Software & Computing status

L. Poggioli, LAL

- End of Run-2
- Towards Run-3
- HL-LHC

Since last LCG-FR: Short period

'Tu trouves pas que c'est un peu rapproché ?'

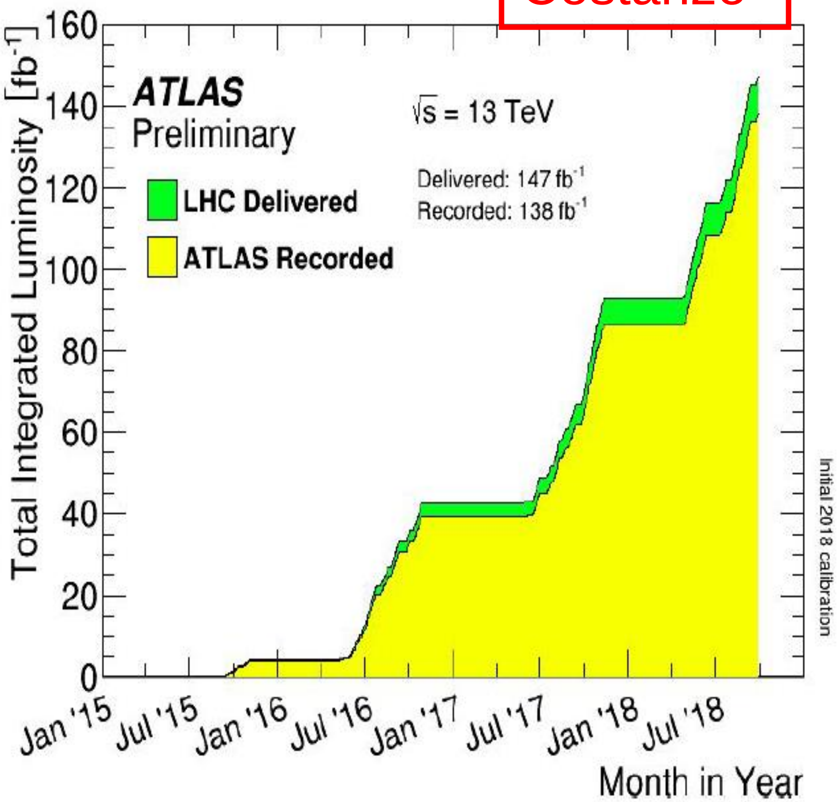
But big progress!

End of Run-2

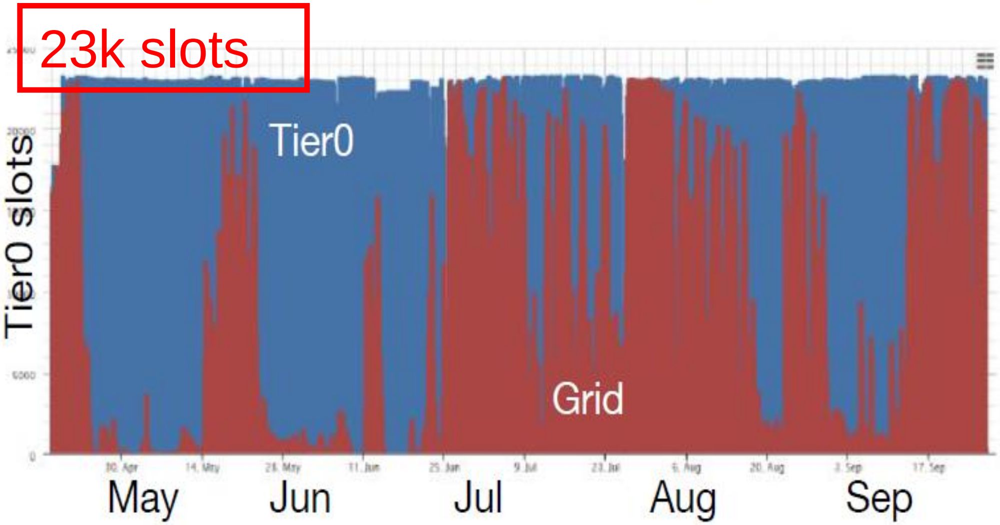
Run-2: A lot of data!!

Costanzo

- A large dataset, and also a lot of MC
- S&C smooth operations in 2018
  - ○ Data and MC needed for Physics analysis ready ahead of time
  - ○ Smooth operations at Tier0
- Stable release 21 was key to this
- Heavy Ion run ahead of us, still!



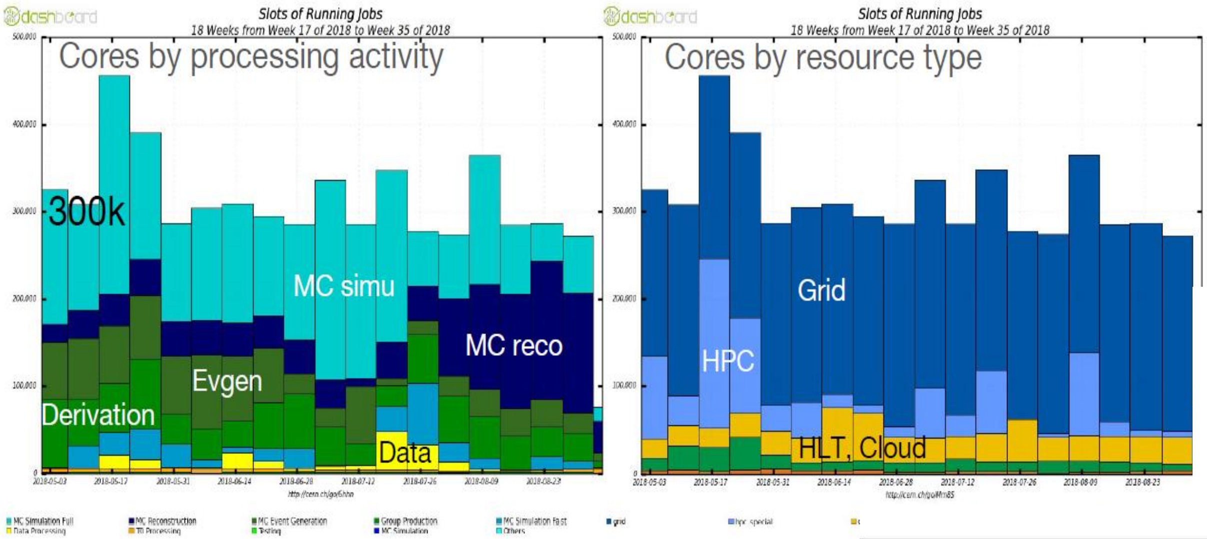
23k slots



- Collected, (re)processed 22pB raw & 22B evts for analysis
- Same amount of MC evts

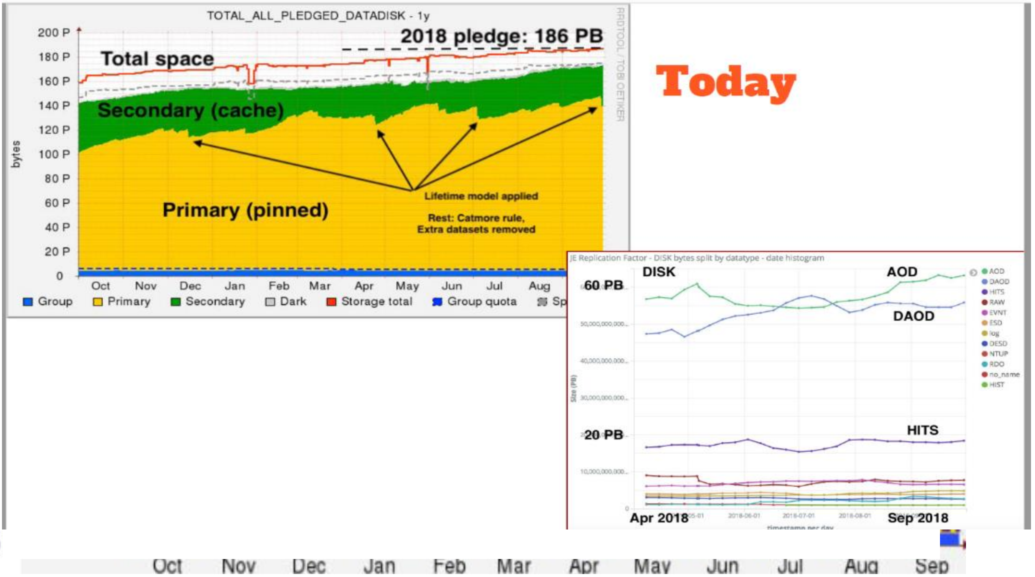
CPU & Disk usage

Costanzo, Girolamo



- Disk remains tight
- Dominated by AOD & dAOD

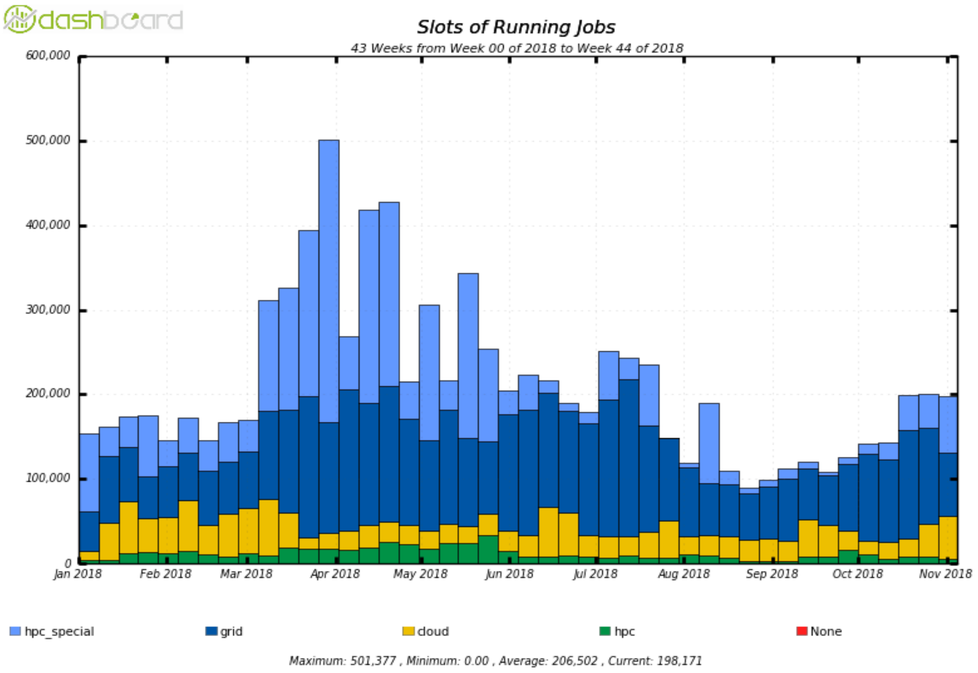
- Smooth Tier0 running on 23k cores
  - Bphysics stream spillover submitted
- Production on more than 300k cores
- Exhausted HPC allocations.
  - Waiting for more
- Move >1PB, >20GB/s, 1.5-2M files/day



Davide Costanzo IO review introductio

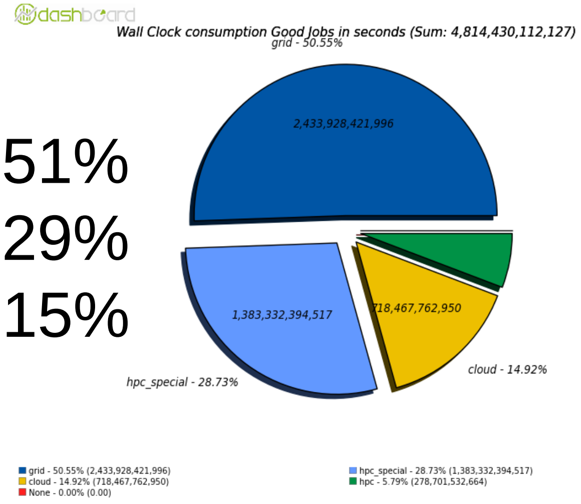
Non-grid resource (MC sim only)

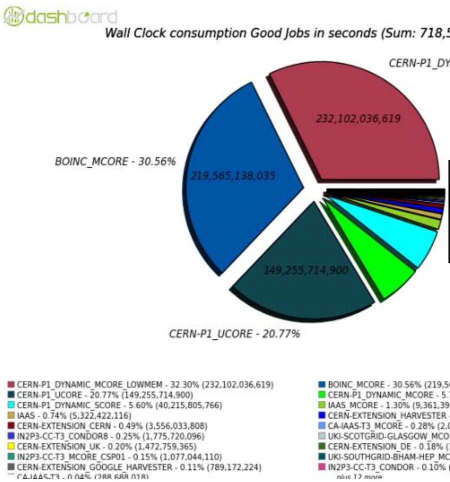
Running slots (January->Now)



CPU (Jan->Now)

GRID	51%
HPC	29%
CLOUD	15%





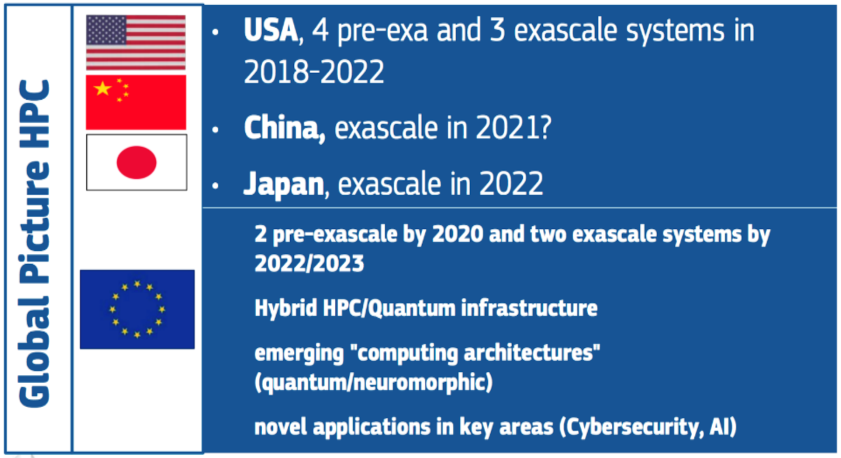
Cloud

- US-HPC: Lower now
  - Allowed slots exhausted
- CLOUD

•HLT 64%, **BOINC** 31%!!

Looking for extra-resources

- CLOUDS (commercial)
  - Google 'Data Ocean': GCE integration with Panda, GCS interface to Rucio, first 'solid' cost discussions
  - HNSciCloud project ongoing
- BOINC
  - Volunteer base (@home) stable but not really increasing
  - Recent increase in resources mainly from spare CERN machines and grid backfilling (up to 25/% for a site)
- HPC
  - Exascale ~ 2022 in US, China Japan, Europe (EUHPC)
  - Architecture moving -> GPU
    - ATLAS today: No GPU software
    - -> Start looking in GPU & ML applications



Global Picture HPC

- **USA**, 4 pre-exa and 3 exascale systems in 2018-2022
- **China,** exascale in 2021?
- **Japan**, exascale in 2022

2 pre-exascale by 2020 and two exascale systems by 2022/2023

Hybrid HPC/Quantum infrastructure

emerging "computing architectures" (quantum/neuromorphic)

novel applications in key areas (Cybersecurity, AI)

Activities (1)

- Software
  - AthenaMT, manpower in better shape
  - ACTS (Stand alone tracking library) **manpower!!**
  - FastSim
- DDM
  - RUCIO adopted by CMS!!
  - Rucio mover unified way to interact with data
  - Protocols progress xRootD, WebdaV
  - Caches progress: Xcache (XrootD)
    - XCache dedicated XrootD server. User access cached data thru XCache server from upstream XrootD server
  - Towards Run4: Tape carousel, DOMA, QOS

Activities (2)

- WFM
  - Pile-up premixing (à la CMS) and overlay
  - Harvester sw to interface various platforms
    - HPC, Grid,...
  - Event Service also for sites
  - Global shares (UCORE queues)
    - Unified score/mcore queues to better handle EVTGEN
  - R&D project (eg with Google)
  - ATLAS@home
  - Data carousel mode of operation with tapes
    - R&D ongoing to use tapes more efficiently, eg producing directly Derivations from AOD on tape

Activities (3)

- Databases
  - Condition DB: Prepare migration from COOL to REST (Representation State Transfer) -> CREST for Run-3
  - Frontier Analytics progress
    - To understand bottlenecks of the overlay production on the grid (squid-Frontier caching)
  - Essential for efficient pileup treatment
- Monitoring
  - Progress using Kibana Elastic search
  - Unified CERN tools, eg GRAFANA

For sites (1)

- Unified Queues (score/mcore->UCORE)
  - Brand new way to submit jobs: ATLAS controls its internal priorities to run (eg EVTGEN)
  - Only 1 queue & submit w/ job params from scouts
  - ALL French sites have now UCORE queues
- Harvester
  - Unified way to submit jobs wrt resources: Grid, Cloud, HPC
  - Ongoing migration for grid from to Harvester
  - Requires Unified queues
  - Migration a priori transparent for sites

For sites (2)

- FAX decommissioned
    - Sites still required to provide xrootd access to storage, BUT no need to have it federated
- DOMA TPC (3rd party copy)
    - Need alternative to gridftp: http/xrootD
- CentOS7
    - No deadlines for migration until early 2019
    - Sites encouraged to upgrade earlier if they can
        - Containers better supported
        - Native CentOS7 releases are now being built and will not run on SL6 nodes
    - Singularity is a requirement for upgrading sites

Lines of effort: Summary

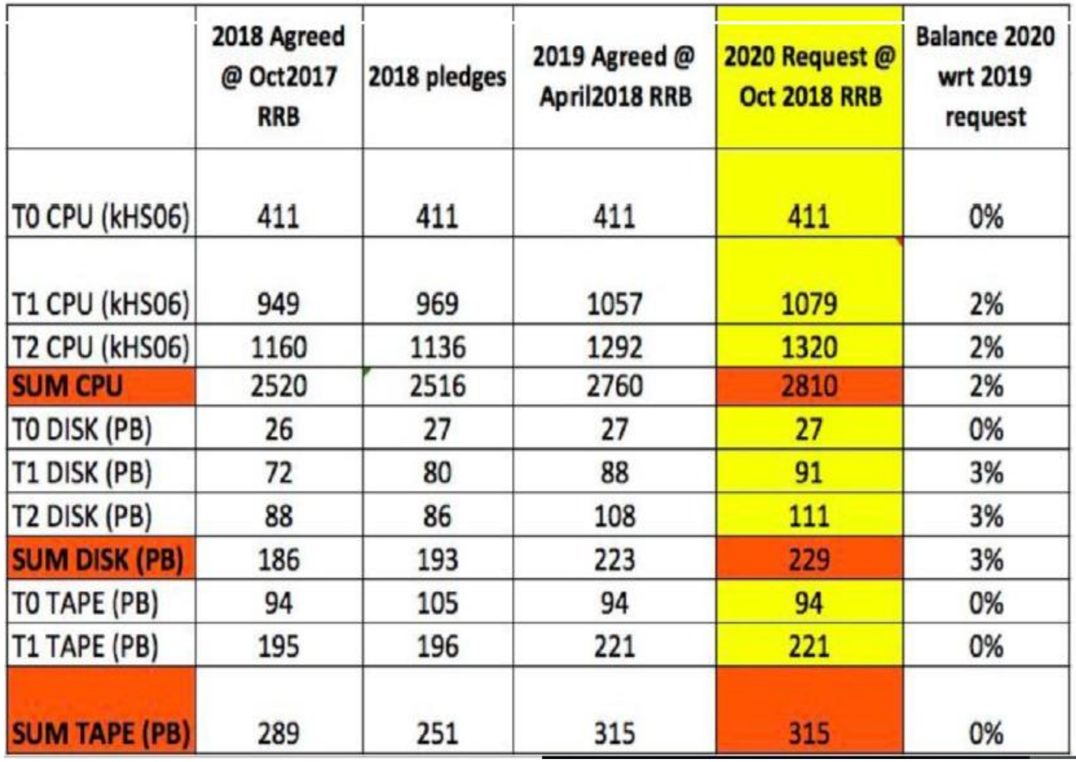
- Software
  - Leverage additional resources (HPC, Boinc, …)
  - Improve software and efficiency (SPOT group)
  - Run less full-simulation (and more fast sim)
  - Promote support for software development
- Workflow
  - T1s continue to exercise and improve perf. of dAOD production from tape inputs
  - Harvester, Event service (ES), Overlay (pileup handling),
  - New: Event Streaming service (ESS)
    - What ES is to computing, ESS is to input data transfer
- Computing Model
  - Nucleus/satellites model
  - T2/T3 consolidation. Check pledges deployment

Preparing Run-3

CRSG

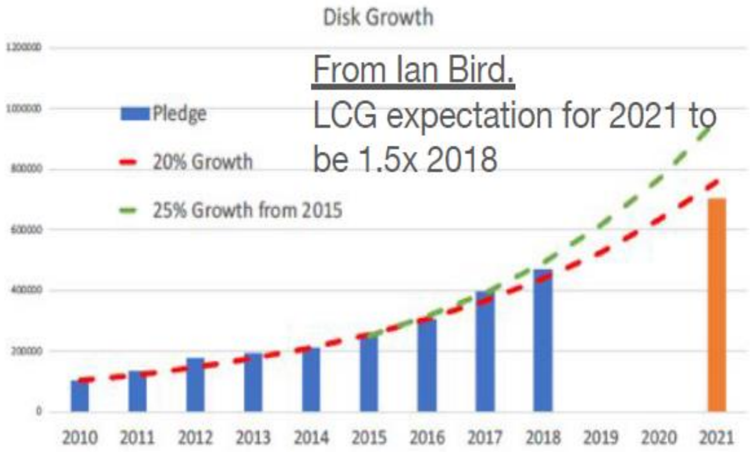
Costanzo

- Report submitted to referees over the summer
  - Resources in 2020 expected to be at the same level as 2019
  - See presentation at ATLAS weekly
- Received a few mild comments
  - We need to prepare for the 2021 request. Large uncertainties from LHC and lumi
  - We need to reduce disk usage (Analysis Model Study Group for Run-3)



	2018 Agreed @ Oct2017 RRB	2018 pledges	2019 Agreed @ April2018 RRB	2020 Request @ Oct 2018 RRB	Balance 2020 wrt 2019 request
T0 CPU (kHS06)	411	411	411	411	0%
T1 CPU (kHS06)	949	969	1057	1079	2%
T2 CPU (kHS06)	1160	1136	1292	1320	2%
SUM CPU	2520	2516	2760	2810	2%
T0 DISK (PB)	26	27	27	27	0%
T1 DISK (PB)	72	80	88	91	3%
T2 DISK (PB)	88	86	108	111	3%
SUM DISK (PB)	186	193	223	229	3%
T0 TAPE (PB)	94	105	94	94	0%
T1 TAPE (PB)	195	196	221	221	0%
SUM TAPE (PB)	289	251	315	315	0%

Target an increase such as R(2021)/R(2018) = 1.5



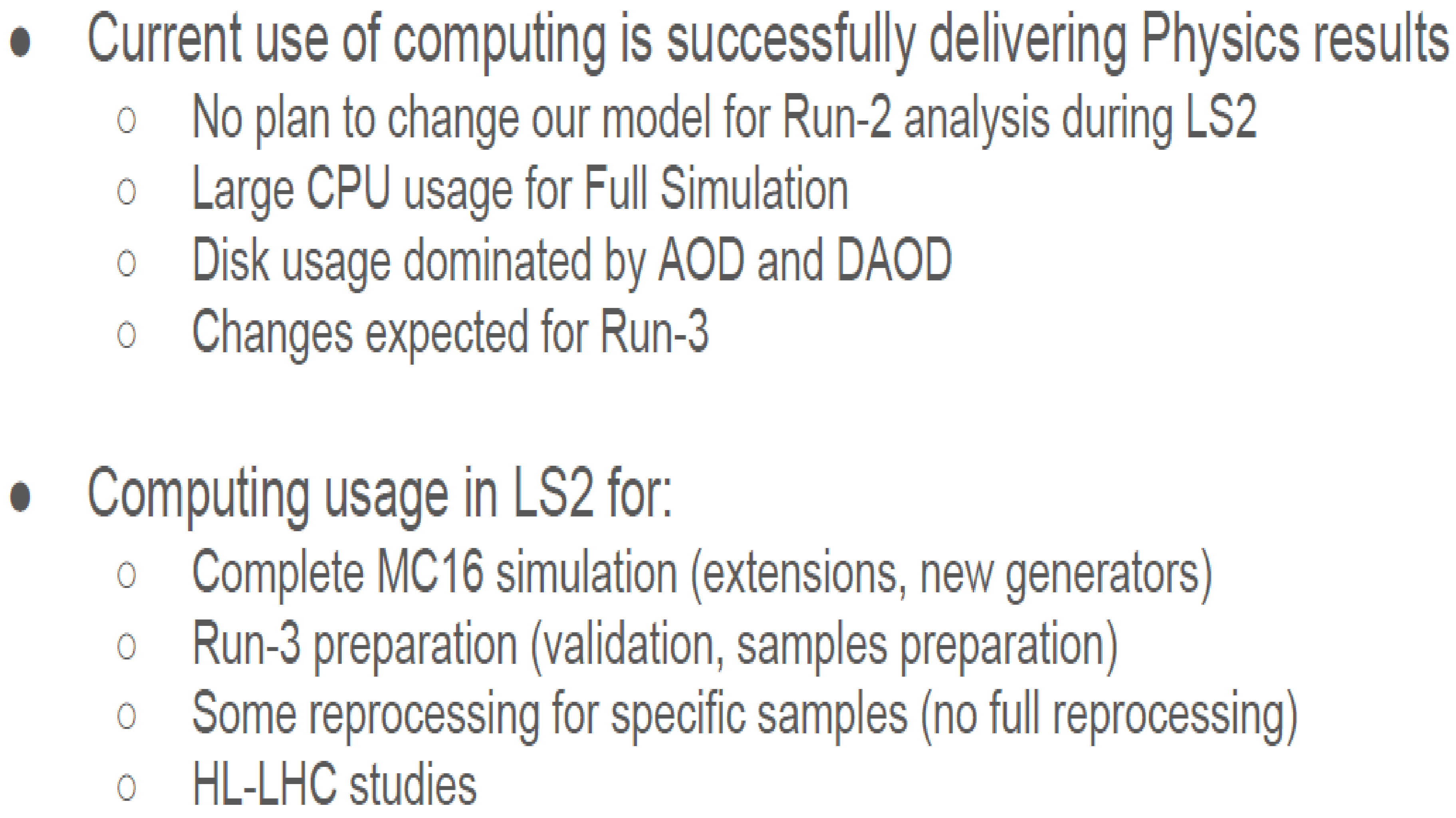
From Ian Bird.
LCG expectation for 2021 to be 1.5x 2018

27-Sep-2018

7

Computing Model during LS2

Costanzo

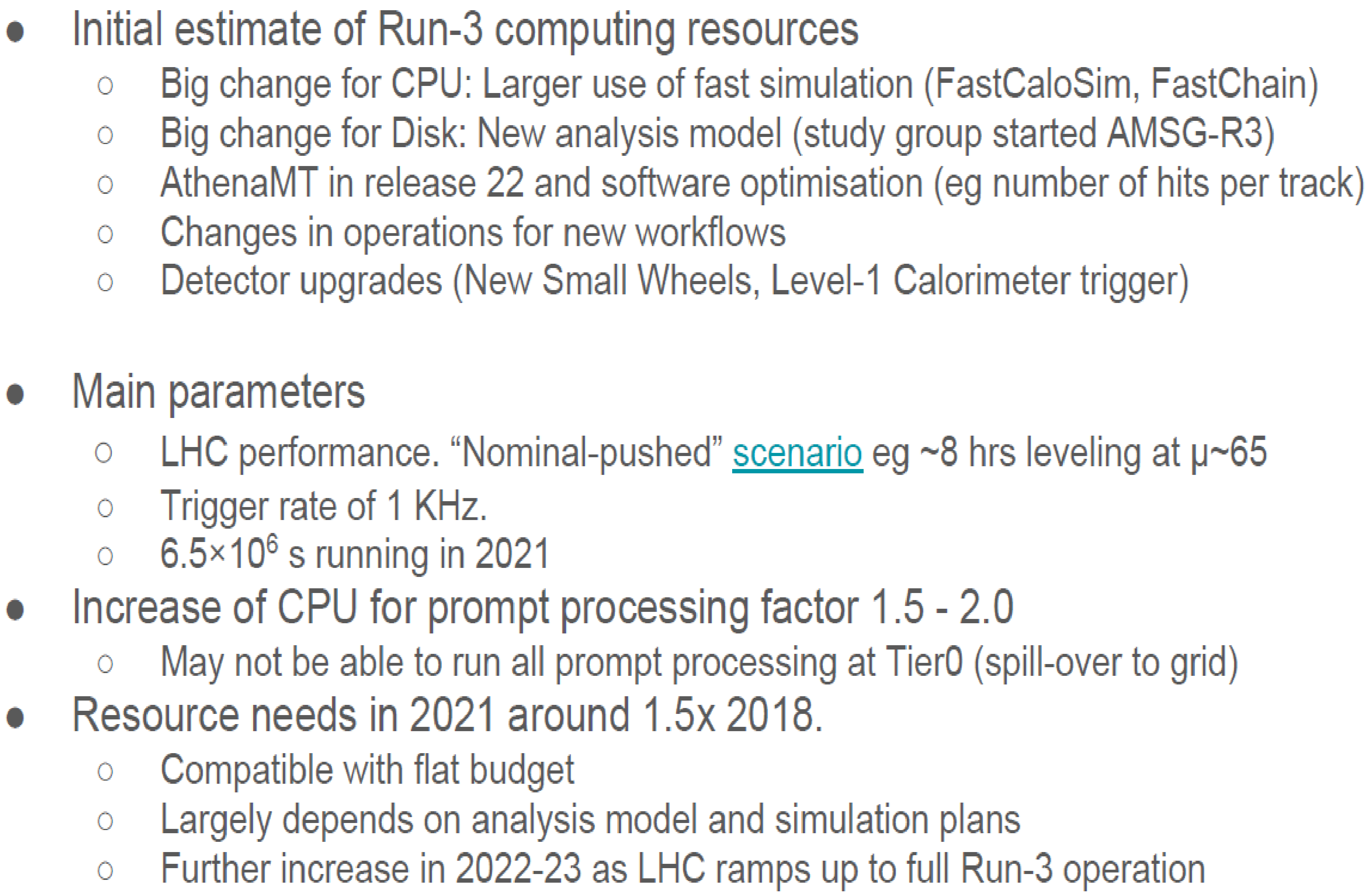


- Current use of computing is successfully delivering Physics results
  - ○ No plan to change our model for Run-2 analysis during LS2
  - ○ Large CPU usage for Full Simulation
  - ○ Disk usage dominated by AOD and DAOD
  - ○ Changes expected for Run-3

- Computing usage in LS2 for:
  - ○ Complete MC16 simulation (extensions, new generators)
  - ○ Run-3 preparation (validation, samples preparation)
  - ○ Some reprocessing for specific samples (no full reprocessing)
  - ○ HL-LHC studies

Initial plans for Run-3

Costanzo



- Initial estimate of Run-3 computing resources
  - Big change for CPU: Larger use of fast simulation (FastCaloSim, FastChain)
  - Big change for Disk: New analysis model (study group started AMSG-R3)
  - AthenaMT in release 22 and software optimisation (eg number of hits per track)
  - Changes in operations for new workflows
  - Detector upgrades (New Small Wheels, Level-1 Calorimeter trigger)

- Main parameters
  - LHC performance. “Nominal-pushed” scenario eg ~8 hrs leveling at μ~65
  - Trigger rate of 1 KHz.
  - 6.5×10^6 s running in 2021
- Increase of CPU for prompt processing factor 1.5 - 2.0
  - May not be able to run all prompt processing at Tier0 (spill-over to grid)
- Resource needs in 2021 around 1.5x 2018.
  - Compatible with flat budget
  - Largely depends on analysis model and simulation plans
  - Further increase in 2022-23 as LHC ramps up to full Run-3 operation

Software for Run-3

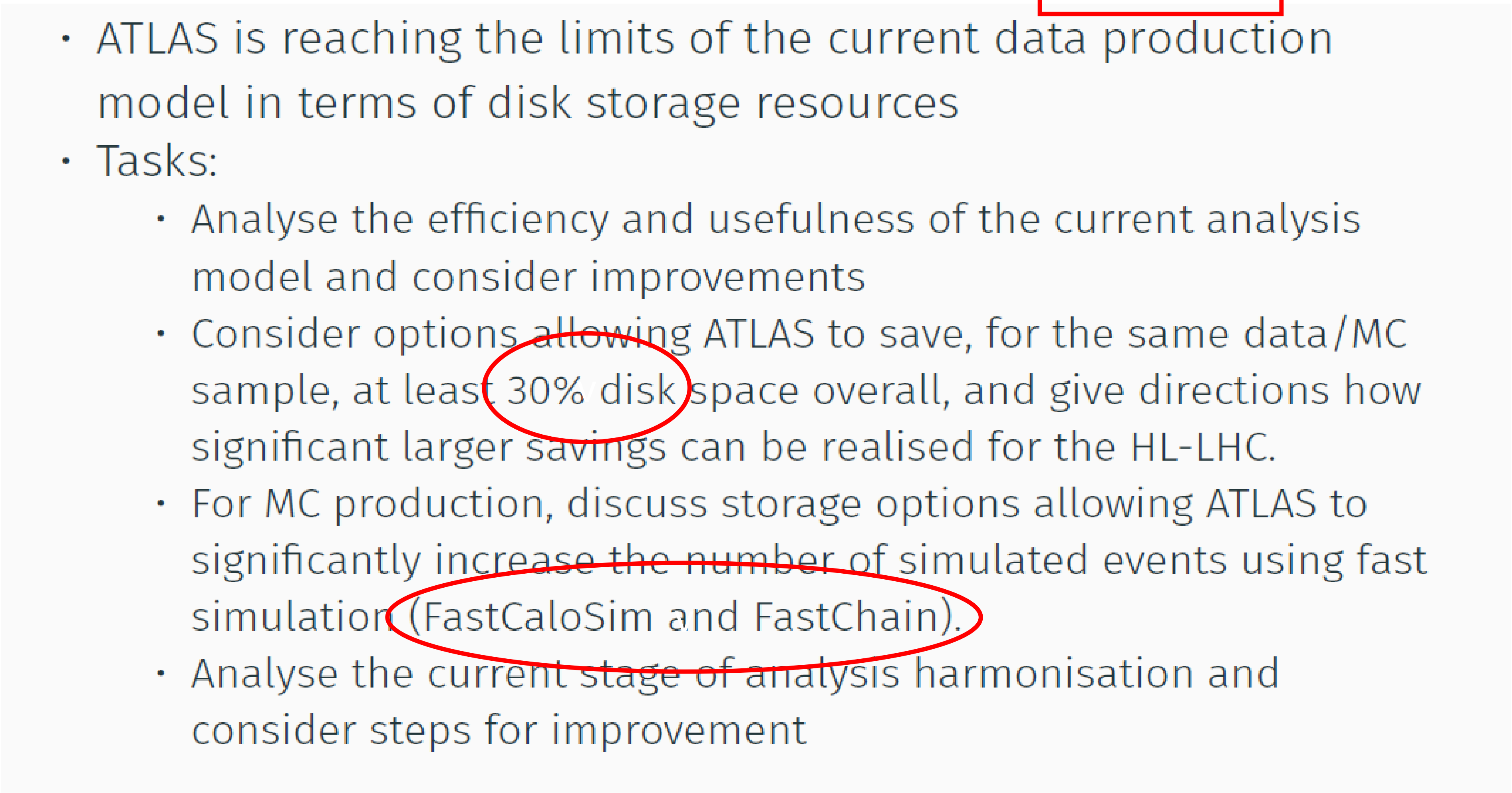
- AthenaMT: Move towards a multithreaded framework to use modern architectures
- FastCaloSim: High priority for ATLAS
- Add new detectors to simulation and reconstruction (NSW)
- ACTS (A Common Tracking Software) for tracking. Streamlined ATLAS software, MT by construction. Recommendation to use some ACTS at end of June
- Lack of developers ~3FTEs missing

Analysis for Run-3

- Run 2 model very successful
  - Many derived AOD (DAOD) formats O(100)
  - AOD use 55 PB of disk / DAOD use 52 PB of disk
- Focus on AOD & dAOD
  - Reduced overall size
  - #versions used
  - Smaller evt sizes?
- Scrutiny group at last RRB
  - ATLAS uses more disk than CMS. Difference is growing
  - Encouraged to look into smaller data formats
- -> Analysis Model Study Group for Run-3
  - Run-3: More MC (FastSim), Bigger evts (μ), Same #data

AMSG-R3 working group

Elmheuser



- ATLAS is reaching the limits of the current data production model in terms of disk storage resources
- Tasks:
  - Analyse the efficiency and usefulness of the current analysis model and consider improvements
  - Consider options allowing ATLAS to save, for the same data/MC sample, at least 30% disk space overall, and give directions how significant larger savings can be realised for the HL-LHC.
  - For MC production, discuss storage options allowing ATLAS to significantly increase the number of simulated events using fast simulation (FastCaloSim and FastChain).
  - Analyse the current stage of analysis harmonisation and consider steps for improvement

- ESSENTIAL: Gathers input for physics & performance groups
- https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/AnalysisModelStudyGroupRun3

Towards HL-LHC

Towards HL-LHC: Challenges

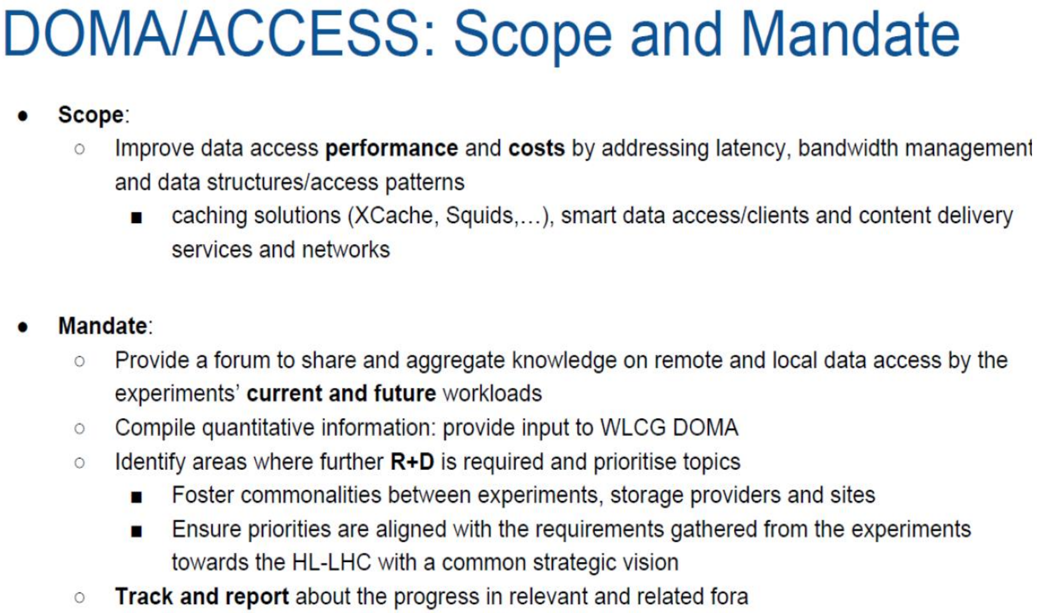
- Inputs
  - Trigger rate 10kHz. Increase total evt numbers
  - <μ> ~200. Increase in CPU & storage needs
- Today
  - X 3 missing in CPU. Seems doable (many ideas)
    - R&D inside HSF, Accelerators (GPU, FPGA), Extra-resources (HPC, R&D with Google,…)
    - FastSim, Detecor layout, Machine Learning
  - X 7 missing in storage More critical
- R&D areas
  - DOMA, Software upgrade, HSF technical forum

Possible gains for storage

- Disk usage today: ⅓ AOD, ⅓ dAOD
- -> Extend tape carousel to (d)AOD
  - But Tape means delay, and (d)AOD workflows time critical & very complex
  - But Tape is limited at T1s, while processing resources much more widely distributed
- Also Possible:
  - Make AODs 10x smaller à la CMS
  - Streamline some physics analyses
  - Limitation of # replicas
    - >=1 replicas on disk today, -> dynamic, managed availa'ty of actively used data via Data Lake, replica count <<1

DOMA/ACCESS

Stéphane, October S&C
See L. Duflot’s talk



DOMA/ACCESS: Scope and Mandate

- **Scope**:
  - Improve data access **performance** and **costs** by addressing latency, bandwidth management and data structures/access patterns
    - caching solutions (XCache, Squids,...), smart data access/clients and content delivery services and networks
- **Mandate**:
  - Provide a forum to share and aggregate knowledge on remote and local data access by the experiments’ **current and future** workloads
  - Compile quantitative information: provide input to WLCG DOMA
  - Identify areas where further **R+D** is required and prioritise topics
    - Foster commonalities between experiments, storage providers and sites
    - Ensure priorities are aligned with the requirements gathered from the experiments towards the HL-LHC with a common strategic vision
  - **Track and report** about the progress in relevant and related fora

Activities

- Call for projects over summer

→ googledoc created to collect informations

→ Currently 15 projects (some with subprojects)

- – Displays interest of computing teams
  - Not known by community even if presented in conferences
- – Many contributions from ATLAS members or associated sites
  - Not all ATLAS activities

- 3 main topics
  - – Deploy new setup and measure performances within experiment workflow
    - ‘Caching’ is current hot topic (Ilija’s talk)
  - – Study and measure workflow to estimate gain with new setup
  - – Development of generic tools for bandwidth management and caching simulations
    - – Much better position than CMS
    - – Non ATLAS teams present report on our workflow and make recommendations

- Conveners:
  - Stéphane, I. Vukotic
- Discussed extensively at this workshop

Summary

- ATLAS S&C is in good shape
  - Now able to focus on refinements, performance, and look to future with R&D
- ATLAS is front and center in common R&D (inside HSF community)
- Run-3 a priori OK within flat budget. Key issue is software: AthenaMT & FastSim
- HL-LHC
  - Trend lines are good in CPU (constant progress)
  - Plans in storage to be quantified (today critical)
  - R&D, DOMA, very active and growing