

A short introduction to the Worldwide LHC Computing Grid

September, 2024

The LHC challenge

- 200+ PetaByte / year
 - Current total 2.0+ ExaByte
- Data analysis requires at least $\sim 1M$ typical PC processor cores
- Scientists in tens of countries worldwide
- CERN can provide up to 20-30% of the storage and CPU
- We need a GRID !



What is a grid?

- Relation to WWW?
 - Uniform easy access to shared information
- Relation to distributed computing?
 - Local clusters
 - WAN clusters
- Relation to distributed file systems?
 - NFS, AFS, DFS, ...



- A grid gives selected user communities uniform access to distributed resources with independent administrations
 - Computing, data storage, devices, ...

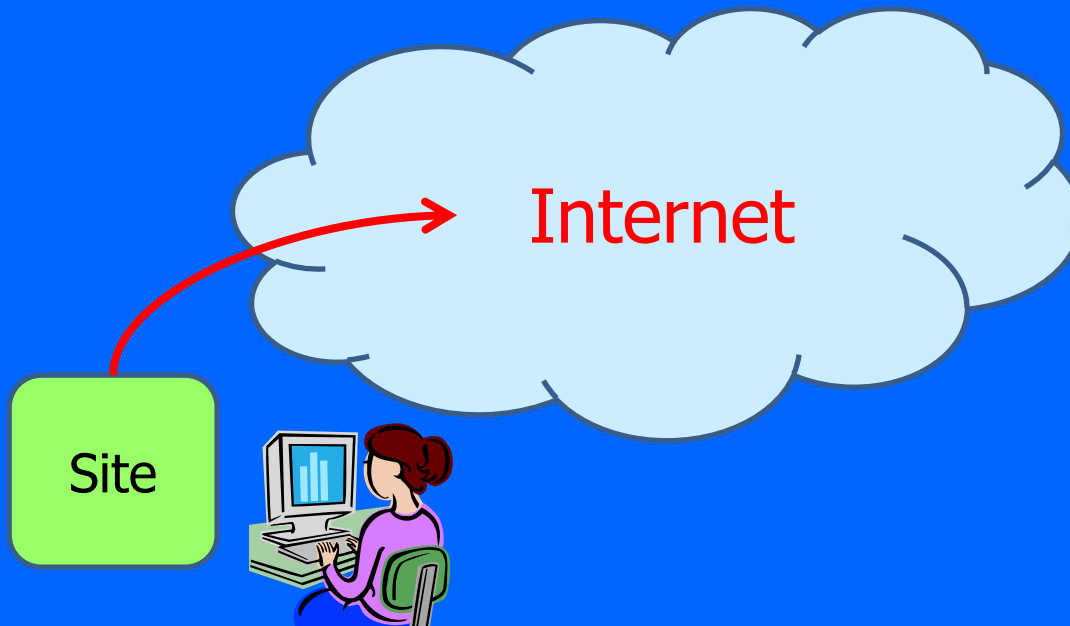
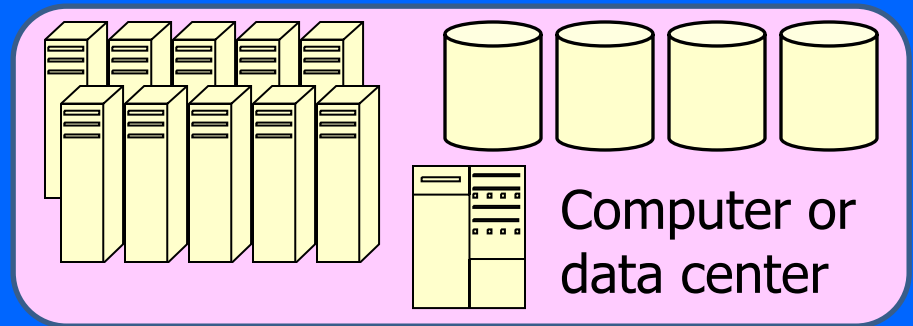
Why is it called grid?

- Analogy to power grid
 - You do not need to know where your electricity comes from
 - Just plug in your devices
- You should not need to know where your computing is done
 - Just plug into the grid for your computing needs
- You should not need to know where your data is stored
 - Just plug into the grid for your storage needs



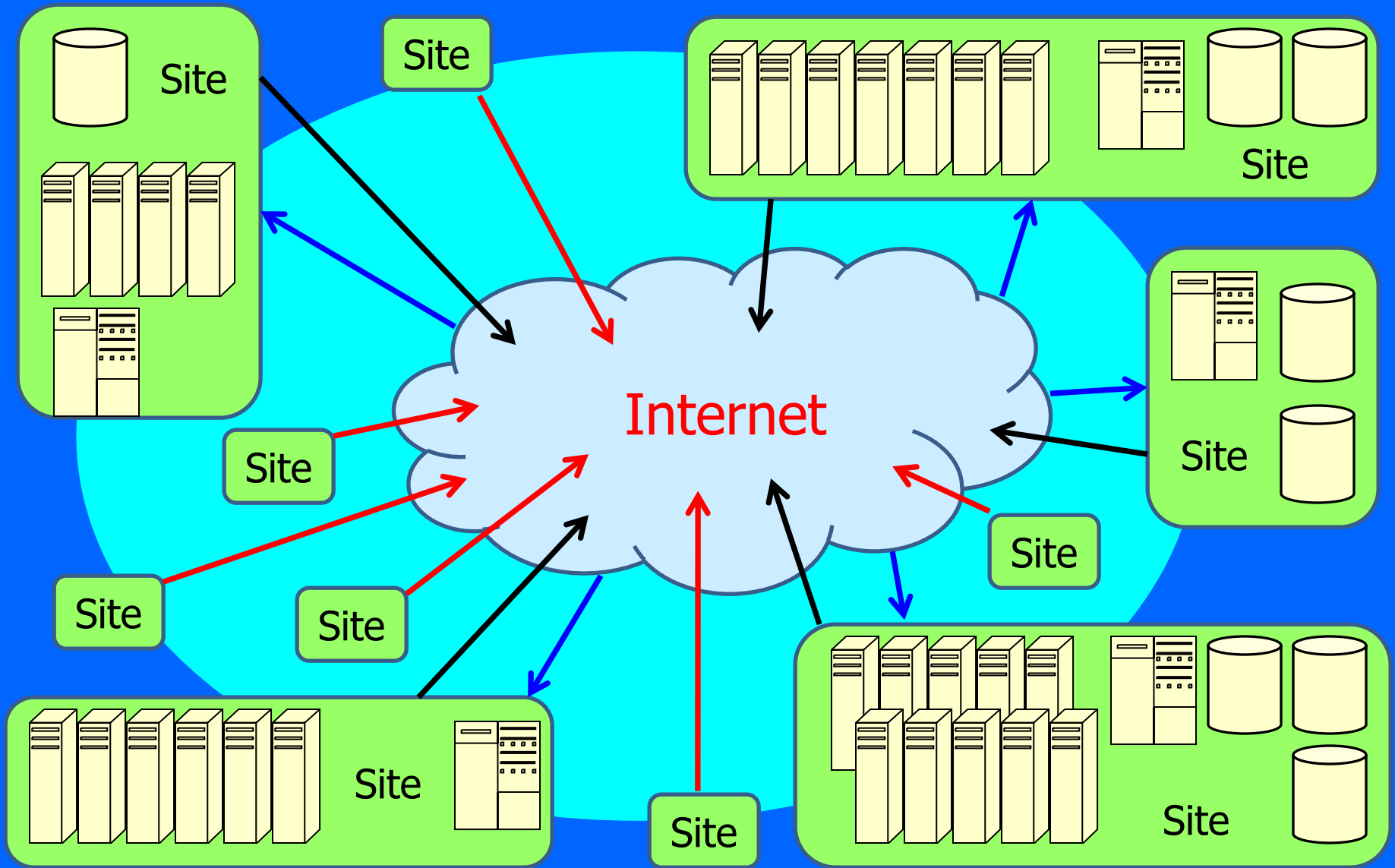
What is cloud computing?

- Transparent use of generic computing resources off-site
 - Dynamically provisioned
 - Metered
- Neutral to applications
 - Rent-a-center



- Amazon EC2, S3
- Microsoft Azure
- Google
- IBM
- Alibaba
- Oracle
- ...

What is grid computing?

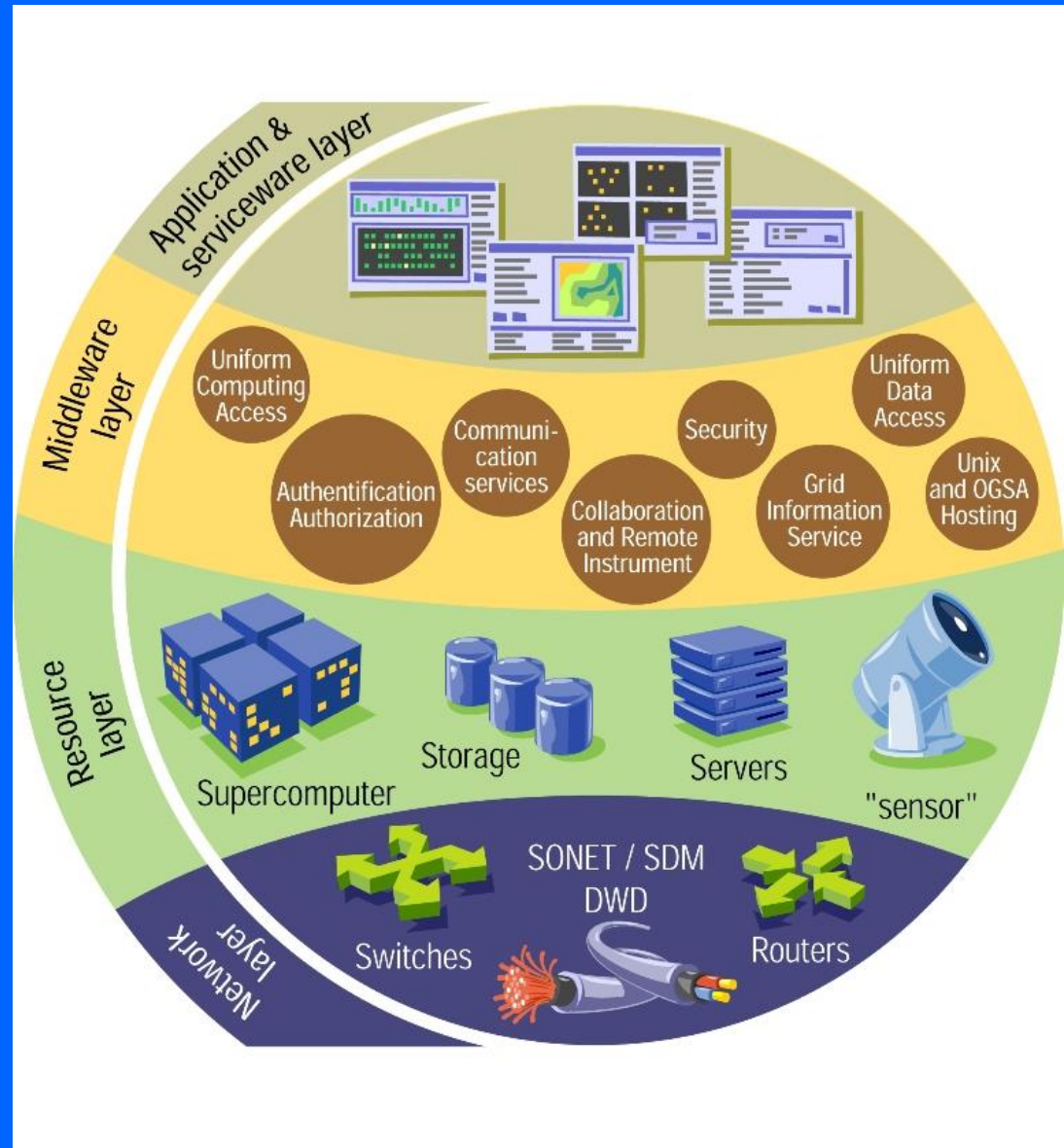


What is grid computing about?

- A grid facilitates collaboration between members of a supported distributed community
 - They can form a Virtual Organization within that grid
- A grid allows distributed resources to be shared uniformly and securely for common goals
 - Computing
 - Data storage
- A grid can support multiple Virtual Organizations in parallel
 - Sites, computer and data centers make selections according to the projects in which they participate
 - The quality of service may differ per VO

How does a grid work?

- Middleware makes multiple computer and data centers look like a single system to the user
 - Security
 - Information system
 - Data management
 - Job management
 - Monitoring
 - Accounting
- **Not easy!**
 - Independent sites
 - Different systems
 - Local policies/priorities
 - Other users



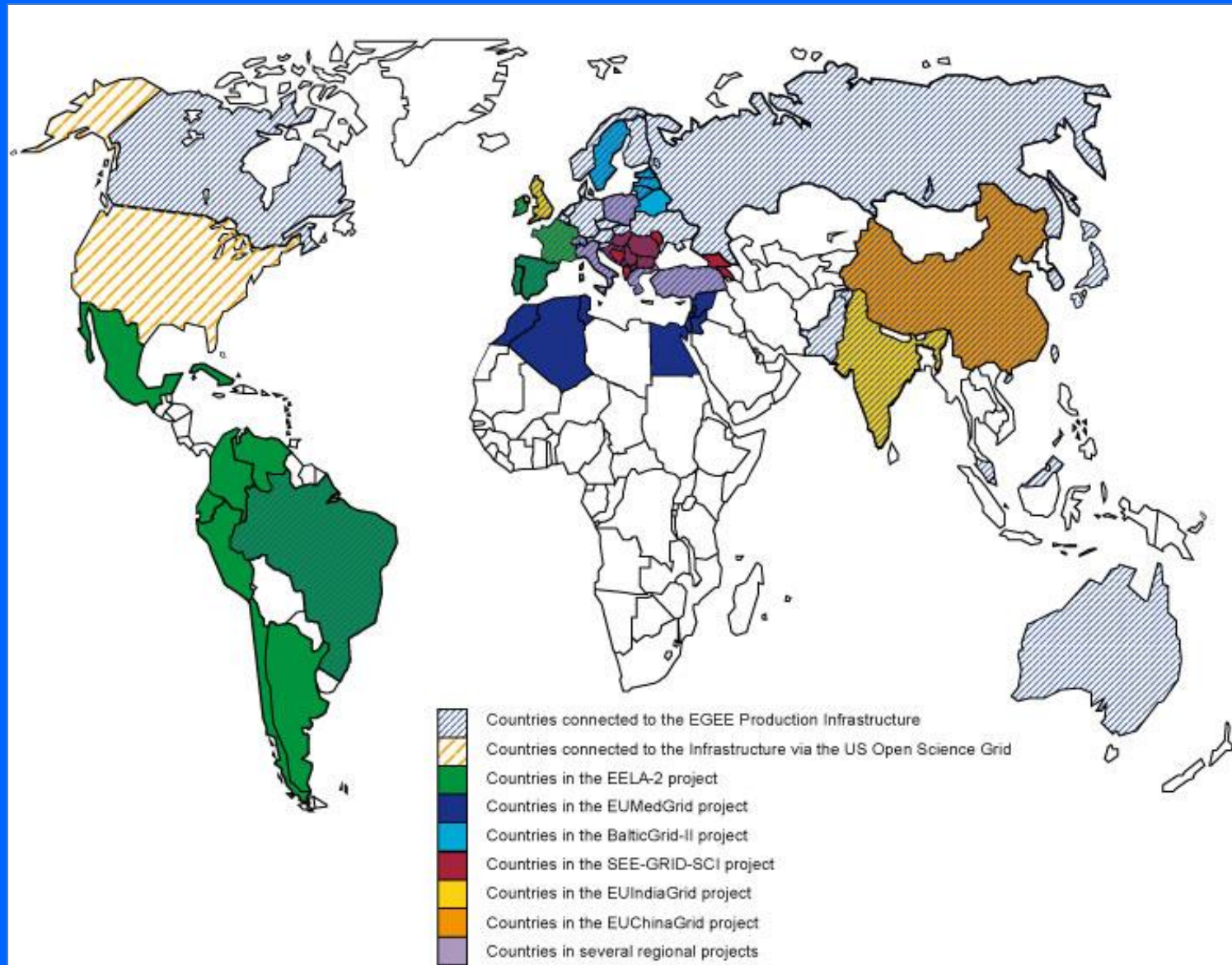
Where can we use grids?

- Scientific collaborations
 - Can also serve in spreading know-how to developing countries
- Industry? Commerce?
 - Mostly cloud computing
- Homes? Schools?
 - Internet Service Providers → cloud computing
 - Secure data sharing technologies?
 - E-learning
 - Social media
- Government? Hospitals? Other public services?
 - Beware of sensitive/private data

Many grids can coexist

- EGI – European Grid Infrastructure
 - Successor of EGEE – Enabling Grids for E-science ← led by CERN
- OSG – Open Science Grid
 - USA and beyond
- National
 - IGI (It), GridPP (UK), DFN (De), France Grilles, ...
- Regional
 - NorduGrid (Nordic countries), BalticGrid (Baltic region), SEEGrid (South-East Europe), EUMedGrid (Mediterranean), ...
- Interregional
 - EELA (Europe + Latin America), EUIndiaGrid, EUAsiaGrid, ...
- WLCG – Worldwide LHC Computing Grid
 - Federation of EGI, OSG, Nordic Data Grid Facility, ...

Projects that collaborated with EGEE



Many communities can coexist

- High-energy physics
- Astrophysics
- Fusion
- Computational chemistry
- Biomed – biological and medical research
 - Statistical analysis of anonymized data
 - “In silico” discovery of new drugs and vaccines
 - ...
- Earth sciences
- UNOSAT – satellite image analysis for the UN
- Digital libraries
- ...

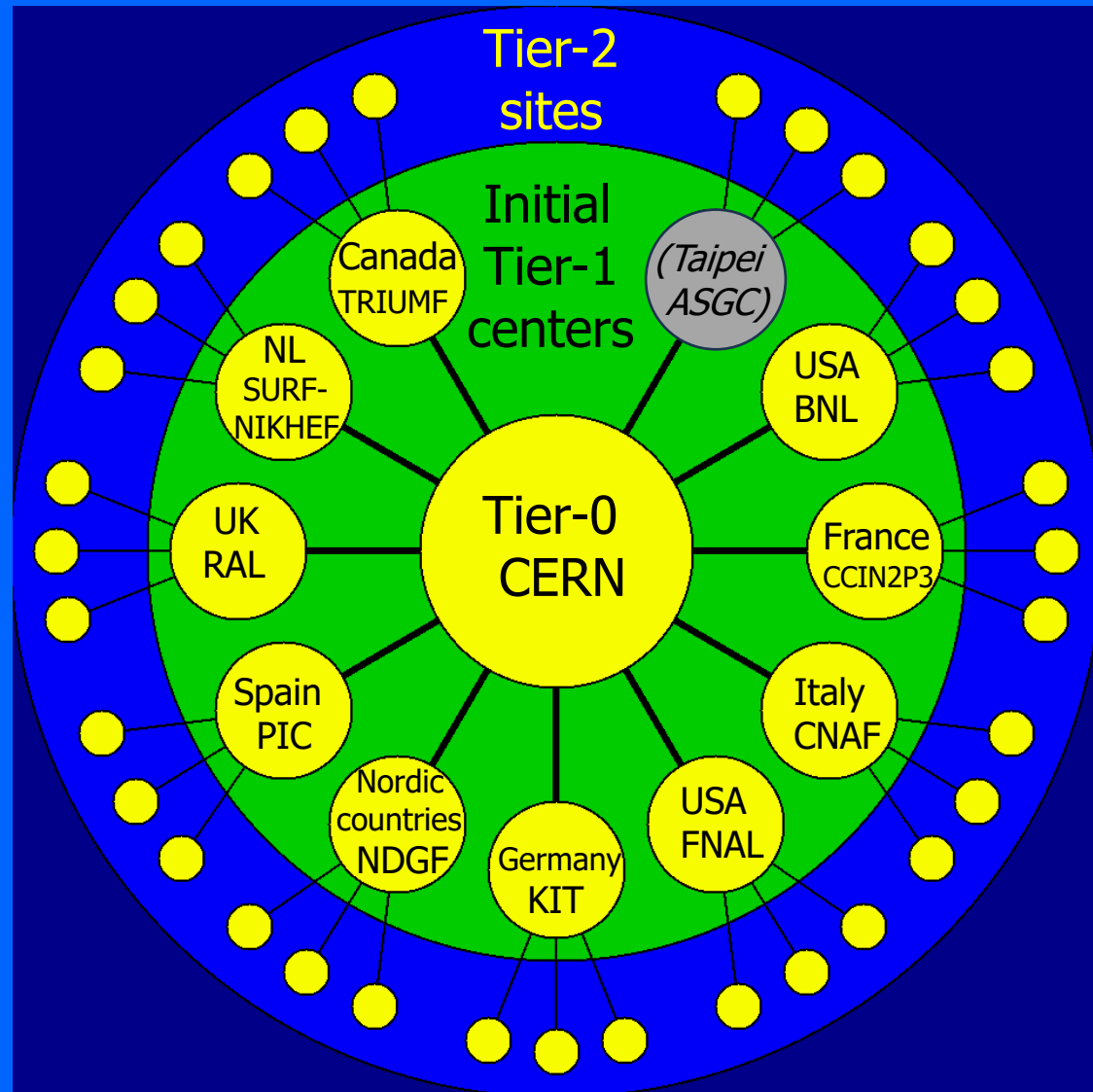
The LHC challenge

- 200+ PetaByte / year
 - Current total 2.0+ ExaByte
- Data analysis requires at least $\sim 1M$ typical PC processor cores
- Scientists in tens of countries worldwide
- CERN can provide up to 20-30% of the storage and CPU
- 70-80% are provided by WLCG partners



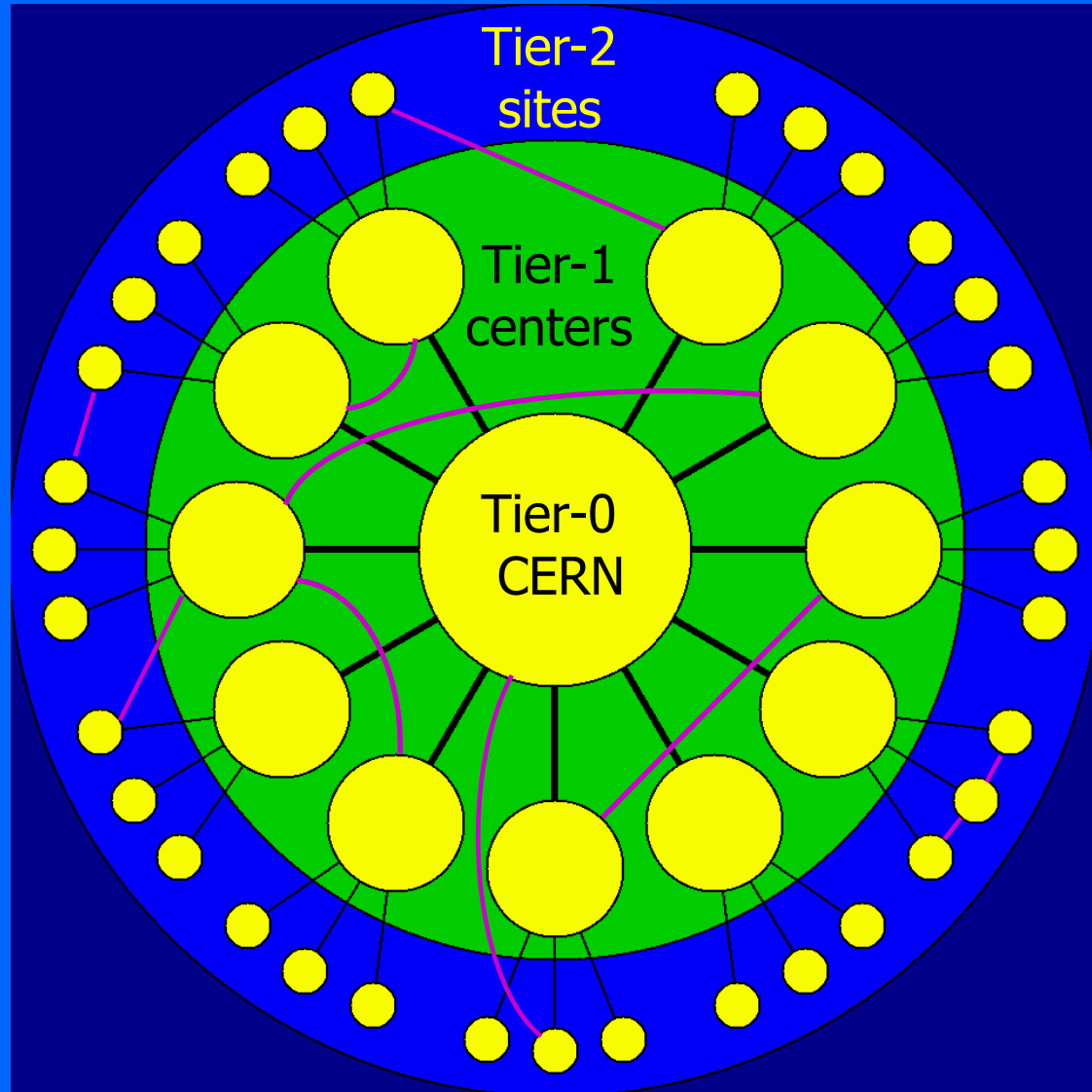
The WLCG answer

- 160+ computing centers
30+ countries
- Hierarchical and regional organization
- 16 large centers for long-term data management
 - CERN = Tier-0
 - 15 Tier-1 centers →
 - South Korea: KISTI
 - Russia: JINR, NRC-KI
 - Poland: NCBJ
 - China: IHEP
 - Dedicated network links
- 60+ federations of 140+ smaller Tier-2 centers
- Tens of Tier-3 sites
 - Resources outside of WLCG policies



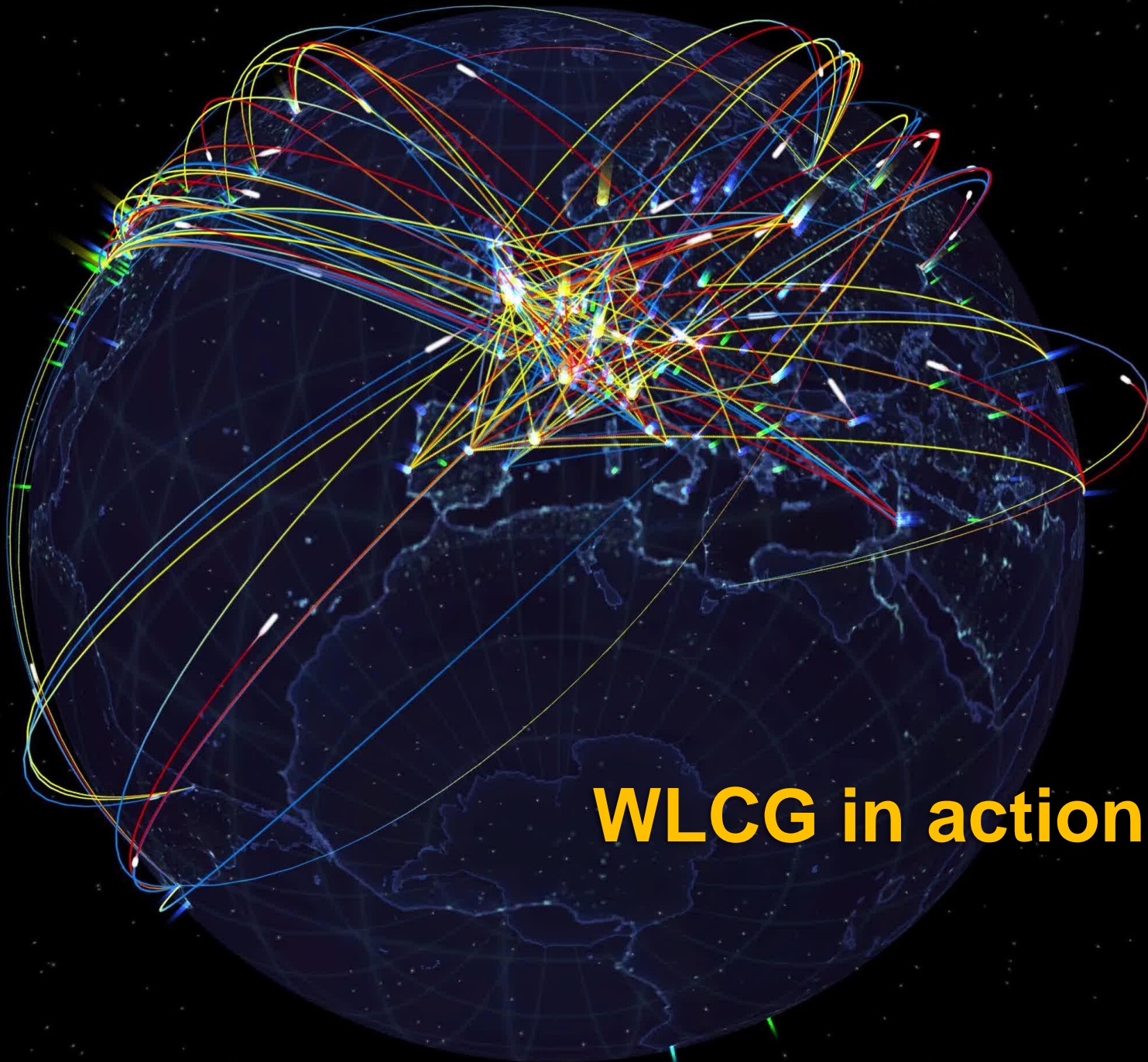
The WLCG evolution

- Fast networks allow for direct transfers between:
 - Tier-1 sites
 - Also providing backup routes since years
 - T2 in the same region
 - T2 and T2/T1/T0 in different regions
- Set of fast network hubs connecting many T2 to many T1 and T0
 - LHCONE = LHC Open Network Environment
- T1 and T2 sites keep their different responsibilities



Locations of the WLCG sites





WLCG in action

Conclusions

- Grids facilitate collaboration between members of supported distributed communities
- Grids allow distributed resources to be shared uniformly and securely for common goals
- Grids may have complex infrastructures
- Grids are useful for many scientific disciplines and projects

The Worldwide LHC Computing Grid is vital for the success of the LHC experiments !

More information



wlcg.web.cern.ch



www.egi.eu



www.opensciencegrid.org



www.ndgf.org