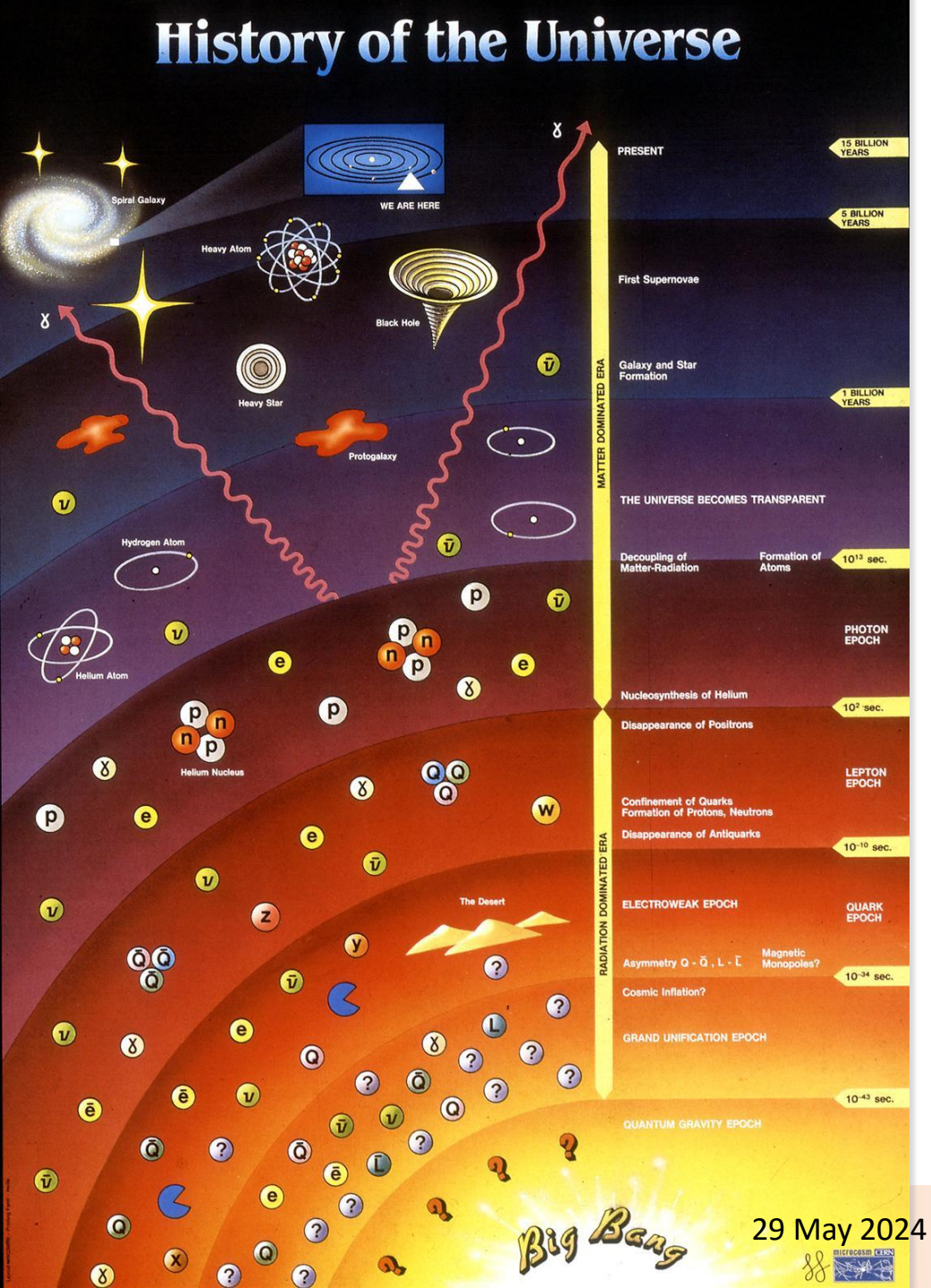
NORDUGRID
Grid Solution for Wide Area Computing and Data Handling

ARC

ARC-CE motivation and use case: LHC computing Tier1

Oxana Smirnova
Lund University / NeIC

neic



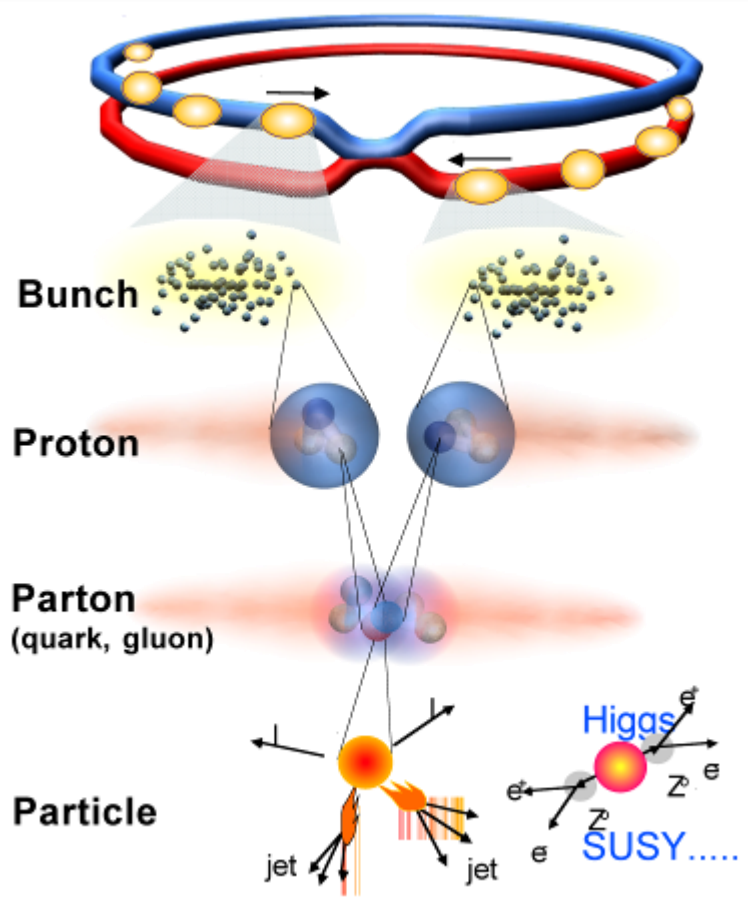
A use case: particle physics at the Large Hadron Collider

The more we know, the more we don't

- Why do particles have the masses we observe? What is the origin of mass?
  - *Discovery of the Higgs particle at CERN helps*
- Are there states of matter at very high density and temperature (Big Bang)?
  - *Indications of the Quark-Gluon Plasma get stronger at CERN*

- How many space-time dimensions do we live in?
- What is the nature of the Dark Matter and Dark Energy that dominate the Universe?
- Can gravity be added to the theory of the other known forces (Standard Model)?
- Are the known elementary particles fundamental or do they have a structure?
- Why is the electrical charge of the electron equal and opposite to that on the proton?
- Why are there three generations of quarks and leptons?
- Why is there overwhelmingly more matter than anti-matter in the Universe?
- Are protons unstable? Or, why are they stable?
- Do the neutrinos have mass, and if so, why are they so light?
- … and many more …

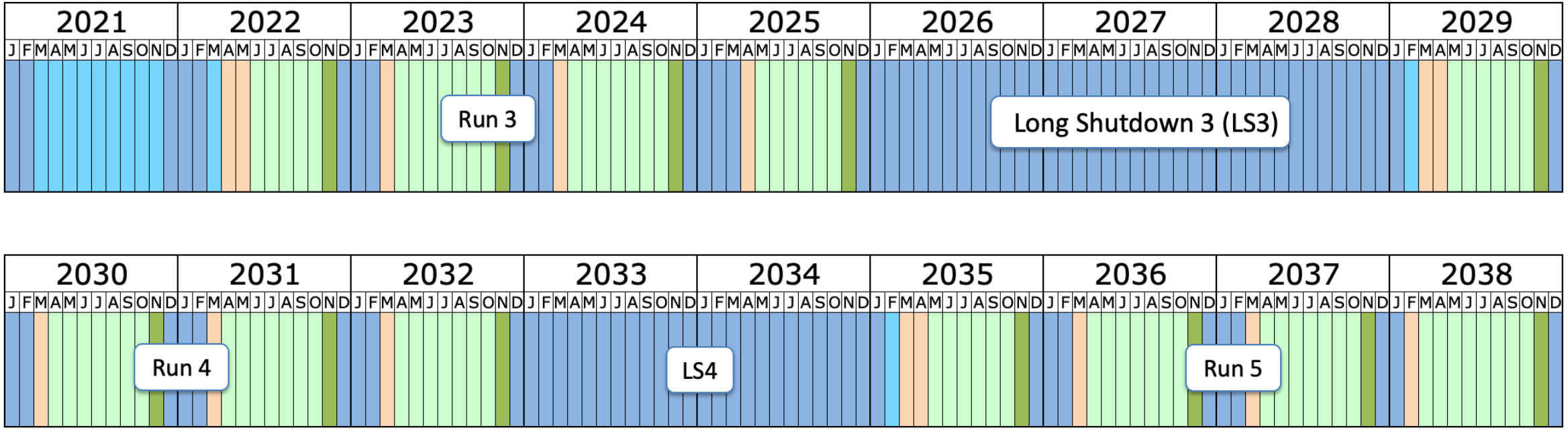
Large Hadron Collider: a discovery machine



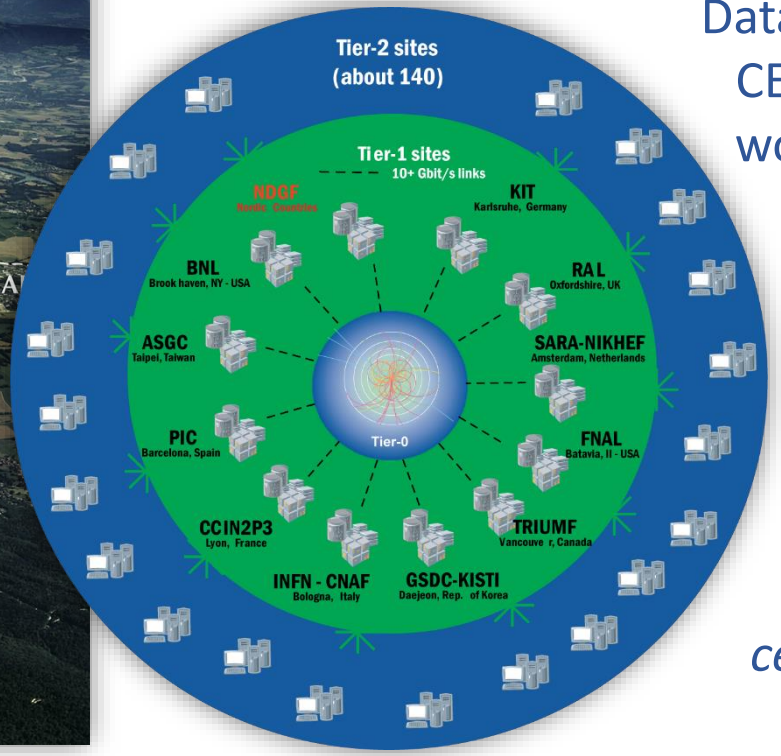
Proton-Proton	2835 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	10^{34} cm^{-2} s^{-1}
Crossing rate	40 MHz
Collisions rate	$\approx 10^{7}$ - 10^{9} Hz
New physics rate	$\approx$ .00001 Hz

Event selection:
1 in 10,000,000,000,000

LHC operations schedule

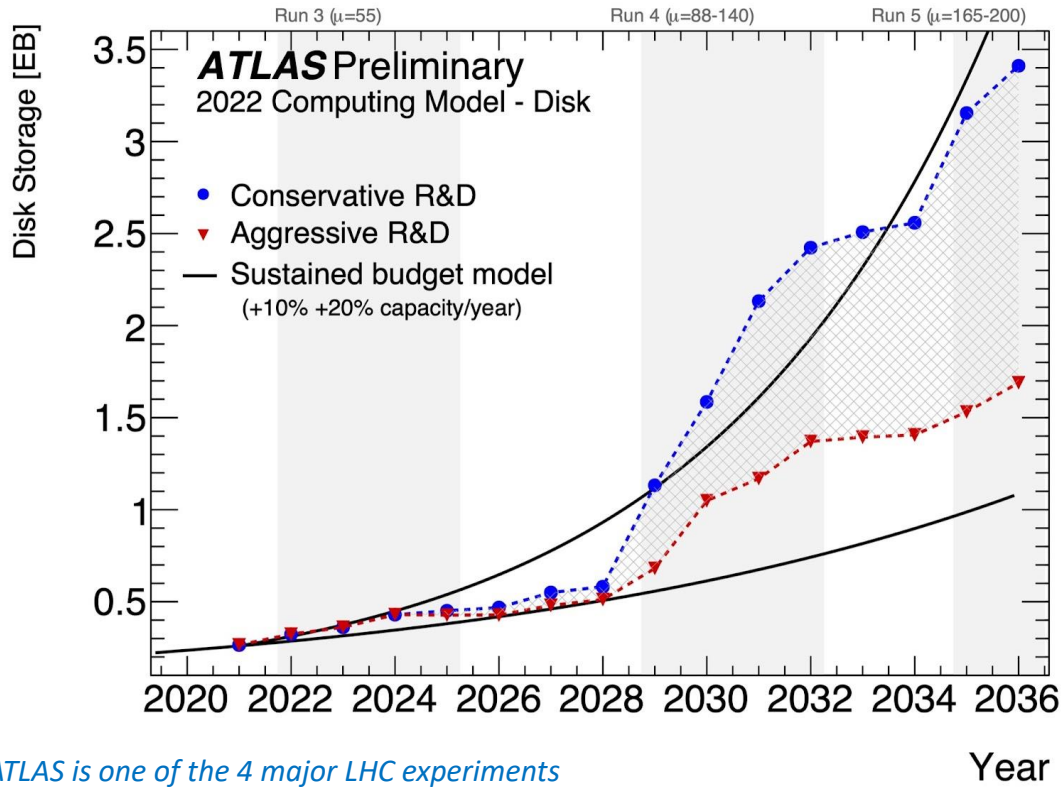


Large Hadron Collider will deliver *Exabytes* of data by 2038



Data are sent from CERN around the world for storage and processing

12 regional Tier1 centres at the core



ATLAS is one of the 4 major LHC experiments

Data is the challenge

- *Must be FAIR: Findable, Accessible, Interoperable and Reusable*
- *FAIRness requires a large and sustainable infrastructure*
- *Contrary to CPU/GPU cycles, storage can not be re-used*

Solution: a *Data Lake* spanning countries
+ special middleware for data-intensive tasks

Nordic Tier1 as a user-driven infrastructure

- Established by LHC user communities in 4 Nordic countries
  - Denmark, Finland, Norway and Sweden, total of 6 sites
  - Storage in Slovenia and Switzerland, too
- 2 LHC experiments: ALICE and ATLAS
  - Different requirements and workflows
  - Other non-LHC users: IceCube, EISCAT_3D (under investigation)
- A unique data center that spreads 4 countries
  - No single Nordic data center could meet all the user needs
    - Not even now: **disk and tape storage** is still a unique requirement
  - Services are pledged by the Nordic countries via 4 dedicated MoUs with CERN

Nordic Tier1 implementation needs special middleware

- Key design principle: CPU cycles decoupled from storage
  - CPU cycles provided via national allocations
  - Storage (disk and tape) provided on a more long-term basis
  - Multiple data centres across several countries provide the services
- Key middleware requirement: Tier1 must present itself as a single site to the users
  - Single storage entry point – implemented by **dCache**
  - Data must be pre-cached for processing – implemented by **ARC-CE**
  - Internal accounting – implemented by **SGAS**
  - Nordic Tier1 contributes to this middleware development

ARC-CE: Compute Element of ARC

- ARC: **Advanced Resource Connector**
  - Implements standard external interface for job submission to a computing resource
  - Complete with data stage-in, stage-out and caching
- ARC-CE is also used as a conventional Grid Compute Element
  - Decoupling from storage and worker nodes makes it suitable for HPC
  - Pluggable interface to any batch system
  - REST interface for external communications
  - Together with ARC Control Tower (aCT) can be used to handle classical HPC workflows

ARC usage around the world