

WLCG Strategy & Computing Model Evolution

Andreas Petzold

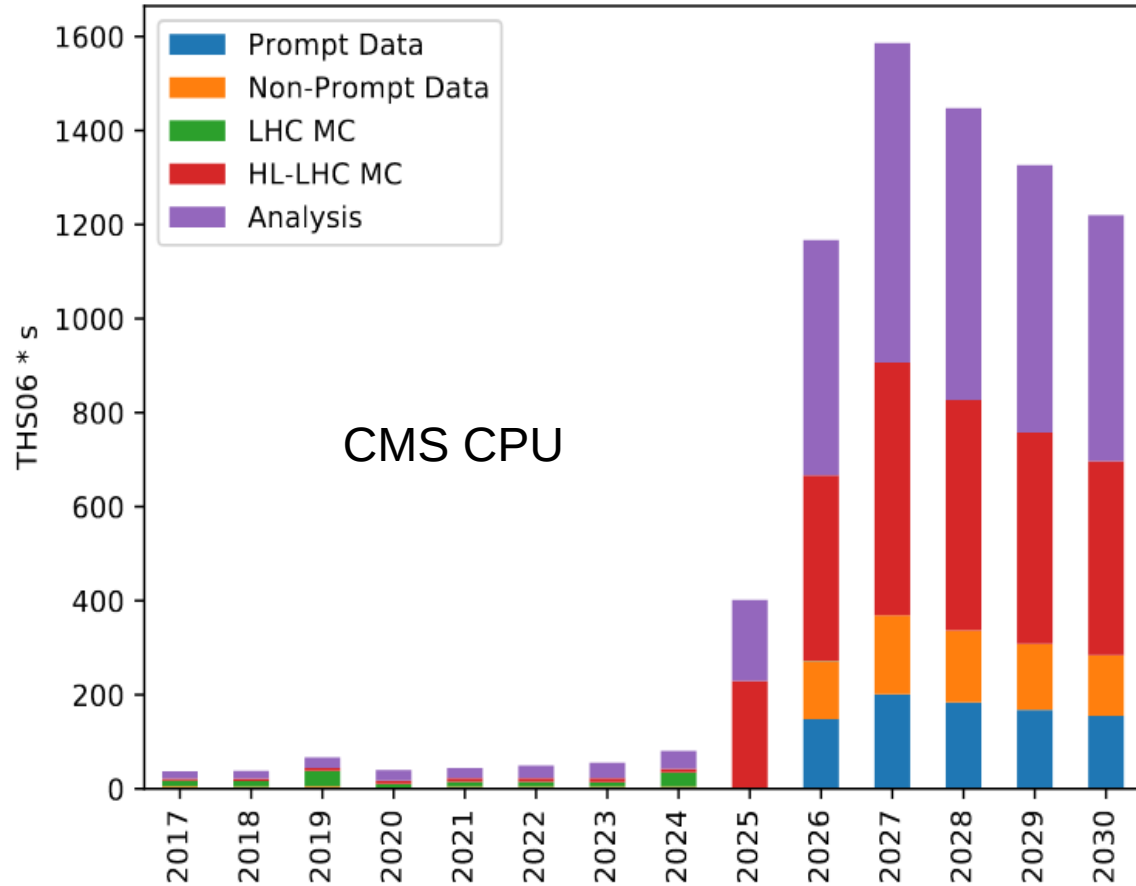
STEINBUCH CENTRE FOR COMPUTING - SCC



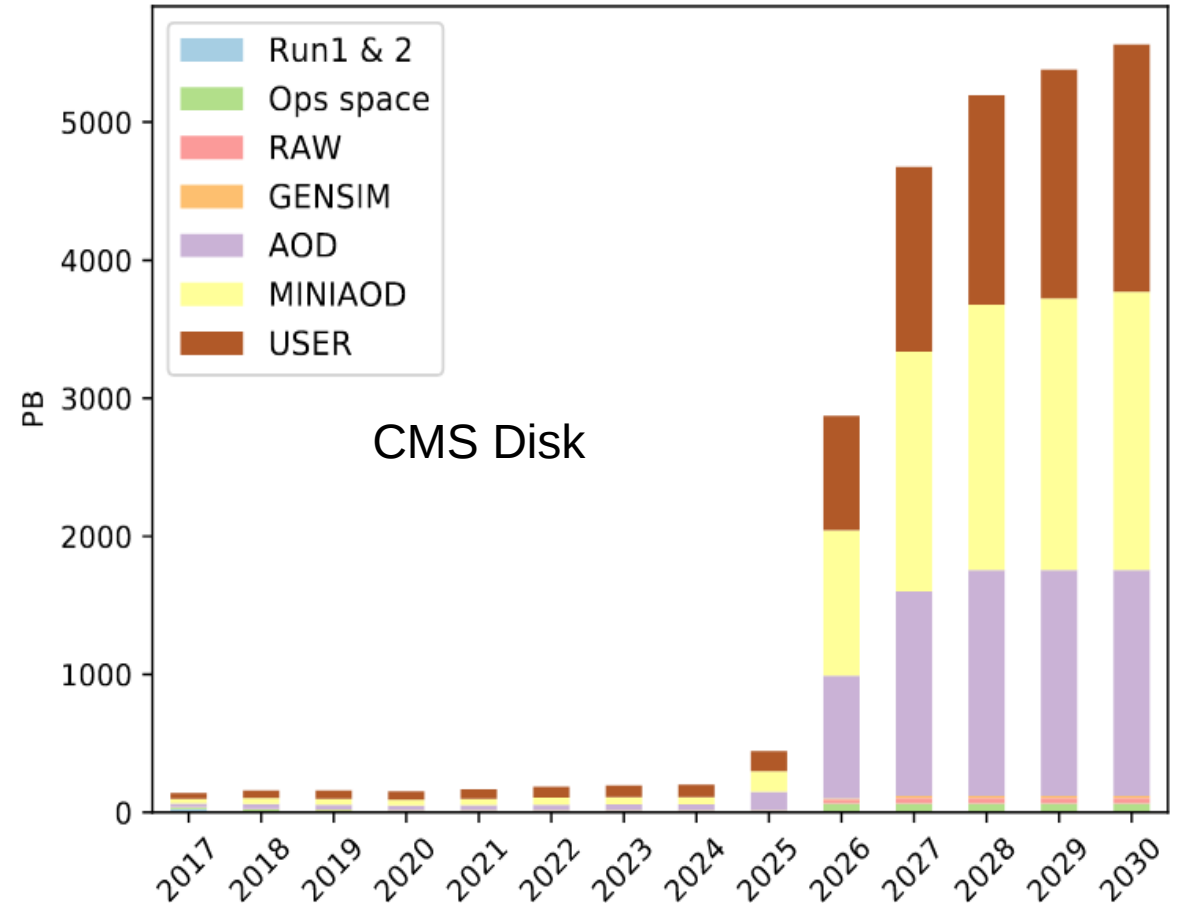
Computing Challenges in HEP

HEP Software Foundation Community
 White Paper [HSF-CWP-2017-01]

CPU seconds by Type

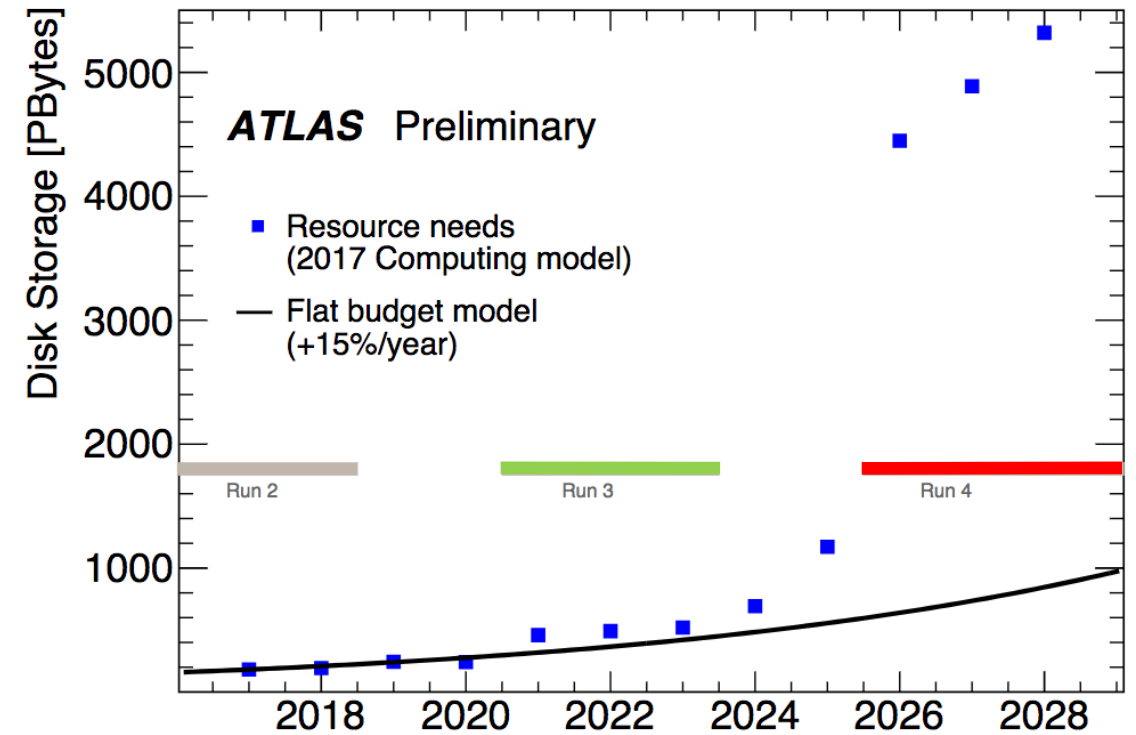
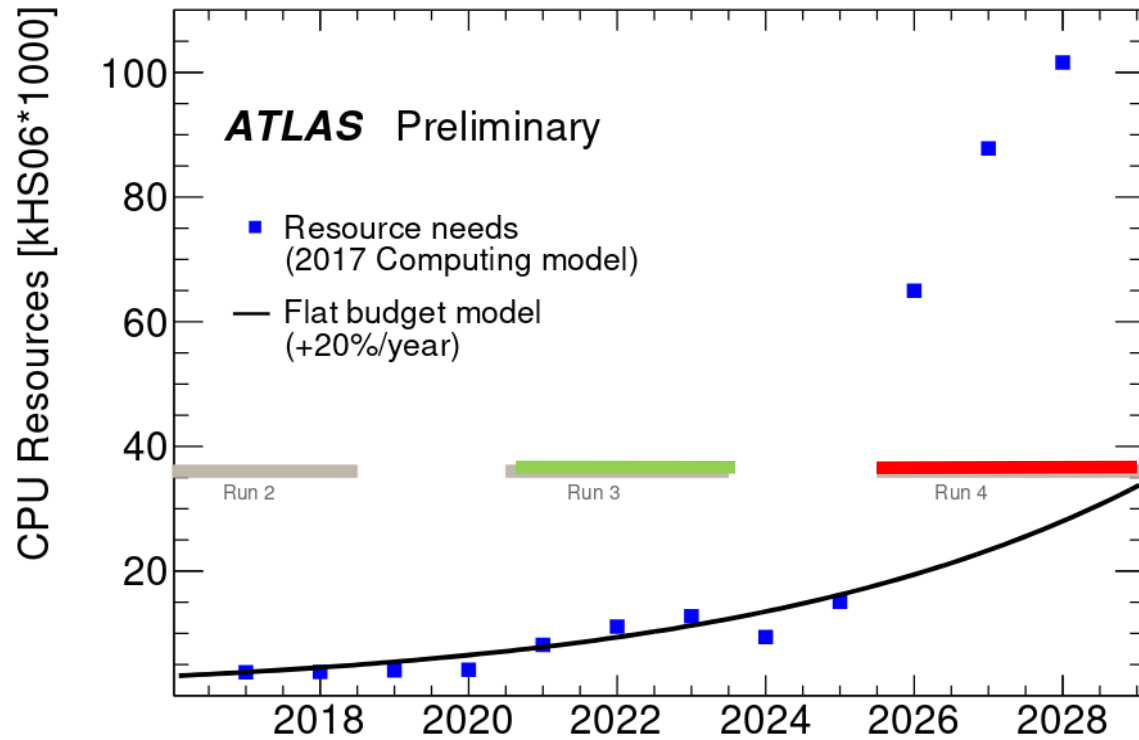


Data on disk by tier



Computing Challenges in HEP

HEP Software Foundation Community
 White Paper [HSF-CWP-2017-01]



Computing Challenges in HEP

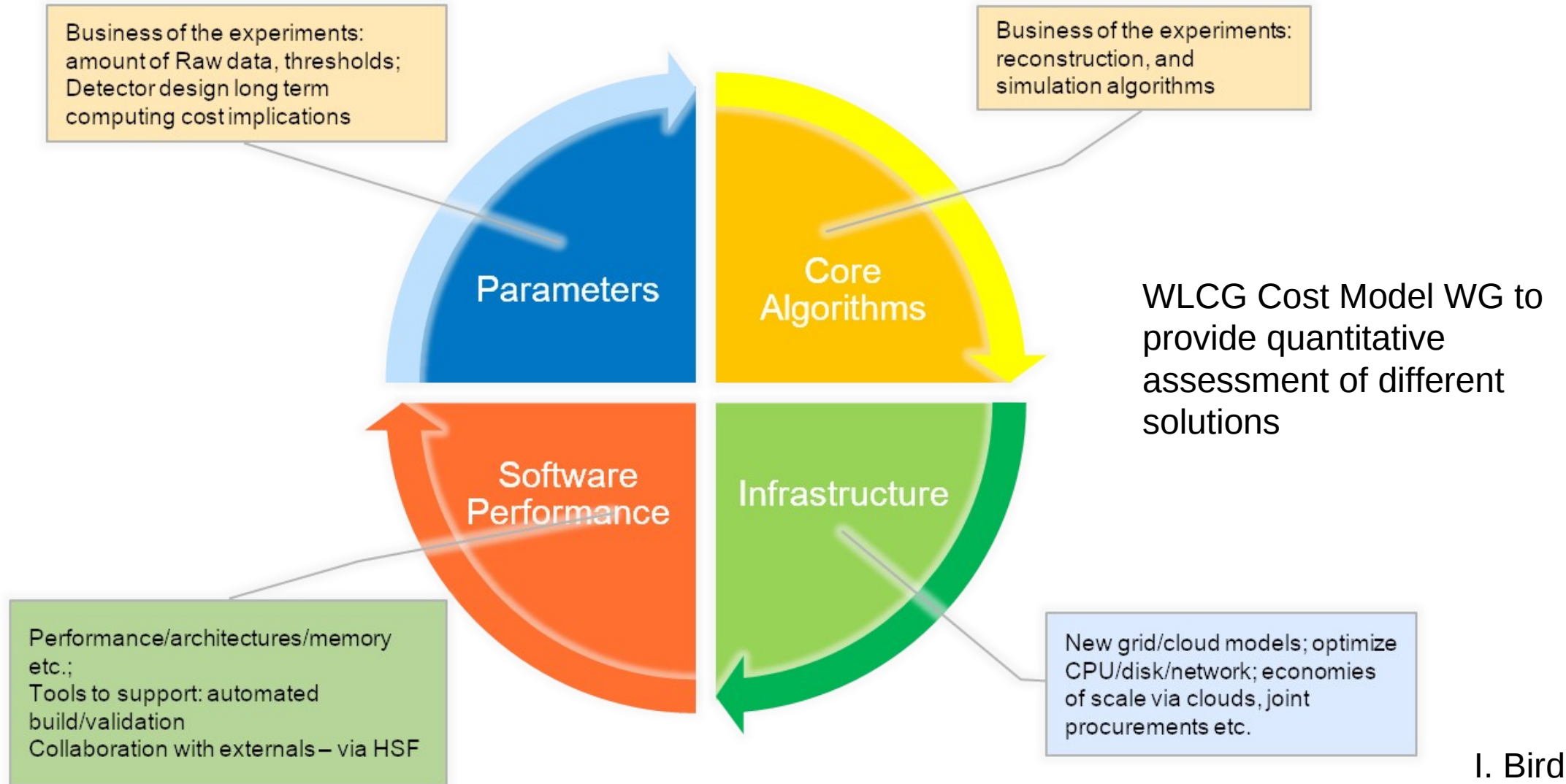
- Gap between projected requirements vs. resource increase achieved by flat budget + technology evolution
 - Run 3: factor 1.3 – 1.5 shortfall
 - Run 4: factor 7 – 10 shortfall
- Storage is the main cost driver
- Required Changes

Evolution for Run 3



Revolution for Run 4

HL-LHC Computing Cost Drivers



I. Bird

WLCG Strategy Document

- Select aspects of summary given by Simone Campana to the WLCG OB

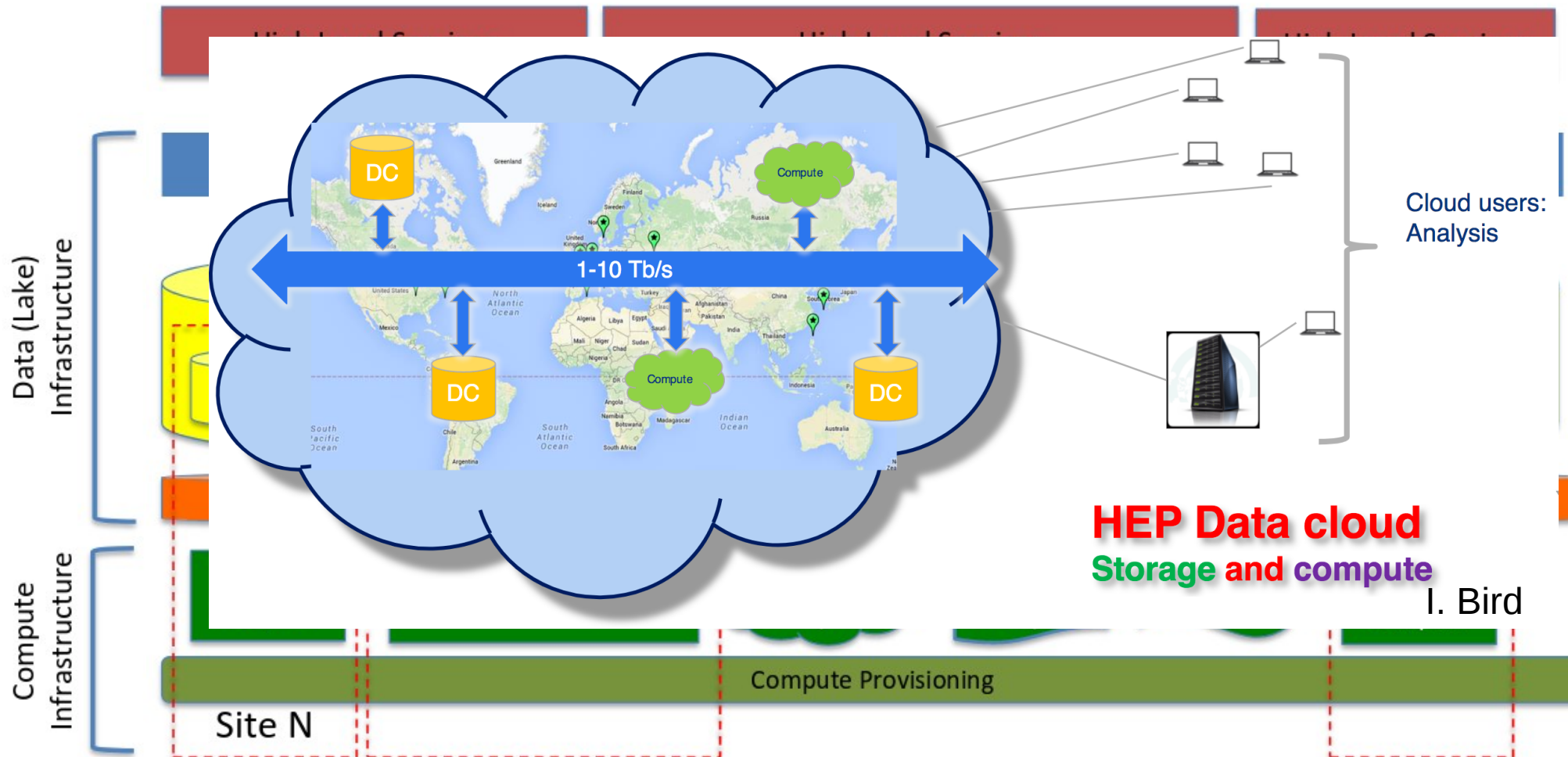
The WLCG Strategy Document

- The HL-LHC computing challenge: provide the computing capacity needed for the LHC physics program, managing the cost
- The WLCG strategy document is a specific view of the CWP, prioritizing R&Ds relevant to the HL-LHC computing challenge
- The prototyped solutions will be the foundation of the WLCG TDR for HL-LHC, planned for 2020. Timing to be re-considered?
- This is a presentation of the content of the strategy document
 - <http://cern.ch/go/Tg79>

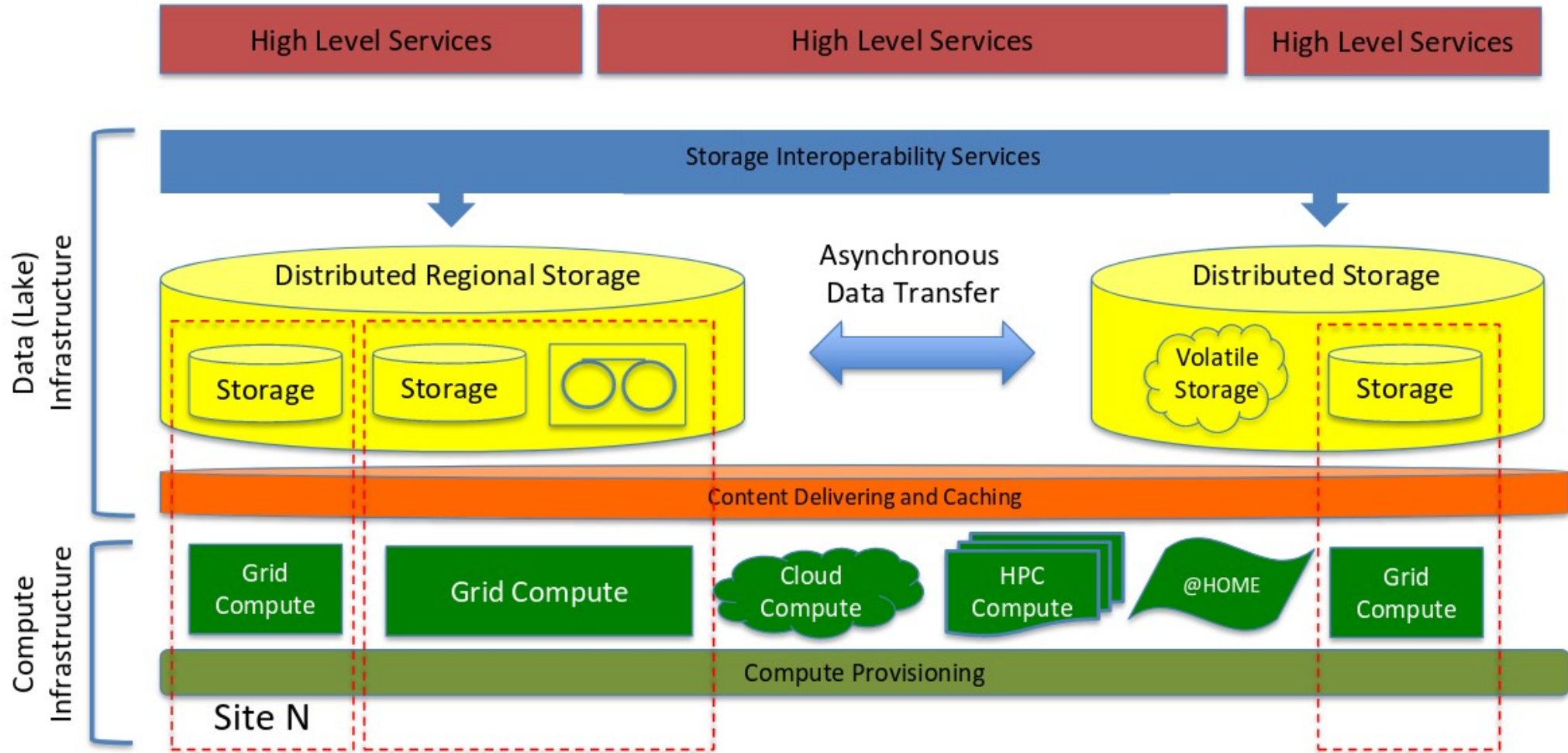
Computing Models

- Understand the HL-LHC running conditions and the input parameters arising from them: trigger rates, # Monte Carlo, seconds of data taking, ..
- Pursue aggressively the reduction of data formats
 - Compression
 - Tiering (AOD->MiniAOD->NanoAOD) or slimming the derived formats (e.g. DAODs from trains)
- Rely on less expensive media (e.g. TAPE) for a larger set of formats (LHC data is generally “cold”).
 - Implies evolution of the facilities and the workflow and data management systems (see later)
- Review the centralized processing and the analysis models. Shift more workload in the direction of organized production.

Data and Compute Infrastructures (I)



Data and Compute Infrastructures (I)



Data and Compute Infrastructure (V)

■ Networks

- Investigate the possibility to use different protocols than TCP for WAN transfers
- Investigate the benefits and deployment model of SDNs in a data lake infrastructure and in any case for a distributed storage
- Evolve the network monitoring system to collect and expose information to be consumed and used for adaptive network usage
- Understand the possibility to deploy a caching layer built into the network (NREN exchange points)
- Study how to attach commercial cloud resources to the WLCG network infrastructure achieving the needed performance

Data and Compute Infrastructure (VI)

- **Compute Resources**
 - Understand how to provide a common provisioning layer for heterogeneous resources
 - Understand how brokering will maximize the probability of CPU-data co-location. Understand cache aware brokering
 - Integrate more workflows on HPC facilities and leverage different hardware types

- **Cloud analysis model**
 - Prototype a quasi interactive analysis facility offering the capability to scale out in a cloud backend (e.g. understand how SWAN based analysis fits the data lake model)

Interoperability and Data Preservation

- Review the security model and evolve it toward federated identities. Move away from X509, prototype a token-based solution ensuring interoperability and sustainability
- Favor common solutions across all the stack (from high level services to infrastructure). A very strong message in this direction from all funding agencies: little or no support in the future for experiment specific solutions.
- The previous bullet is the basis for a data and analysis preservation strategy.

Conclusions

- We will face two main challenges for HL-LHC computing: data management and offline software (algorithms, frameworks, I/O, data models)
- We launched a WLCG DOMA (Data Organization, Management, Access) project to address the first
 - Create a forum to discuss ideas and present results
 - Track progress, review status, evolve the strategy, prioritize
 - Discussed in the WLCG/HSF workshop, agreed in WLCG management and Grid Deployment Board, kick off on June 4th and 5th
- We are discussing an equivalent initiative for Offline Software and the HSF is the natural umbrella for this
- Manpower is very short, particularly in software, also for the current core tasks
 - Need to leverage commonality at all costs, this is why the HSF plays a huge role here
 - Need to address the problem of recognition and career opportunity for people working on software related tasks