost has decreased exponentially

- A study of a group might be a few hundred individuals
  - ~10GB file per person for whole exome (A study of about 1% of your DNA) Mutations here can have severe impact on the rest
  - ~200GB file per person for whole genome sequencing. Modern machines can sequence the entire genome

- Raw data in the few TB range for exome and few hundred TB for full genome to a few PB.



Genome research outlook



DOI:10.1371/journal.pbio.1002195

Comparison

Table 1. Four domains of Big Data in 2025. In each of the four domains, the projected annual storage and computing needs are presented across the data lifecycle.

Data Phase	Astronomy	Twitter	YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

doi:10.1371/journal.pbio.1002195.t001

How do you storage and serve big data?

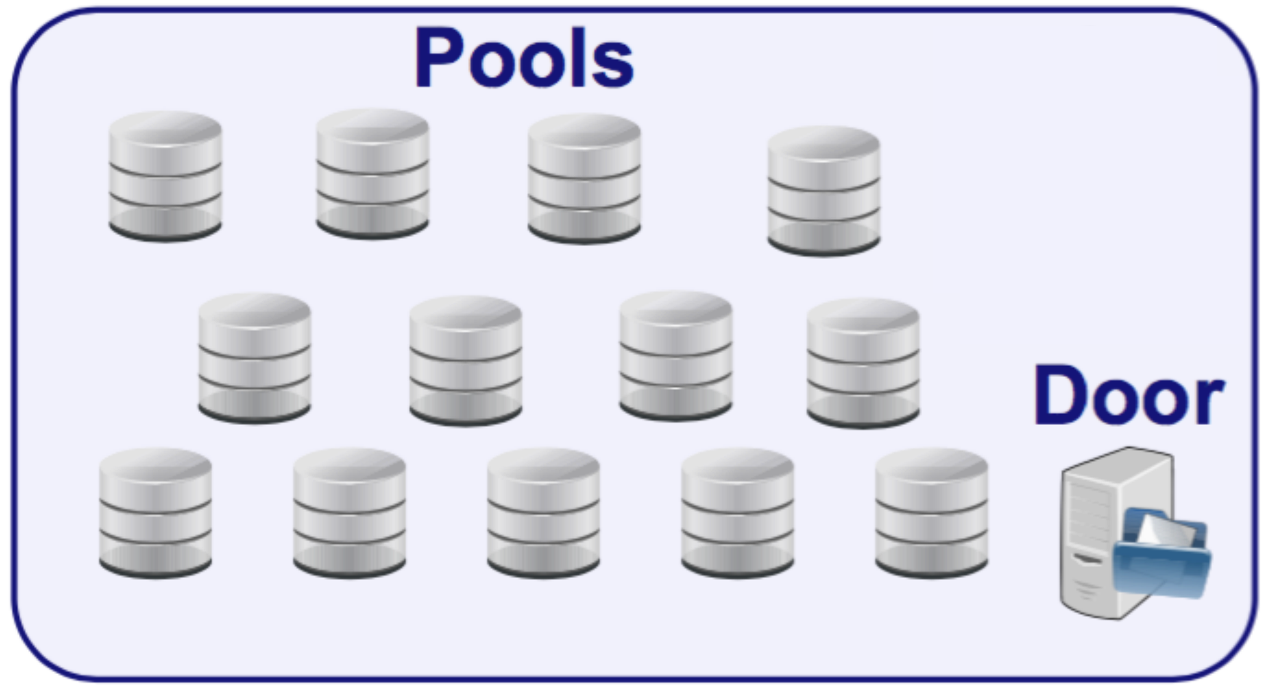
Technology

Disk

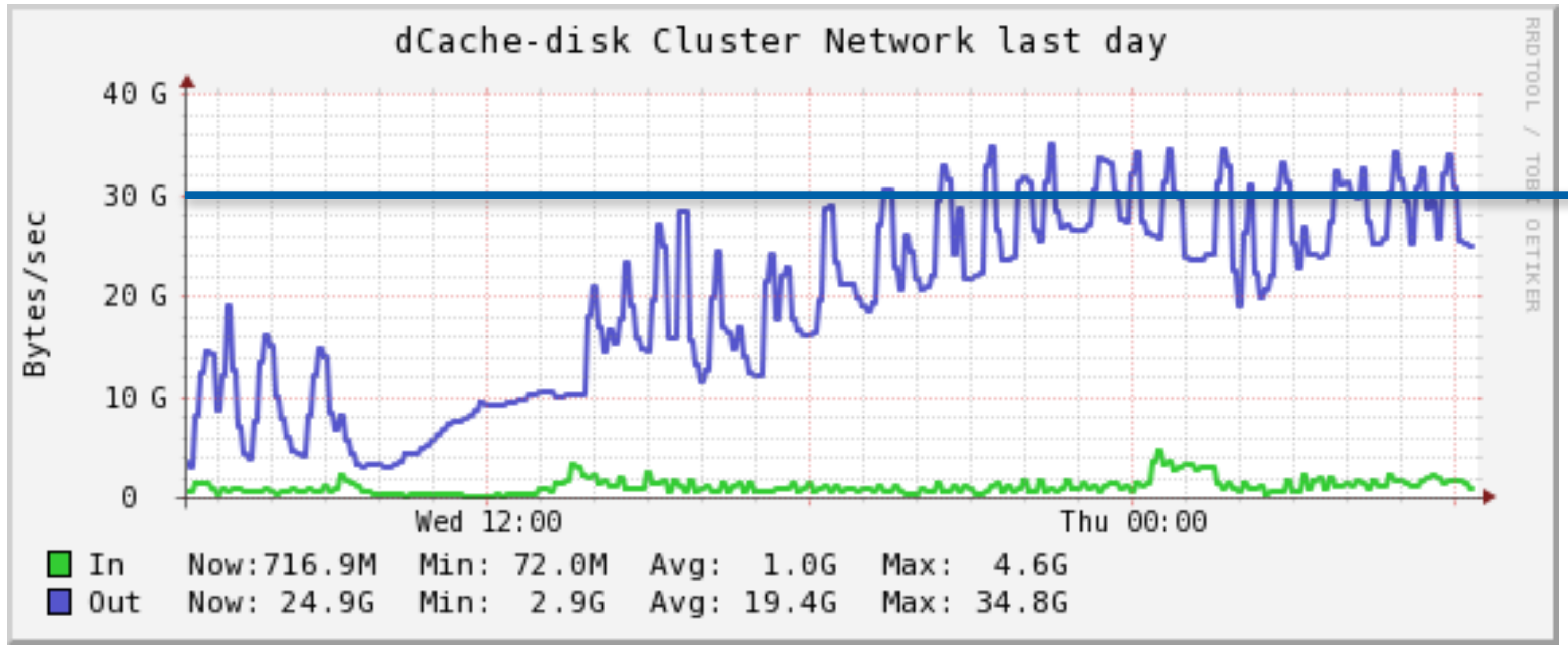
Harddrive

Diskpool

Storage System

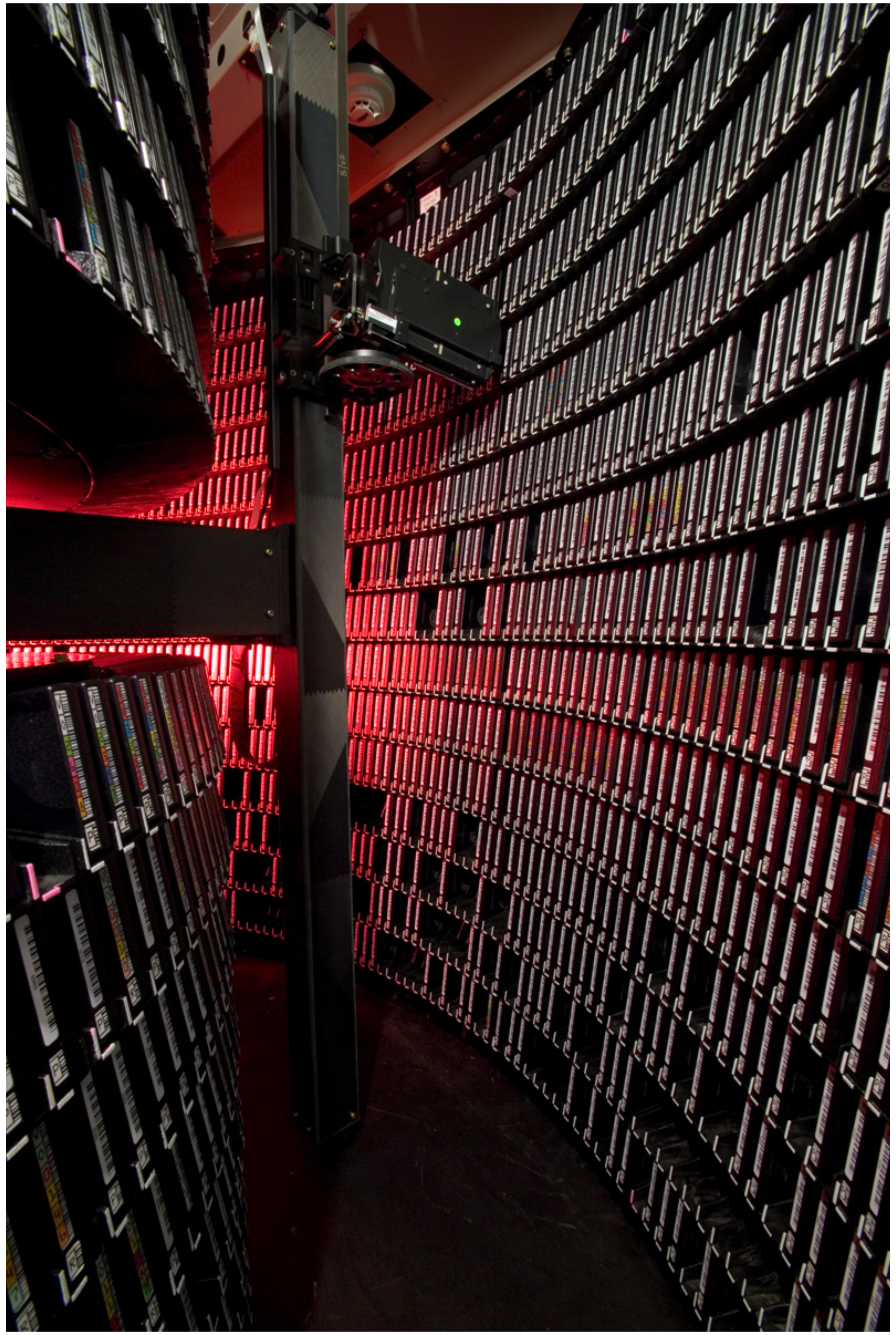


- Access to a lot of data from a lot of CPUs is provided through disk systems
  - Harddrive ➜ Diskpool ➜ Storage Systems
- We operate massive installations of many Petabytes that look like a single hard drive to the user
  - Storage systems handle access, provide replication
  - Access is not like you access as file on your harddrive (POSIX) but through special protocols (scaling reasons)
  - Using community systems like dCache, EOS but also industry systems like HDFS from Apache Hadoop



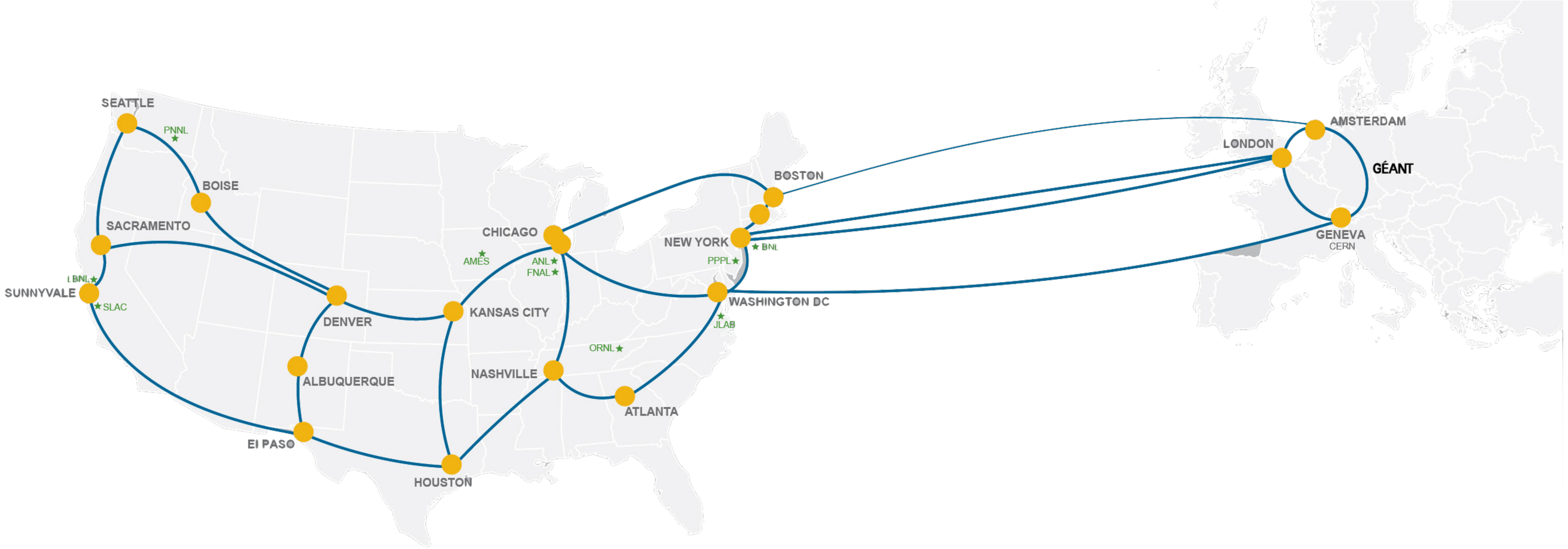
FNAL dCache 30 GB/s (240 Gbps)

Tape



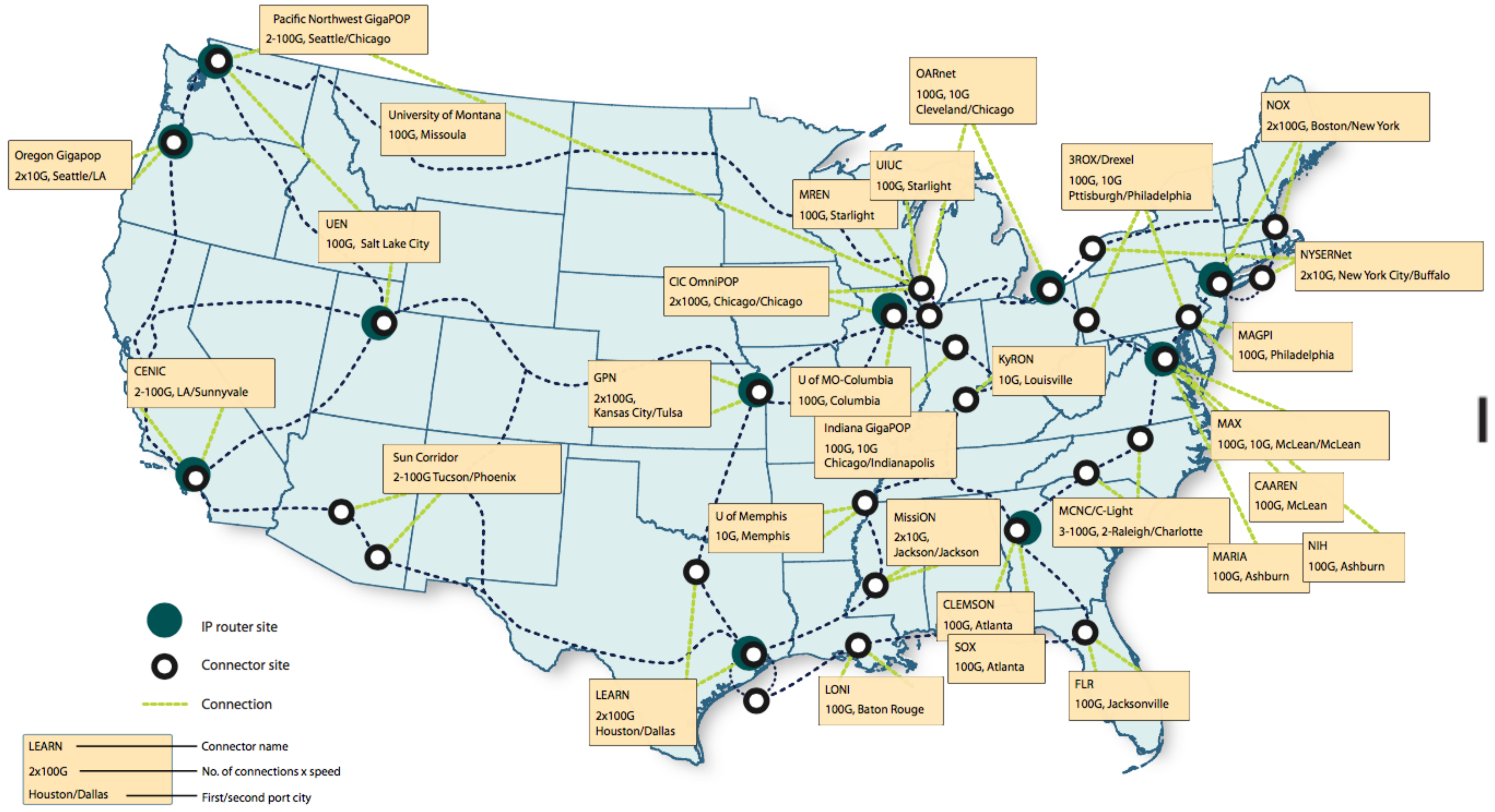
- Longterm storage is handled by magnetic tape
  - Still the cheapest way of storing Petabytes of data ➜ if you don't need to access it
  - For access, data has to be copied ("staged") to disk
- Individual cartridges can currently store up to 10 TB
- Individual robots have 10k - 15k cartridge slots

Strong networks: ESNet



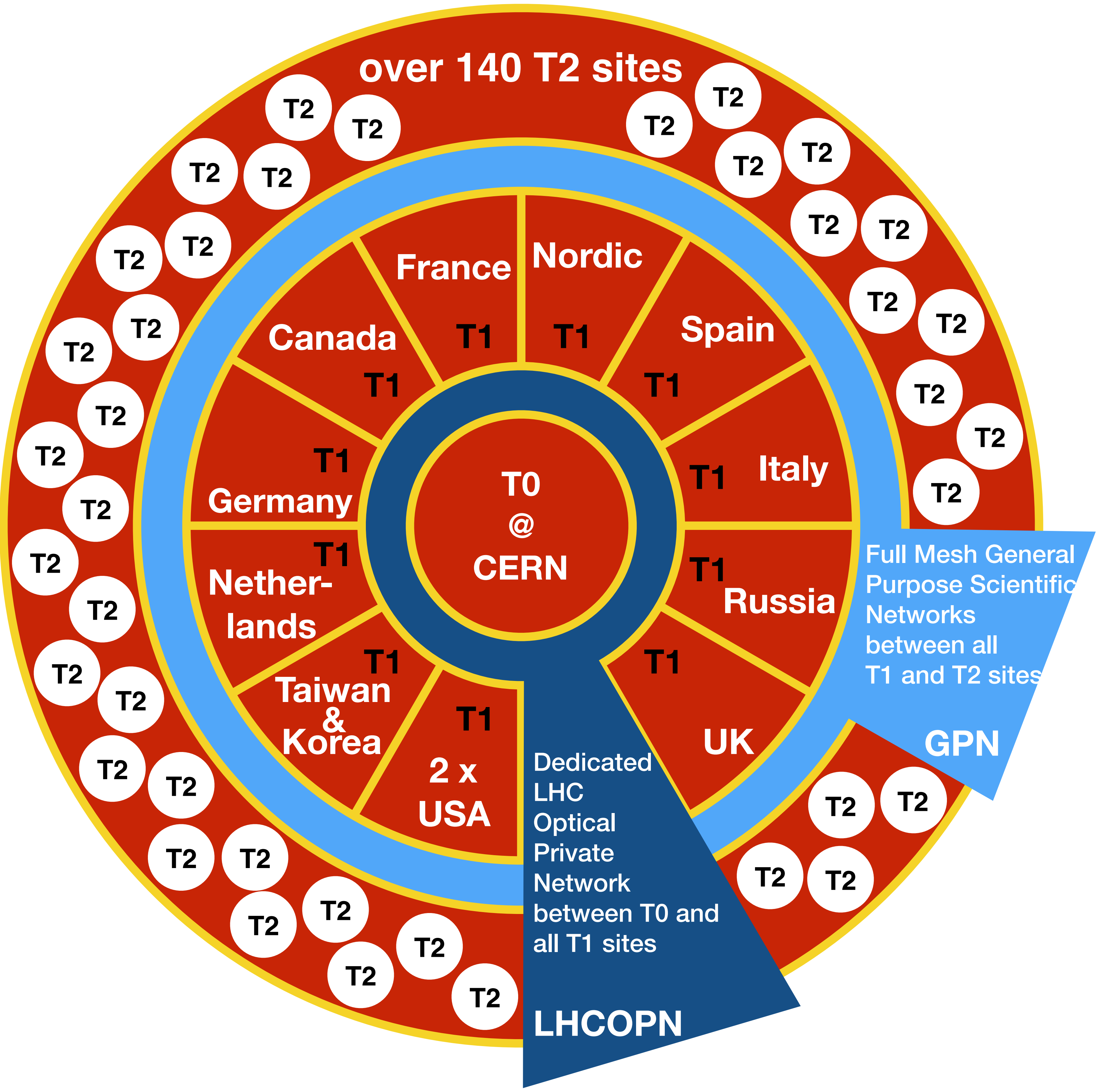
The Office of Science supports:

- 27,000 Ph.D.s, graduate students, undergraduates, engineers, and technicians
- 26,000 users of open-access facilities
- 300 leading academic institutions
- 17 DOE laboratories



Distributed infrastructures and transfer systems

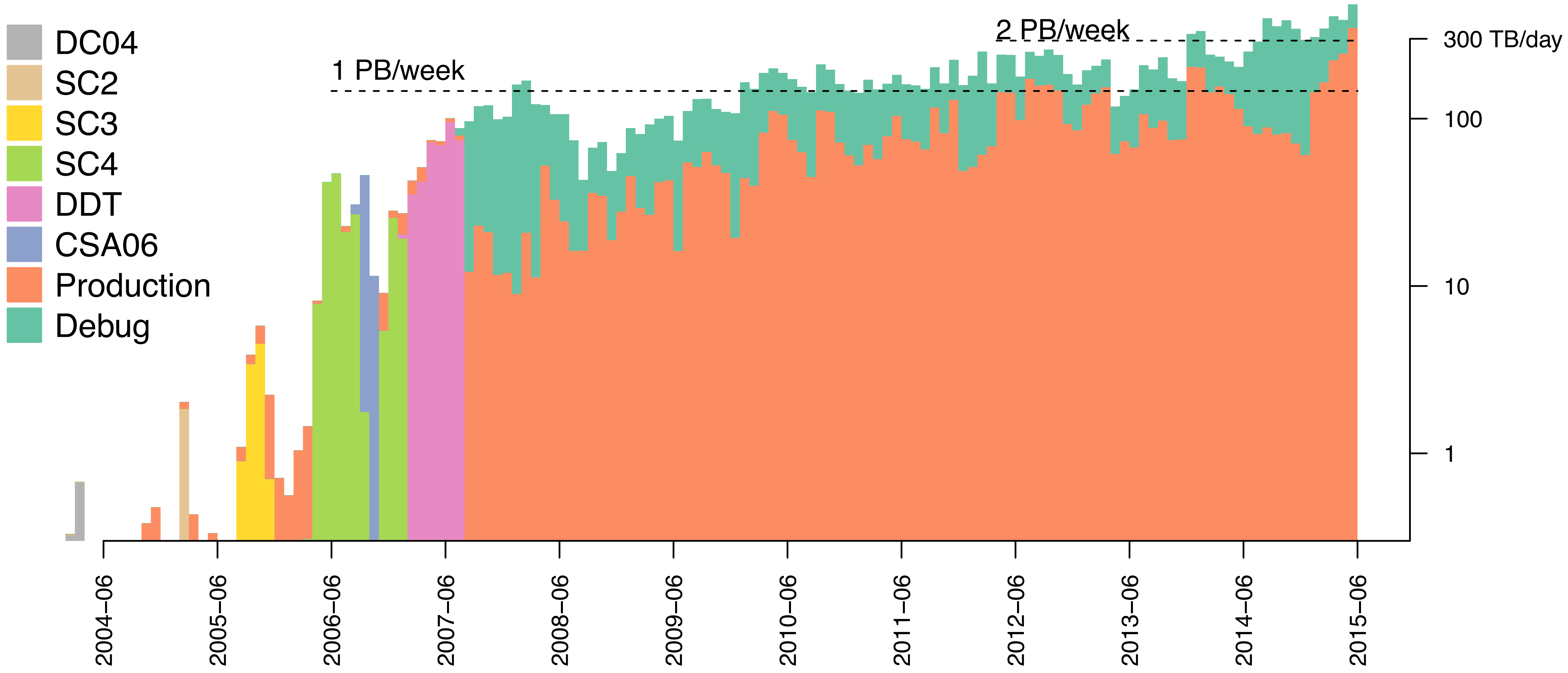
Example: Worldwide LHC Grid (WLCG)



Community uses various solutions to provide distributed access to data:

Experiment specific: Atlas (Rucio), CMS (PhEDEx), …

Shared: SAM (Neutrino and Muon experiments)

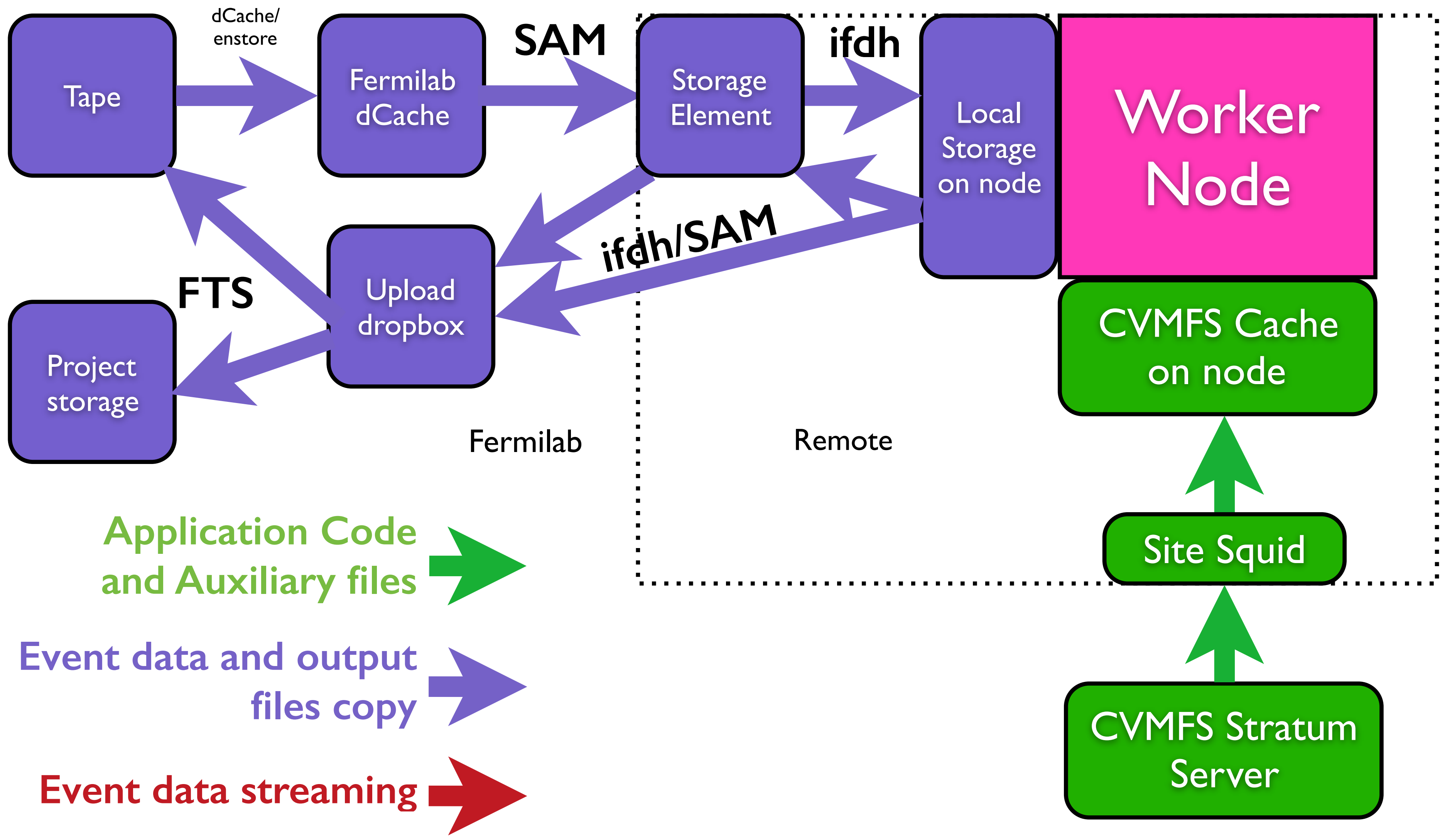


CMS transfers: more than 2 PB per week

Fermilab

Dynamic Data Management

- Subscription based transfer systems
  - PhEDEx (CMS) and Rucio (Atlas)
  - LHC Run 1: mostly manual operations
  - LHC Run 2: **dynamic data management**
    - Popularity is tracked per dataset
    - Replica count across sites is increased or decreased according to popularity



- Fully integrated distribution system
  - SAM (shared amongst Neutrino and Muon experiments)
  - All movement is based on requests for datasets from jobs.
  - Interfaces to storage at sites, performs cache-to-cache copies if necessary

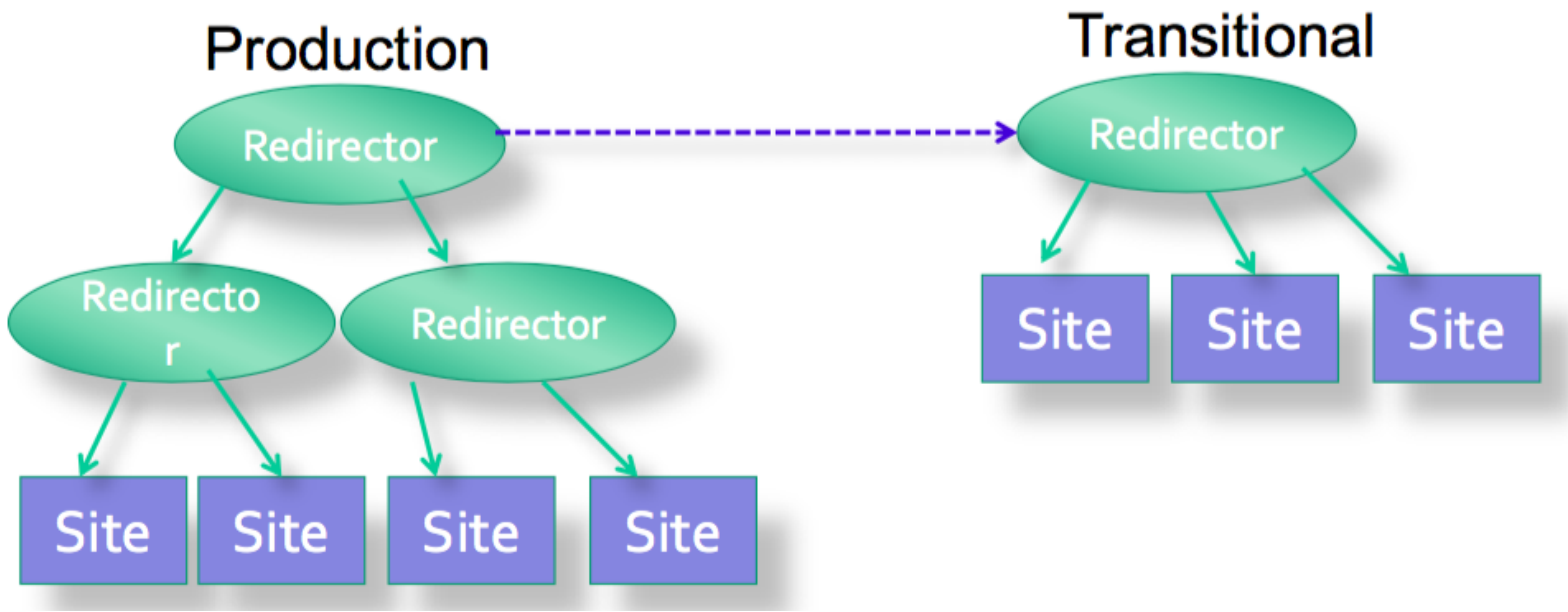
- Data is distributed automatically for the community

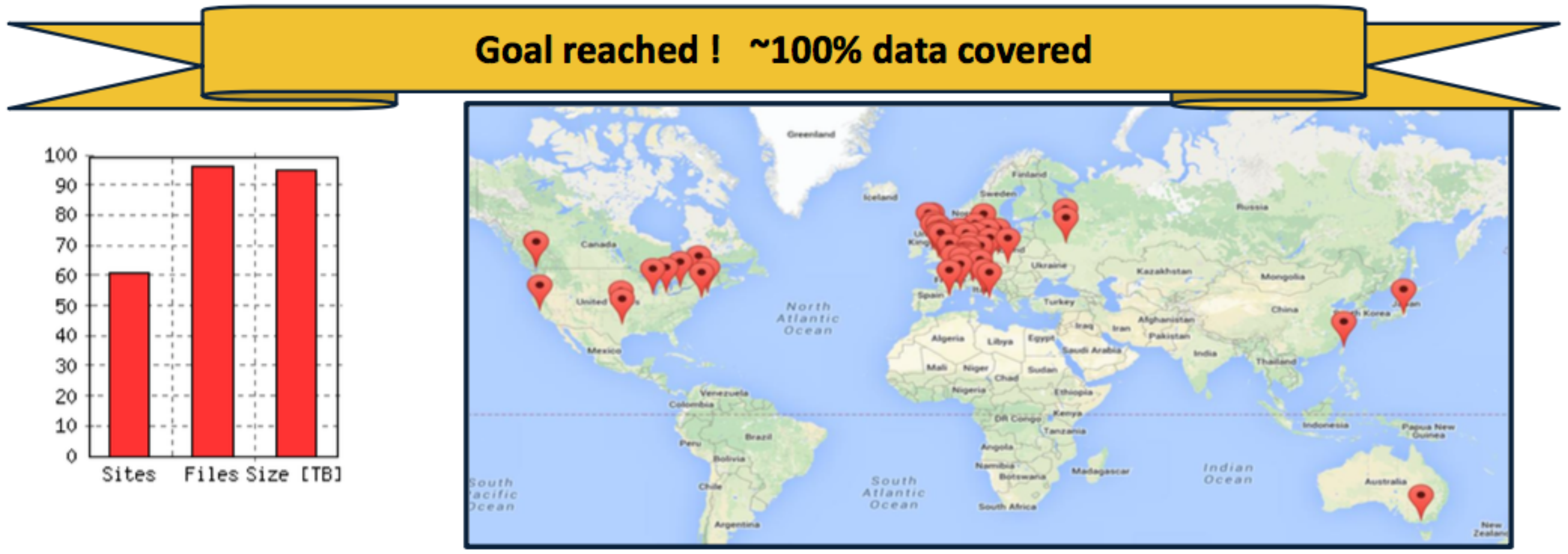
Data Federations

- xrootd: remote access to files

- ALICE based on xrootd from the beginning

- CMS and Atlas deployed xrootd federations
  - AAA for CMS, FAX for Atlas
  - Allows for remote access to all files on disk at all sites
  - Use cases:
    - Fall back
    - Overflow for ~10% of all jobs





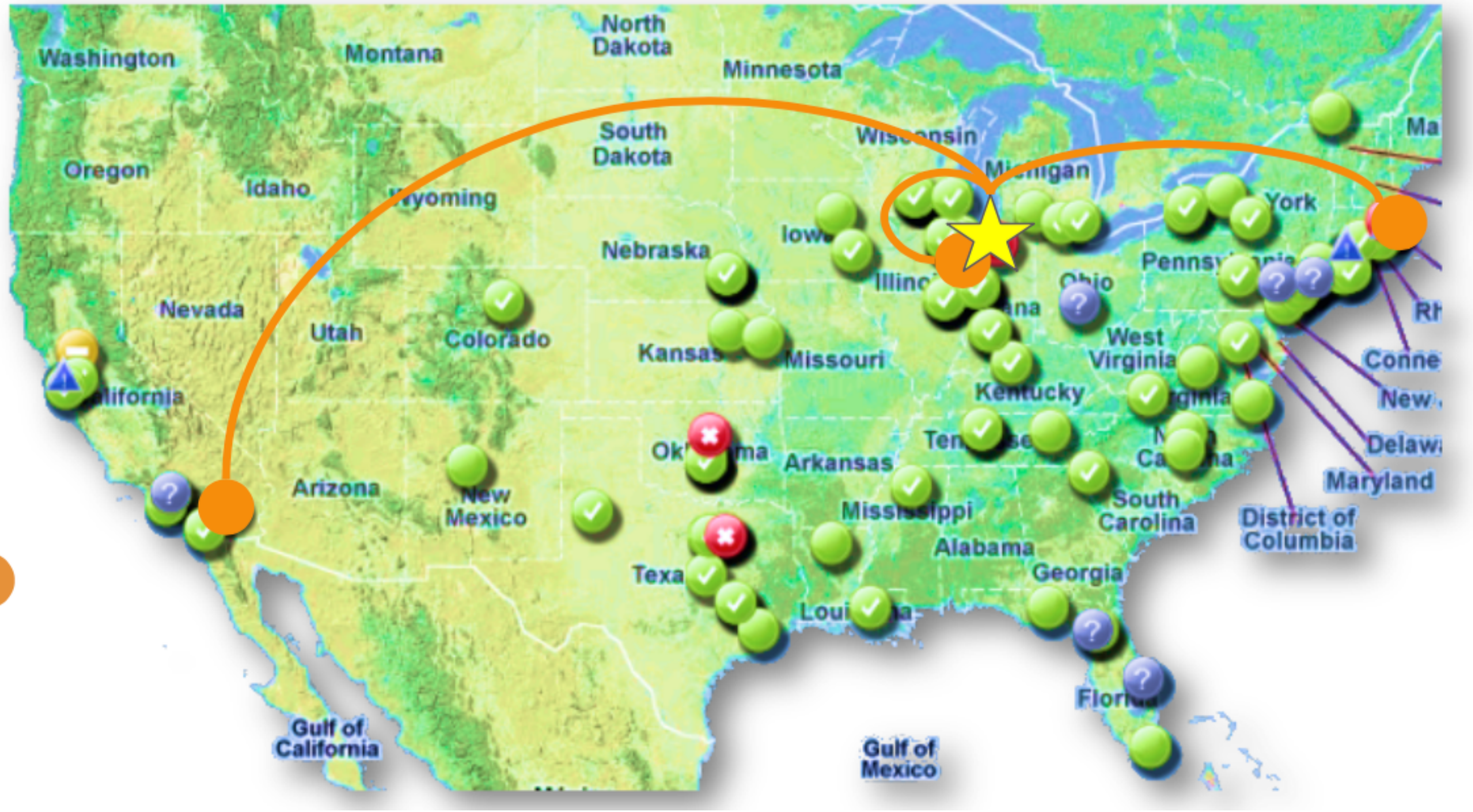
Fermilab

OSG StashCache

- OSG: StashCache
  - Bringing opportunistic storage usage to all users of OSG
  - OSG collaborators provide local disk space
  - OSG is running xrootd cache servers
    - Dynamic population of caches ➜ efficient distributed access to files
      - For users that don't have infrastructures like CMS and Atlas

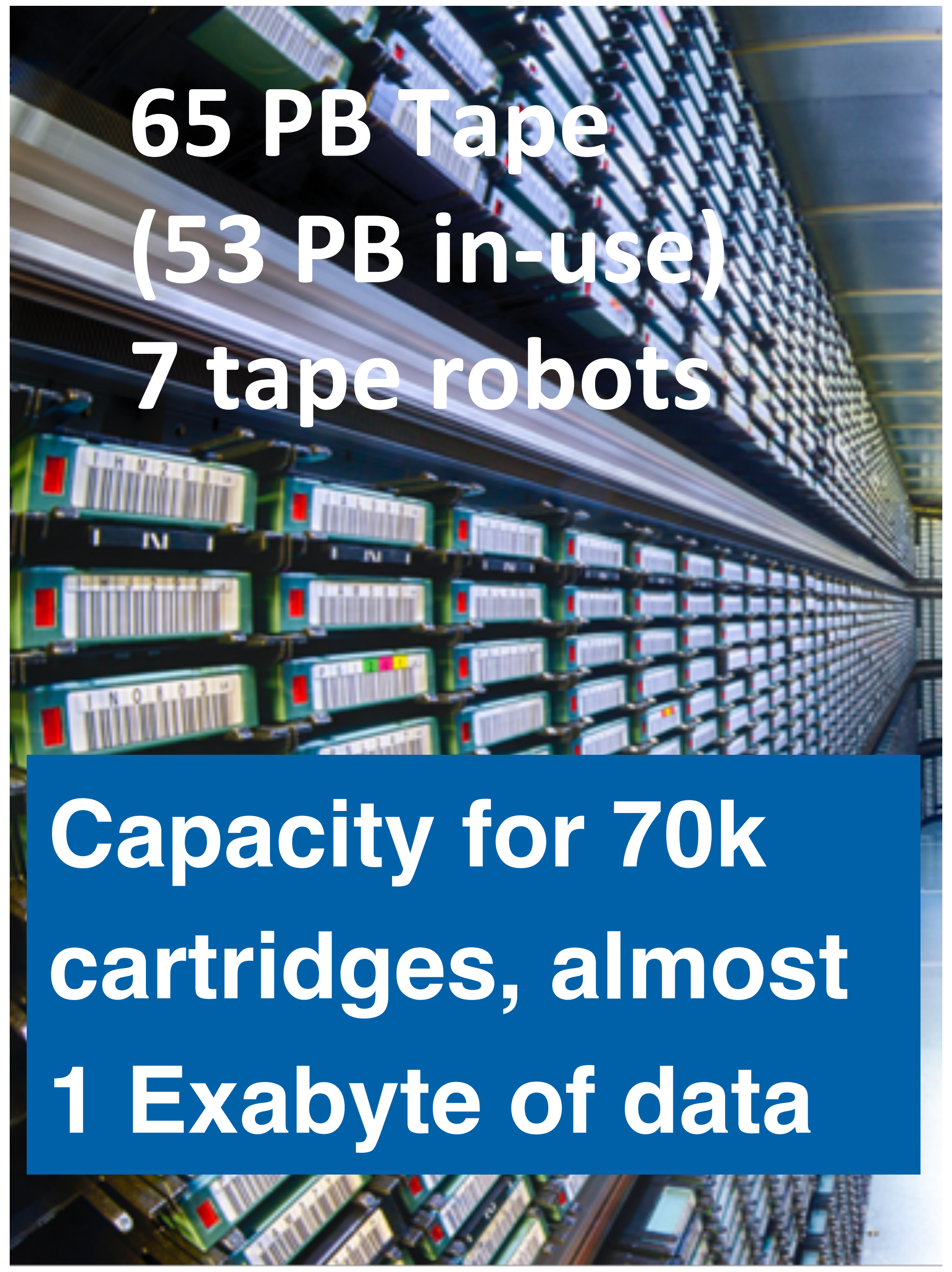
Stash origin: ☆

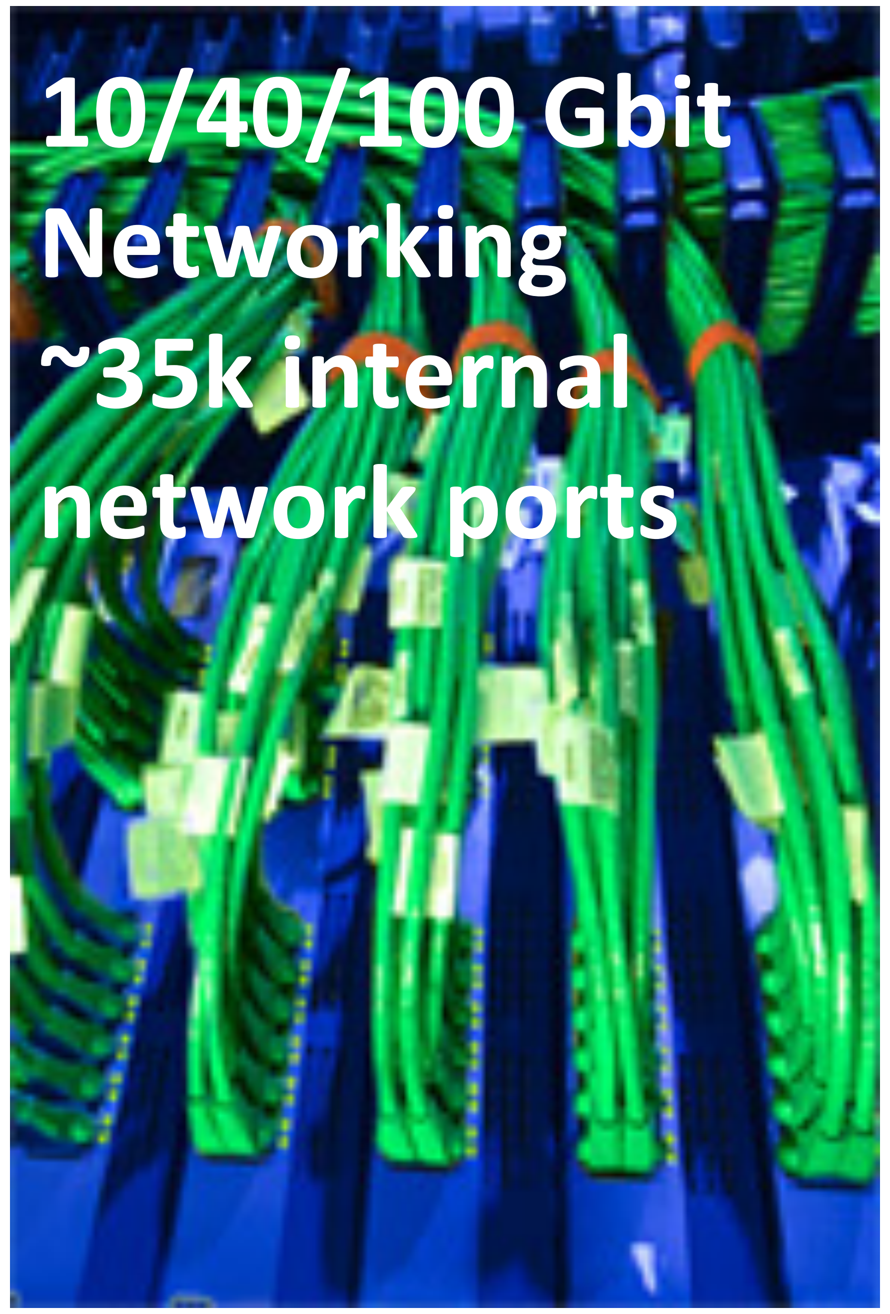
OSG Caches: ●



Fermilab Computing

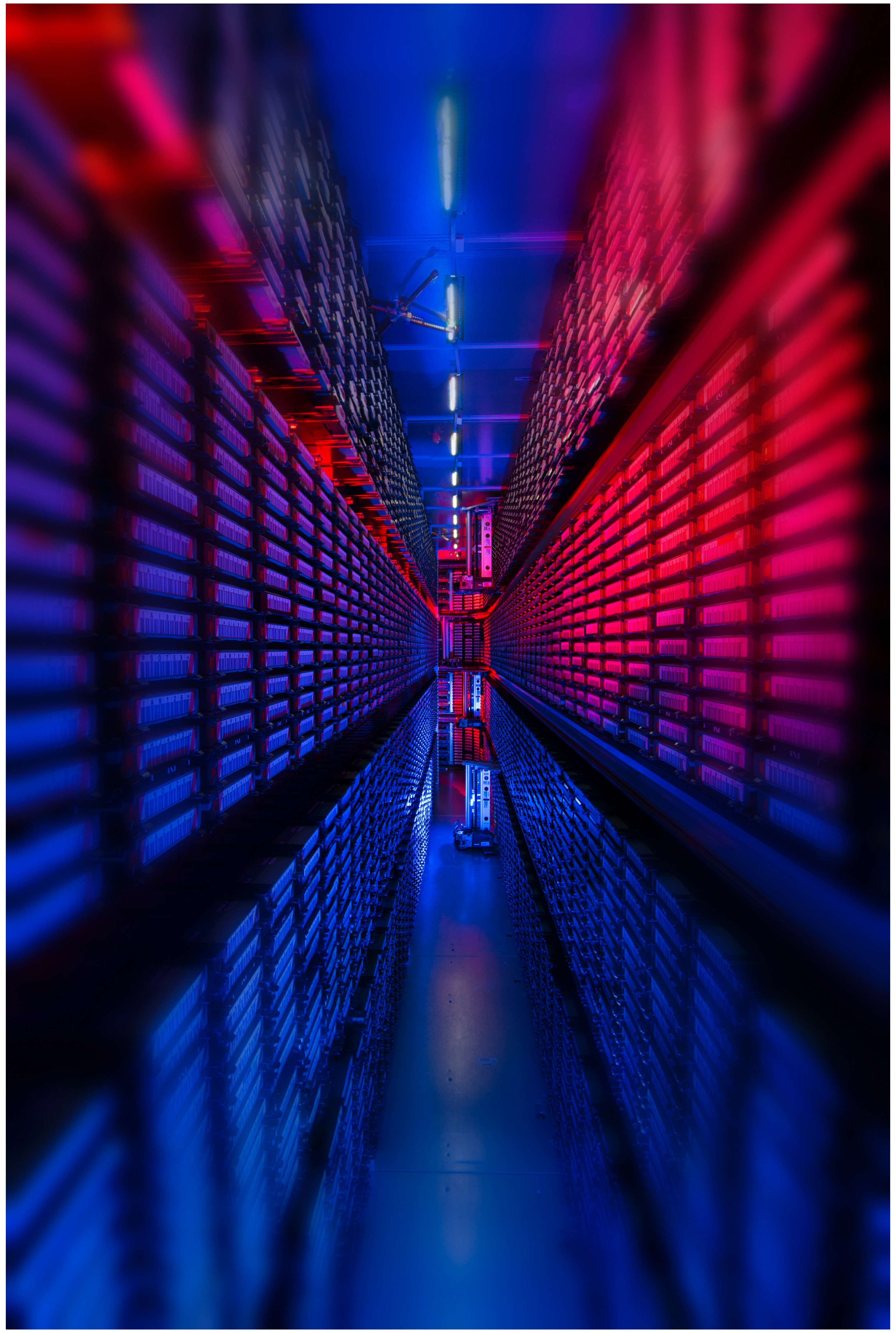
- Provide and manage computing services and resources
- Data recording, storage, access
- Bulk processing, analysis
- Functionality analogous to LHC Tier-0 and Tier-1
- CPU Cores, Online (Disk) and Offline (Tape) Storage, Networking





Active Archival Facility

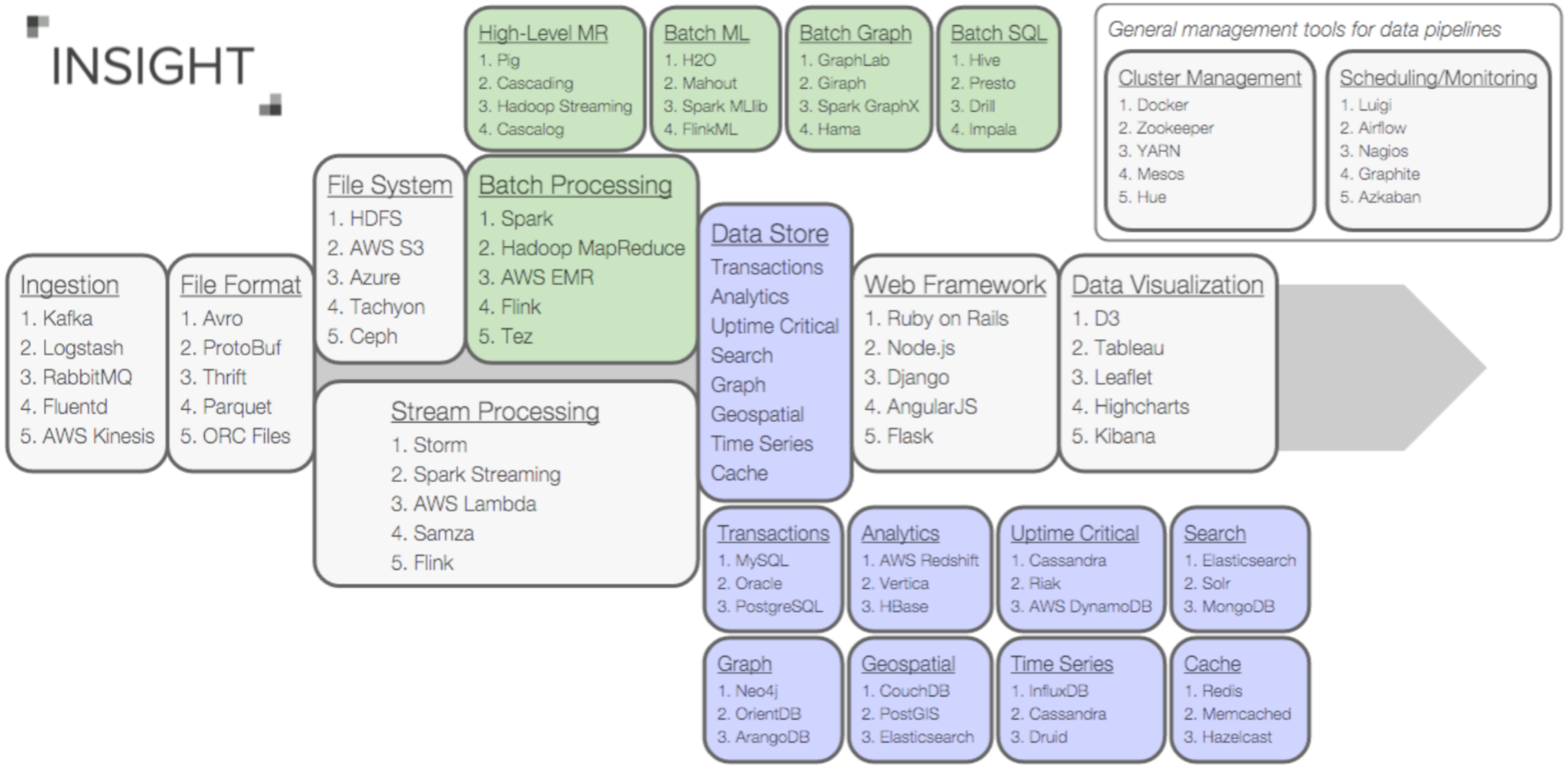
- HEP has the tools and experience for the **distributed exabyte scale**
  - We are “best in class” in the field of scientific data management
- We are working with and for the whole science community
  - To bring our expertise to everyone’s science
  - To enable everyone to manage, distribute and access their data, globally
- Example: Fermilab’s Active Archival Facility (AAF)
  - Provide services to other science activities to preserve integrity and availability of important and irreplaceable scientific data
  - Projects:
    - Genomic research community is archiving datasets at Fermilab’s AAF and providing access through Fermilab services to ~300 researchers all over the world
    - University of Nebraska and University of Wisconsin are setting up archival efforts with Fermilab’s AAF



How do you analyze Exabytes of Data?

Industry

- New toolkits and systems collectively called “Big Data” technologies have emerged to support the analysis of PB and EB datasets in industry.

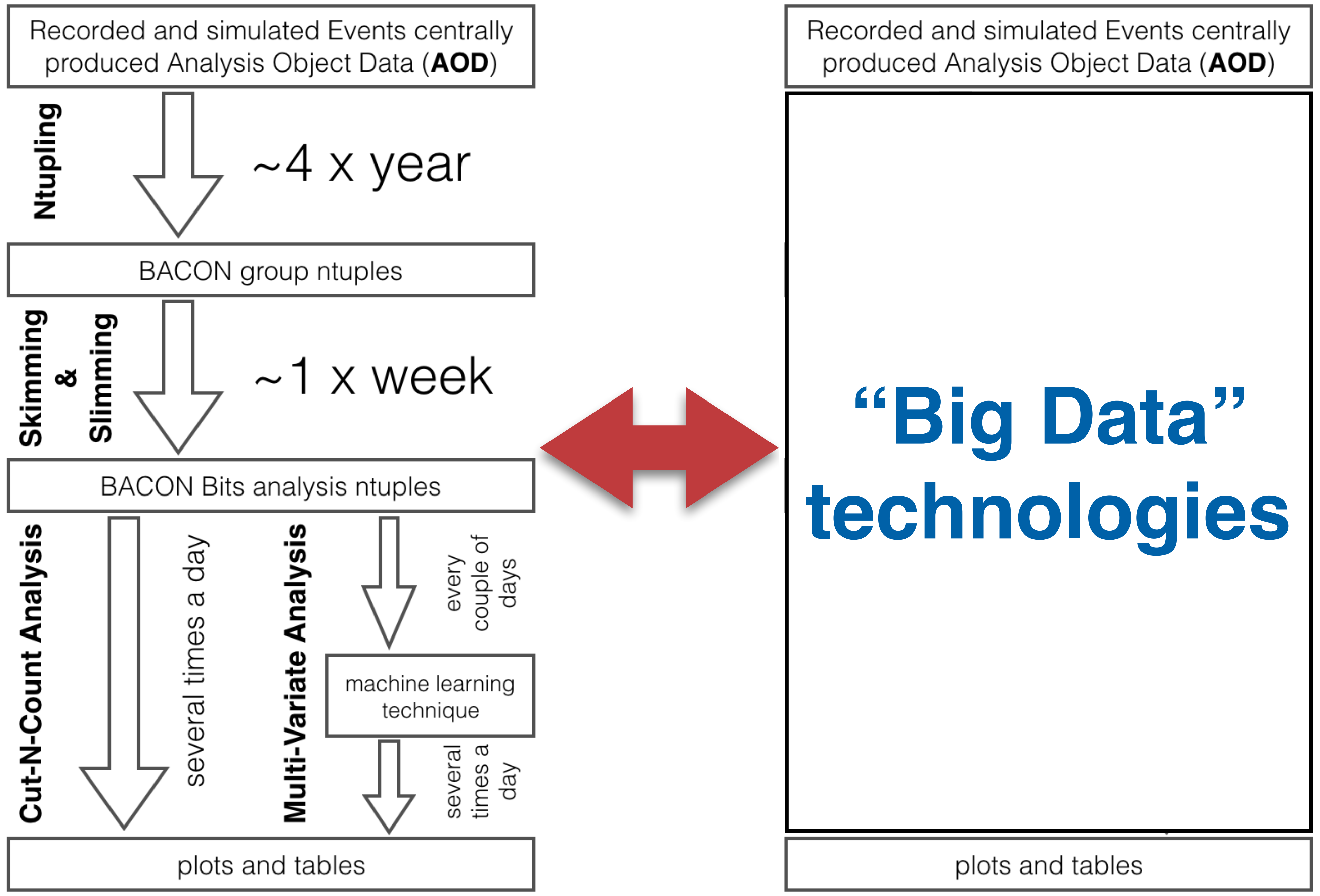


Goals

- Reduce time-to-physics
- Educate our graduate students and post docs to use industry-based technologies
  - Improves their chances on the job market outside academia
  - Increases the attractiveness of our field
- Use tools developed in larger communities reaching outside of our field

A first step: A comprehensive use case study

- Principles of data analysis in HEP have not changed (skimming and slimming experiment-specific data formats)
  - Industry technologies use different approaches and promise a fresh look at analysis of very large datasets and could potentially **reduce time-to-physics with increased interactivity**.
- We want to **use an active LHC Run 2 analysis**, searching for dark matter with the CMS detector, as a **testbed for "Big Data" technologies**



Conclusions

Conclusions

- There is a lot of scientific data!
- The future will bring even more data - exponentially more!
- We have the technology to handle the data of today - will we be able to cope with the data of tomorrow?
- Analysis will be the key challenge in the future ➜ We're working on technologies to analyze Exabytes!


Fermilab
U.S. DEPARTMENT OF ENERGY
Office of Science