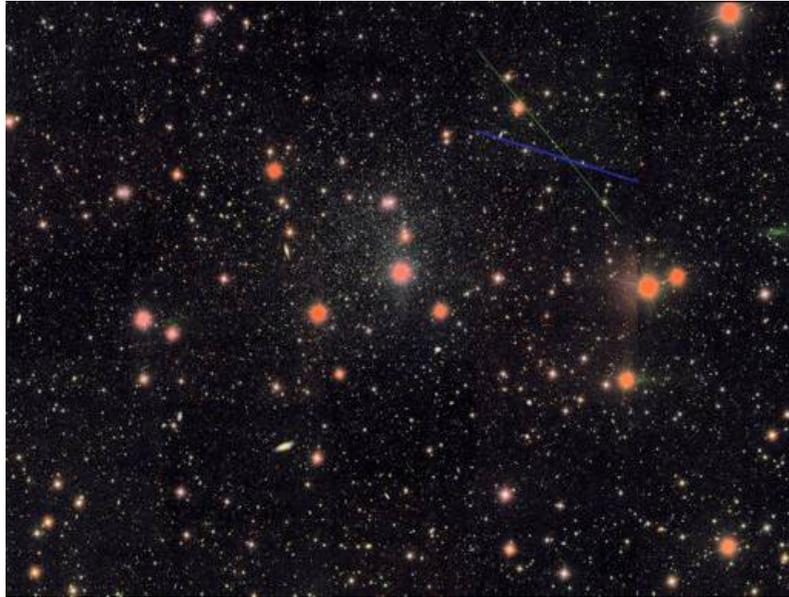
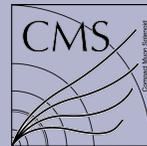


Knowledge Discovery Panel



Standing at the crossing
point between data
analysis and simulation

Dr. M.-C. Sawley
IPP-ETH Zurich



Towards a better comprehension of complex phenomena and systems

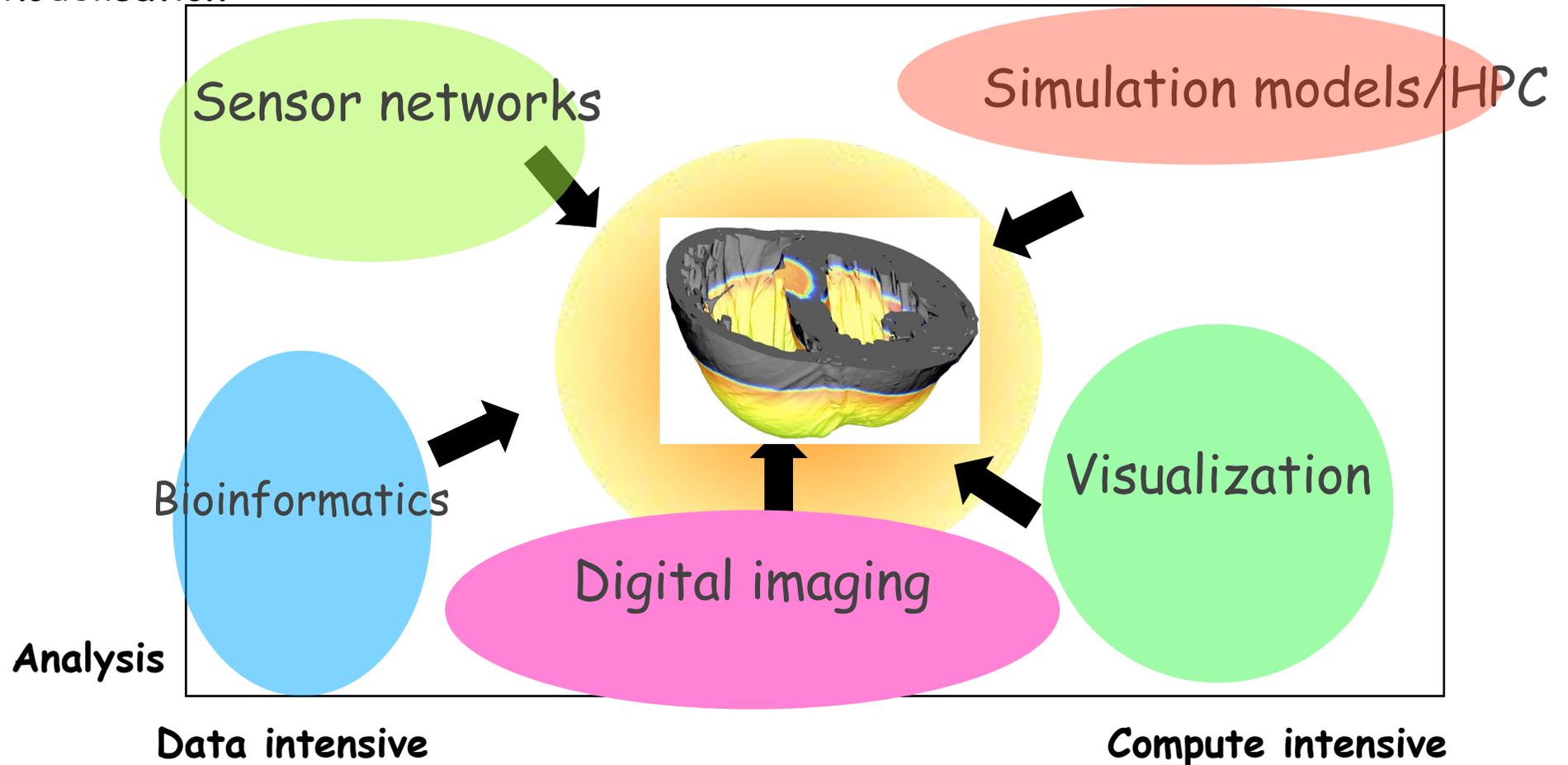
- **Astronomy, astrophysics**
- **Bioinformatics**
- **Systems Biology**
- **Oil & Gas**
- **Meteorology, Oceanography, Volcanology**
- **Bio medical engineering**
- **Automotive, Combustion, Aeronautics**
-(many more to come)



Integration of multi scale, multimodal, multi disciplinary Technologies

Drawing a (far too simple) integration map of available technology

Modelisation



Unravelling the complexity of Nature

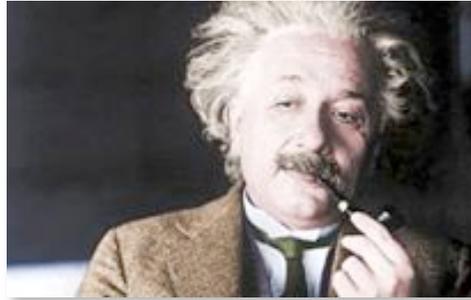
■ Challenges for data driven science

- On-line filtering the deluge of experimental or observational data
- Repacking (data reduction) into high level objects, validation, calibration, quality control
- Analyzing at fine granularity → accessibility, network capacity, data curation, heterogeneity of the systems
- Using data to enrich modelisation, simulation
- Integrating new data, new knowledge on the way

■ Drives and opportunities

- Driven by sensor physics, high resolution images, ...
- Requires setting up of highly complex and heterogeneous infrastructures
 - The chain is as strong as its weakest link
- Balance between
 - collecting, filtering, simulating, distributing and interpreting very large amount of data which comes into large bursts
- Role of HPC?

The art of making knowledge discovery as simple as possible.....

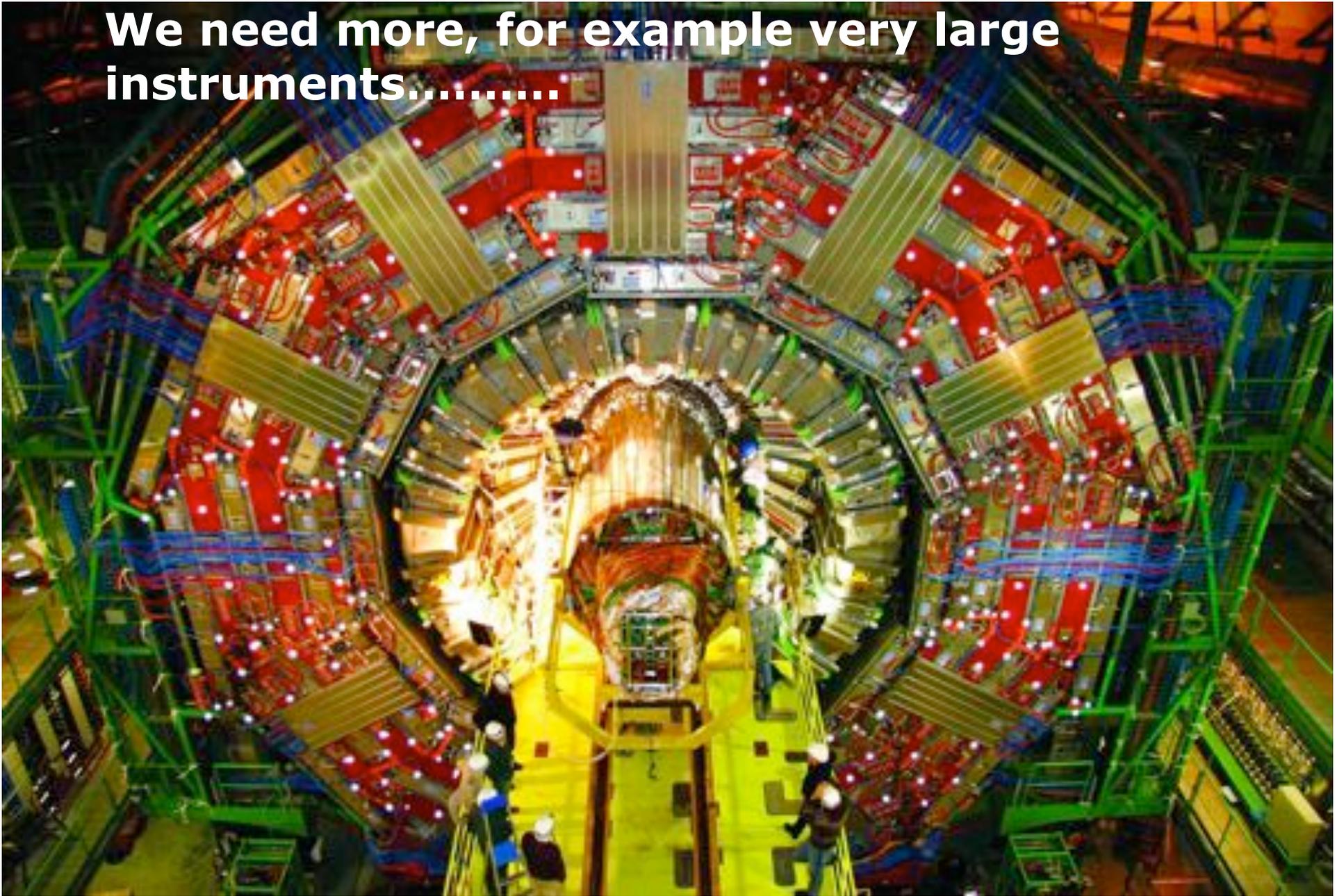


*The most incomprehensible
thing about the Universe is
that it is comprehensible!*

A. Einstein

Not everybody is able to “connect the dots” as quickly as this!

We need more, for example very large instruments.....

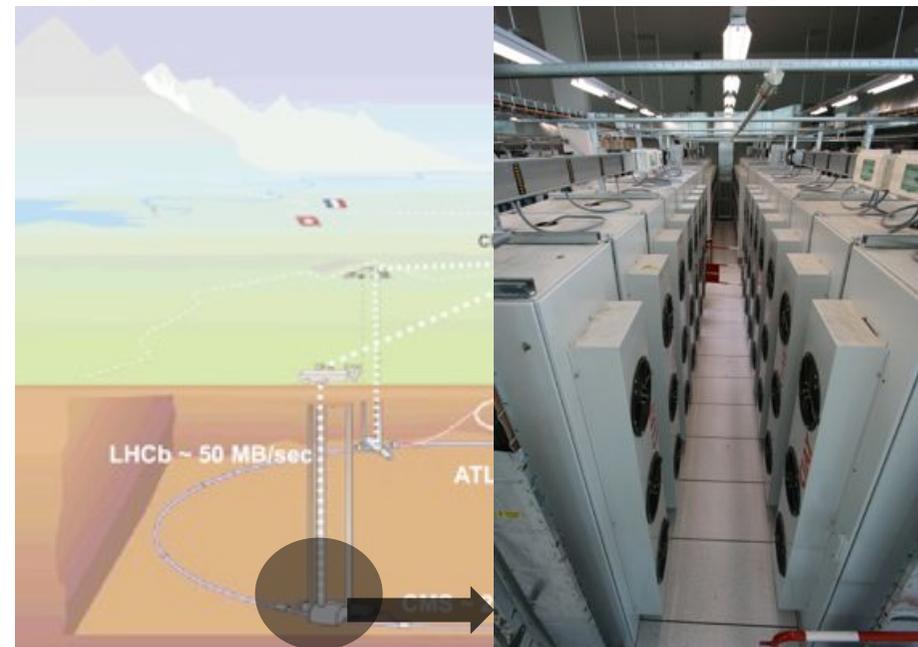


At the detector site: Online computing

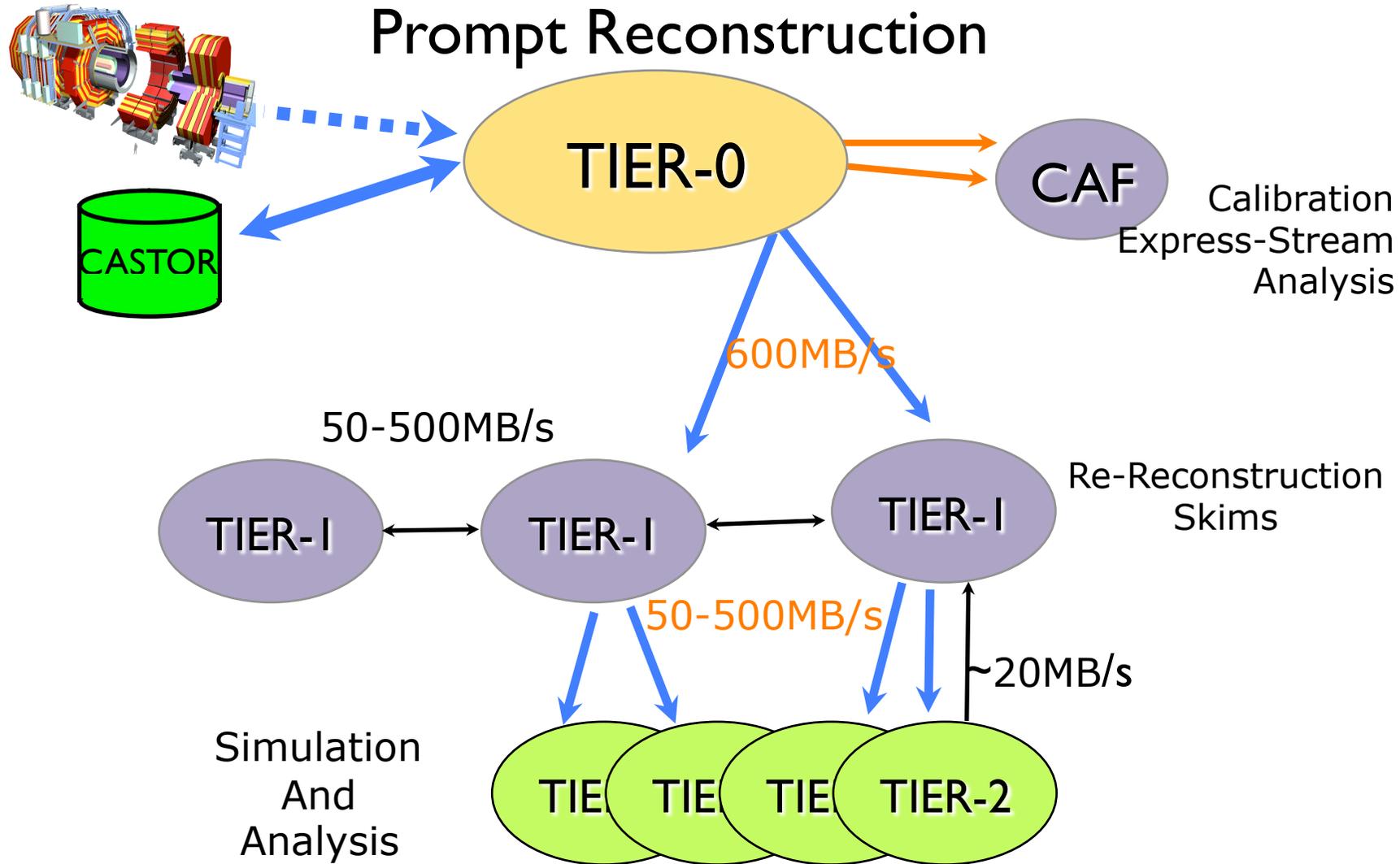
High level trigger:

80 millions electronic channels
X4 (each of them using 4 bytes)
X40 millions (collision rate 40 MHz)
X1/1000 (zero suppression)
X1/100 000 (on line event filtering)
→ O(10) PB/y sent to CERN IT

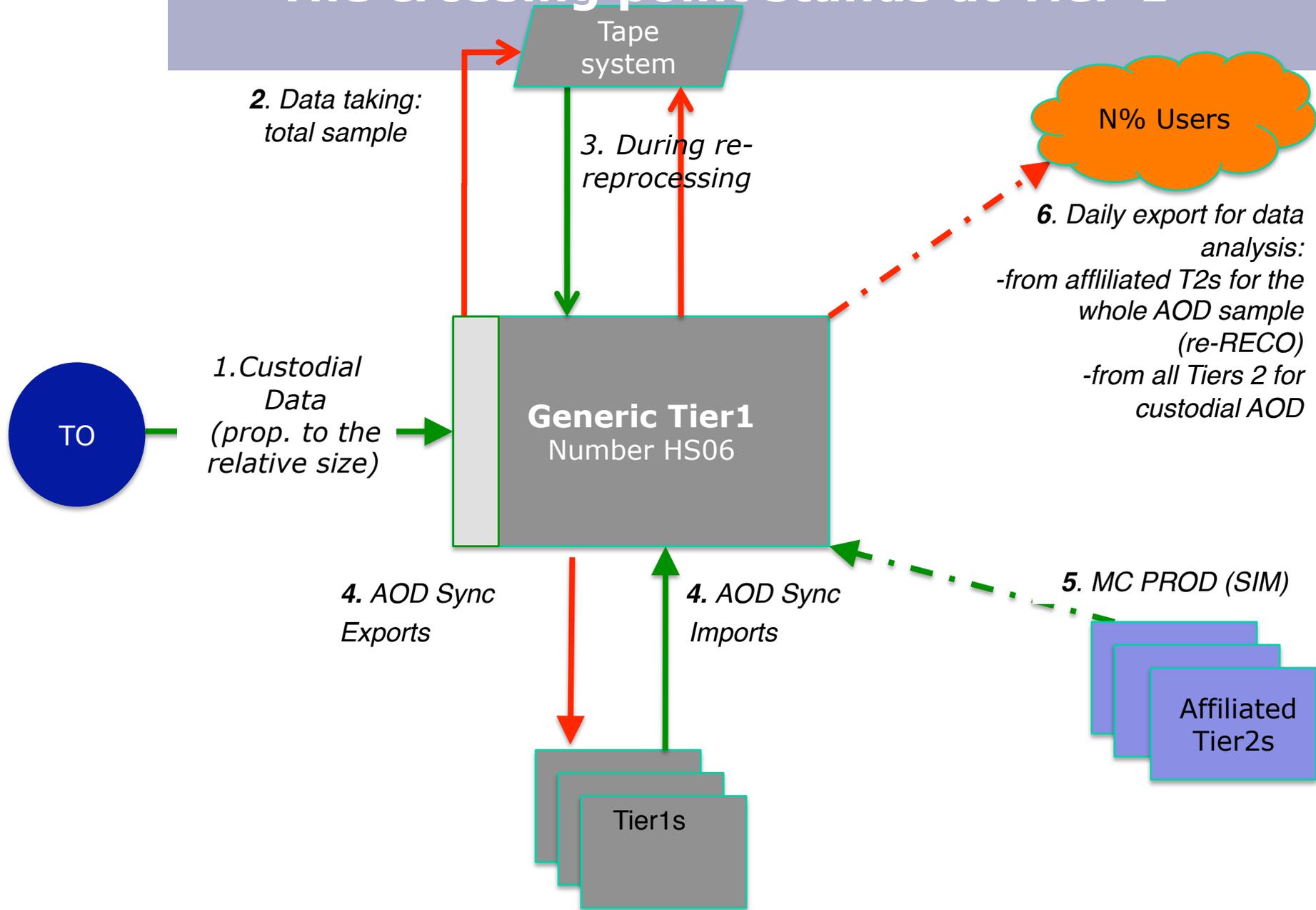
- DELL cluster (33 racks, dualcore Harpertown)
- 230 TB disks acquisition system



Extracting scientific knowledge :CMS computing



The crossing point stands at Tier-1



Participants to the panel

- **Nagiza Samatova, Oak Ridge National Labs**
- **Gerald Kneller, University of Orleans**
- **Ron Oldfield, Sandia National Labs**

Questions for the panel

- **What are the challenges for bridging data analysis and simulation?**
- **What is the role of HPC for mining scientific discovery?**
- **How would you define the “P” of HPC : Performance, Productivity, Portability, Pain, ...?**
- **What would your wish list to the HPC community consist of?**
- **How do you expect the data wealth and complexity to influence simulation models and methods?**