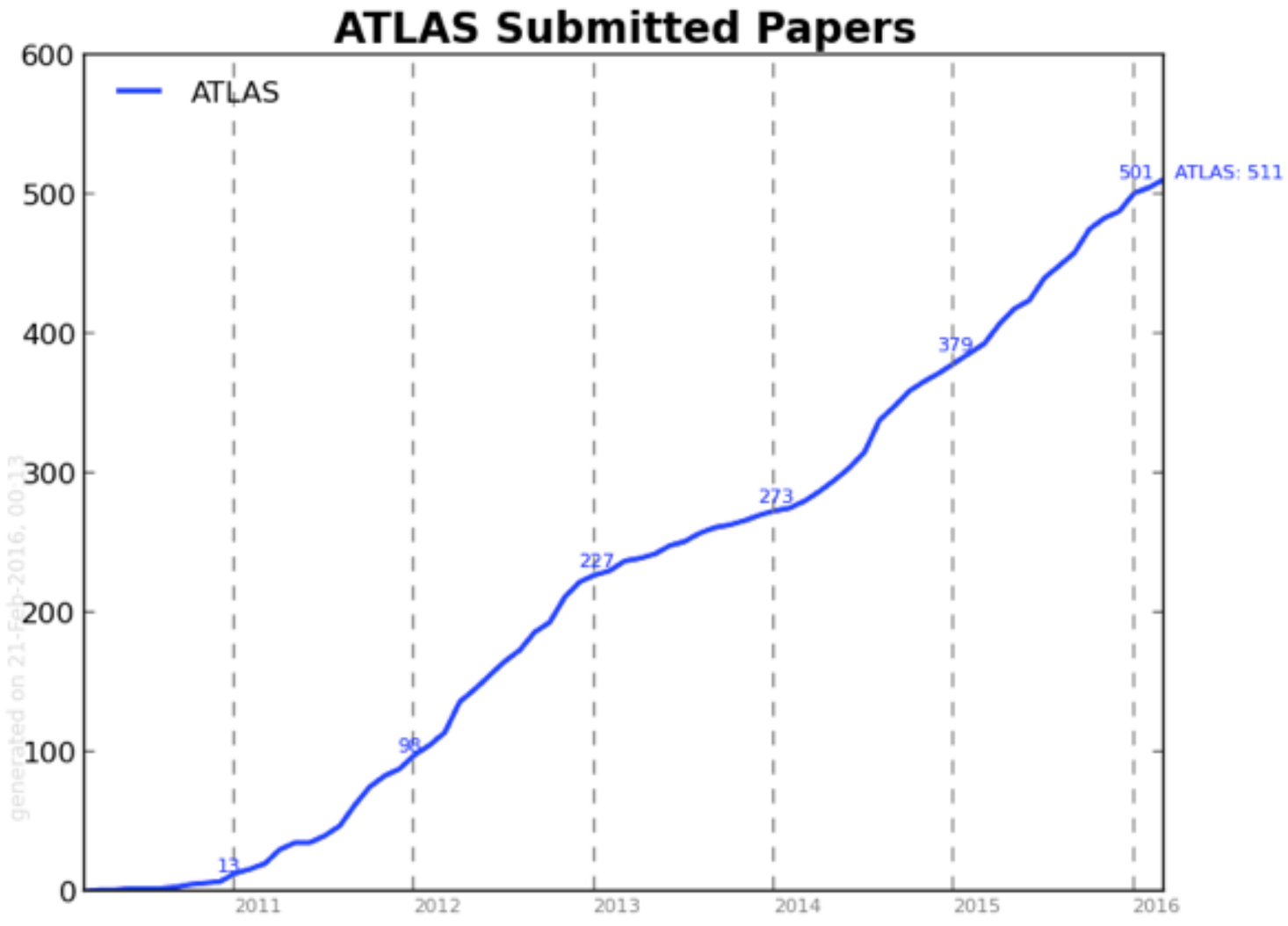
ATLAS computing report

Eric Lançon, Simone Campana
LHCC , March 1st 2016

Publications

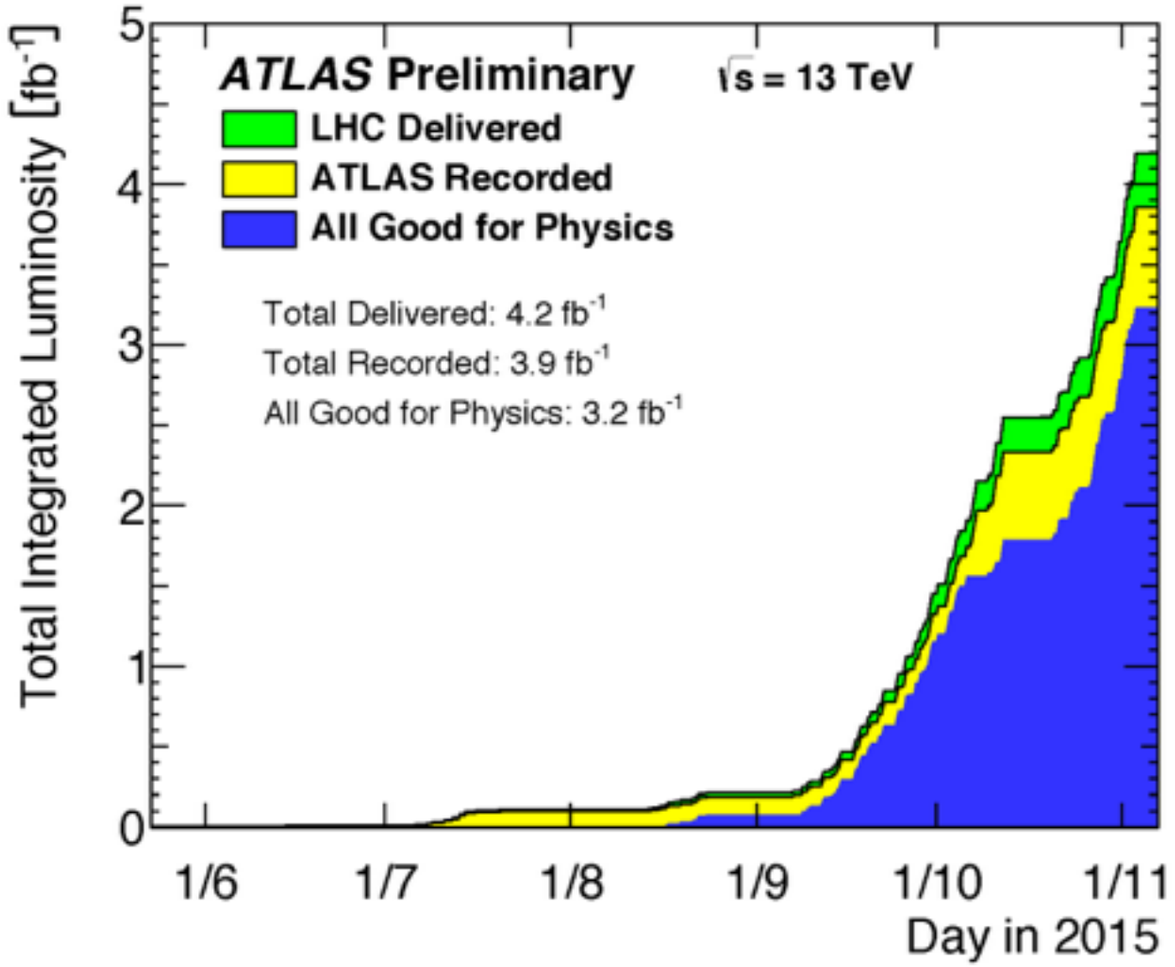
- 2015 finished with **500** ATLAS publications
  - 122 in 2015
- Still ~**50** Run 1 publications in progress
- Already from 2015 data :
  - 7 papers
  - 24 conference notes



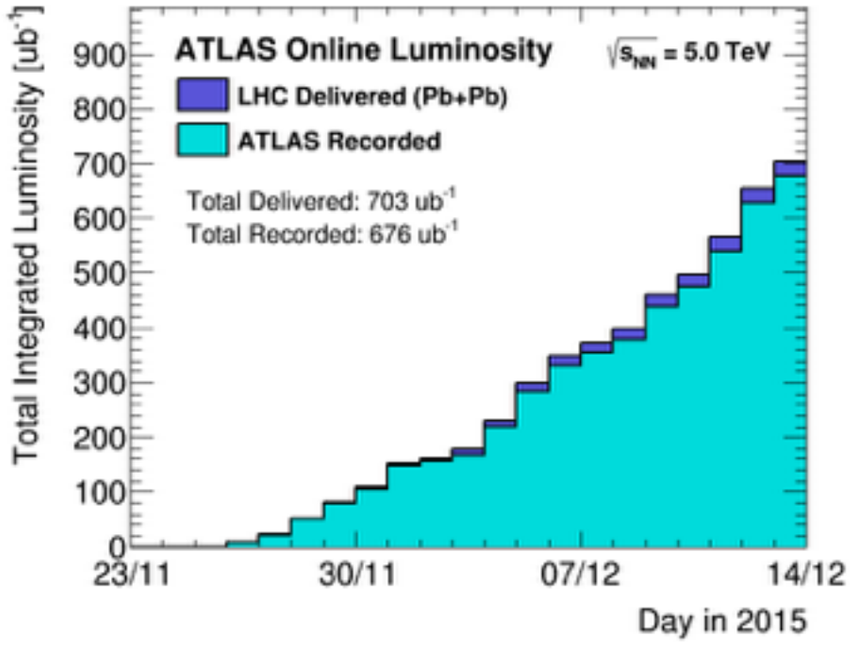
2015 data taking

- 3.5 fb-1 of 25 ns data good for physics
  - 3.2 fb-1 if IBL (new innermost Pixel layer) is required
  - Data quality efficiency: 87% - 93%
- HI : 0.67 nb-1 recorded

Proton-proton

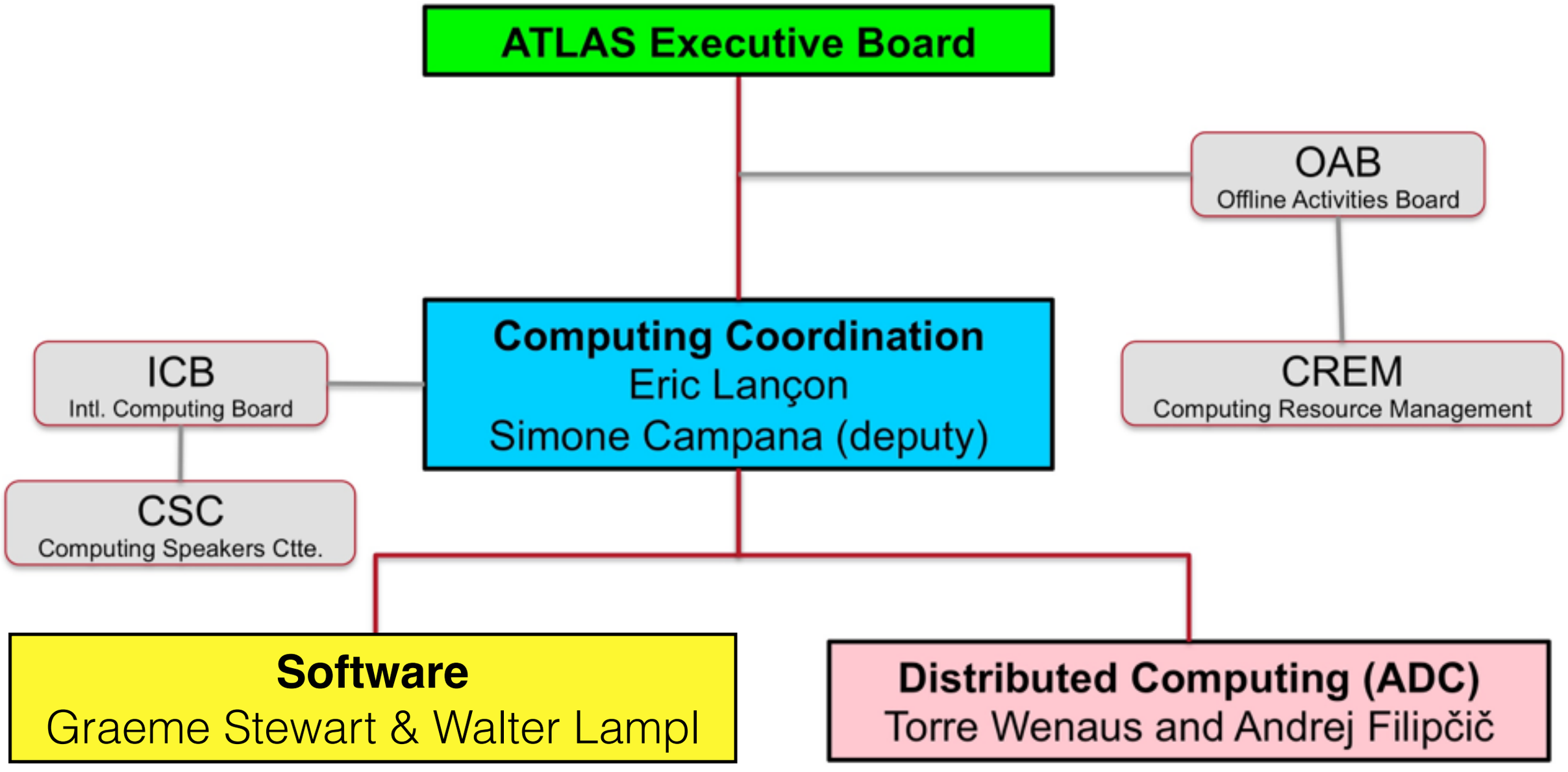


Heavy ion



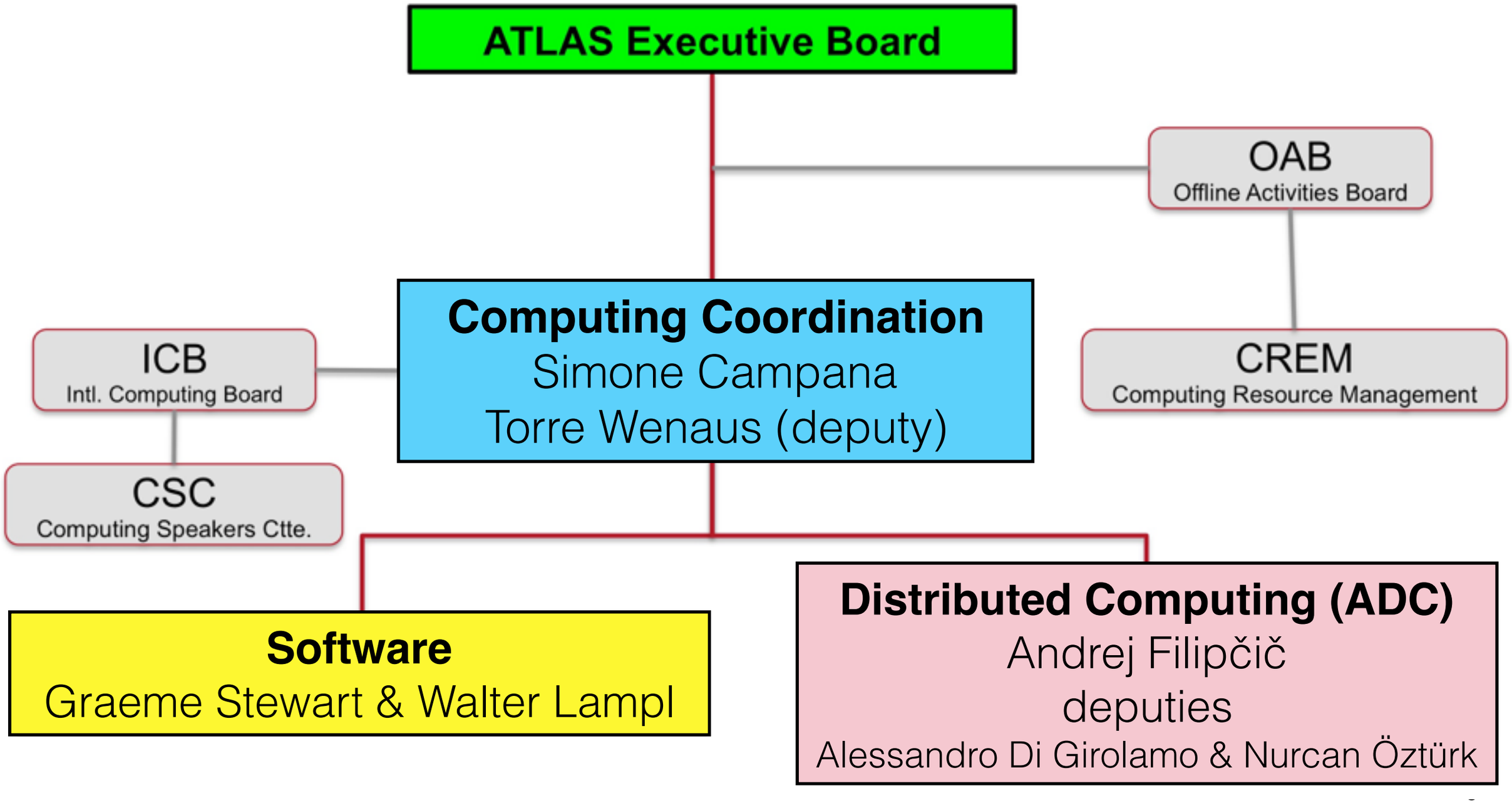
Changes in ATLAS S&C

Before March 1st 2016



Changes in ATLAS S&C

After March 1st 2016

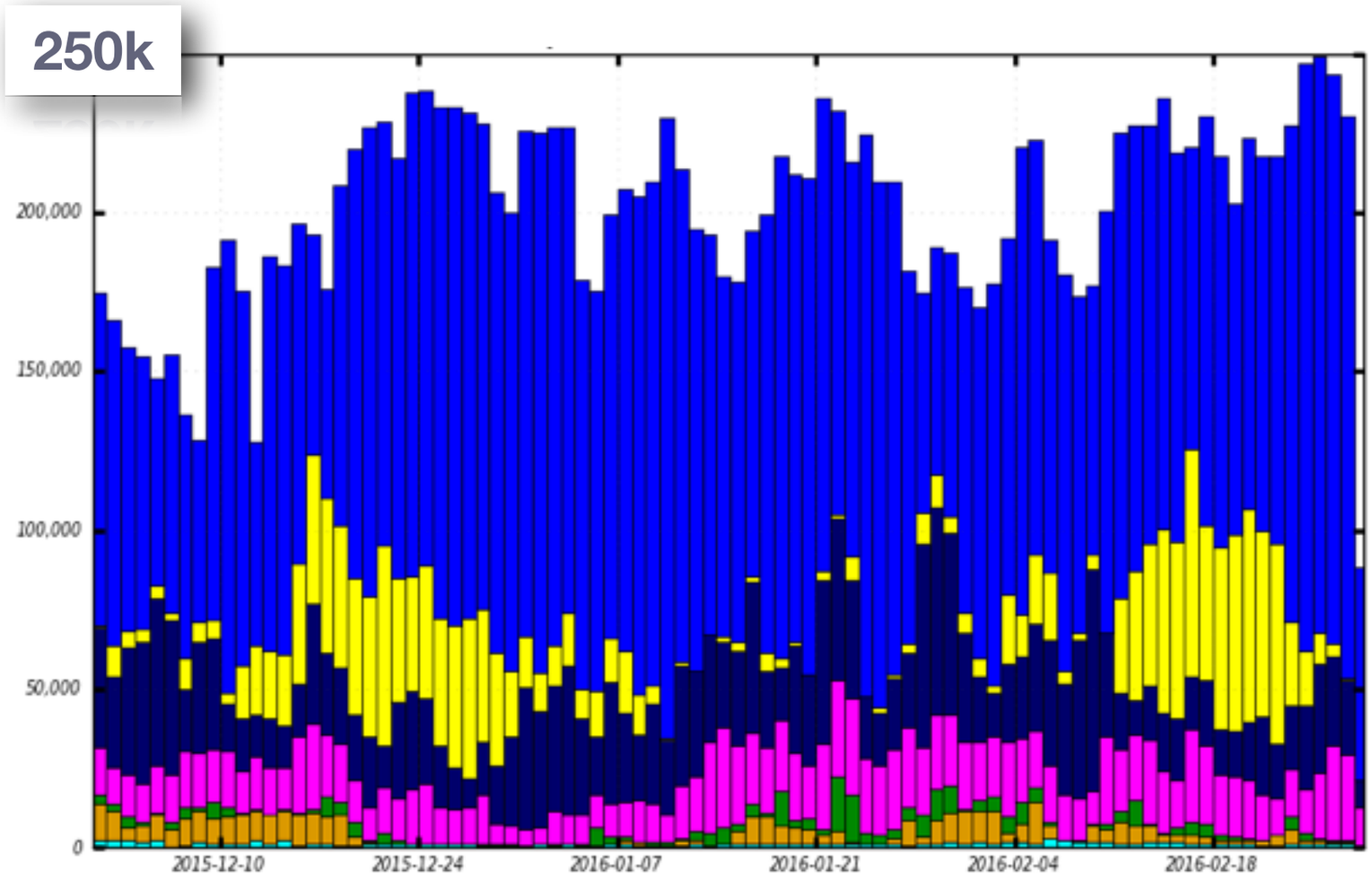


Usage of resources

Grid activity since last meeting

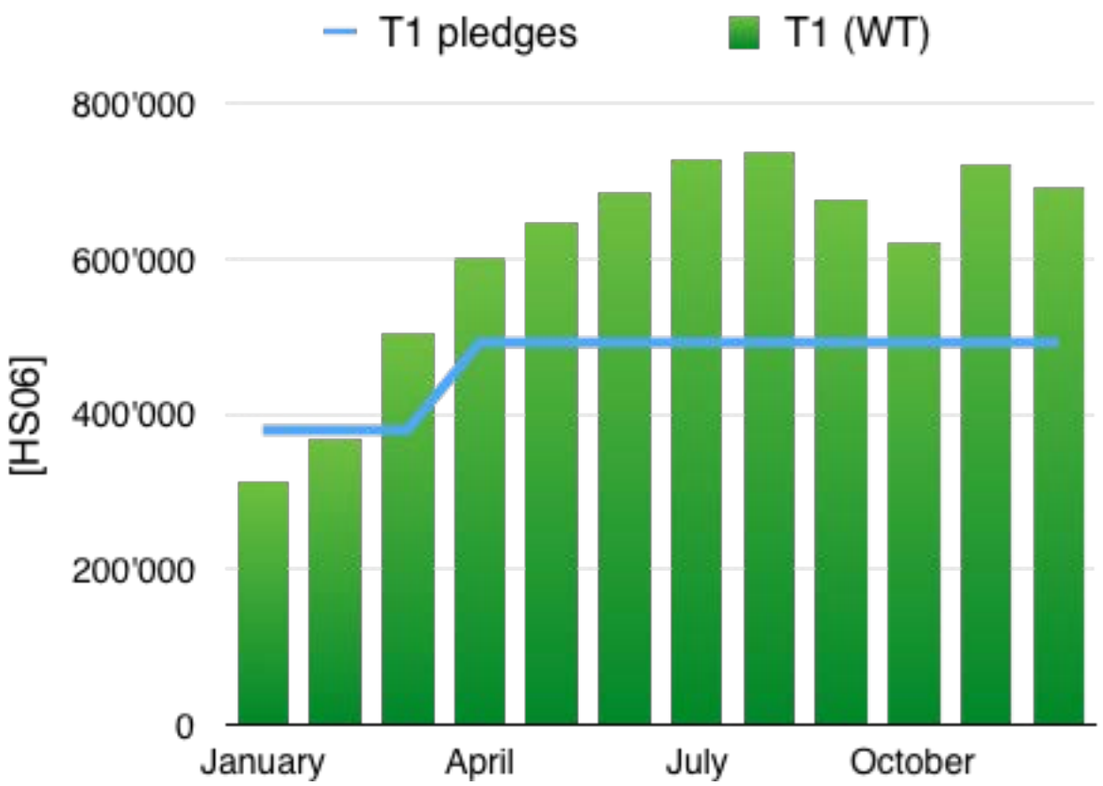
- Again record of activity
  - 250k core used
  - well above pledges
- 2015 pp data have been reprocessed twice
- HI HardProbe stream has not been processed fully yet
  - software developments to improve on speed and memory consumption
  - almost finished

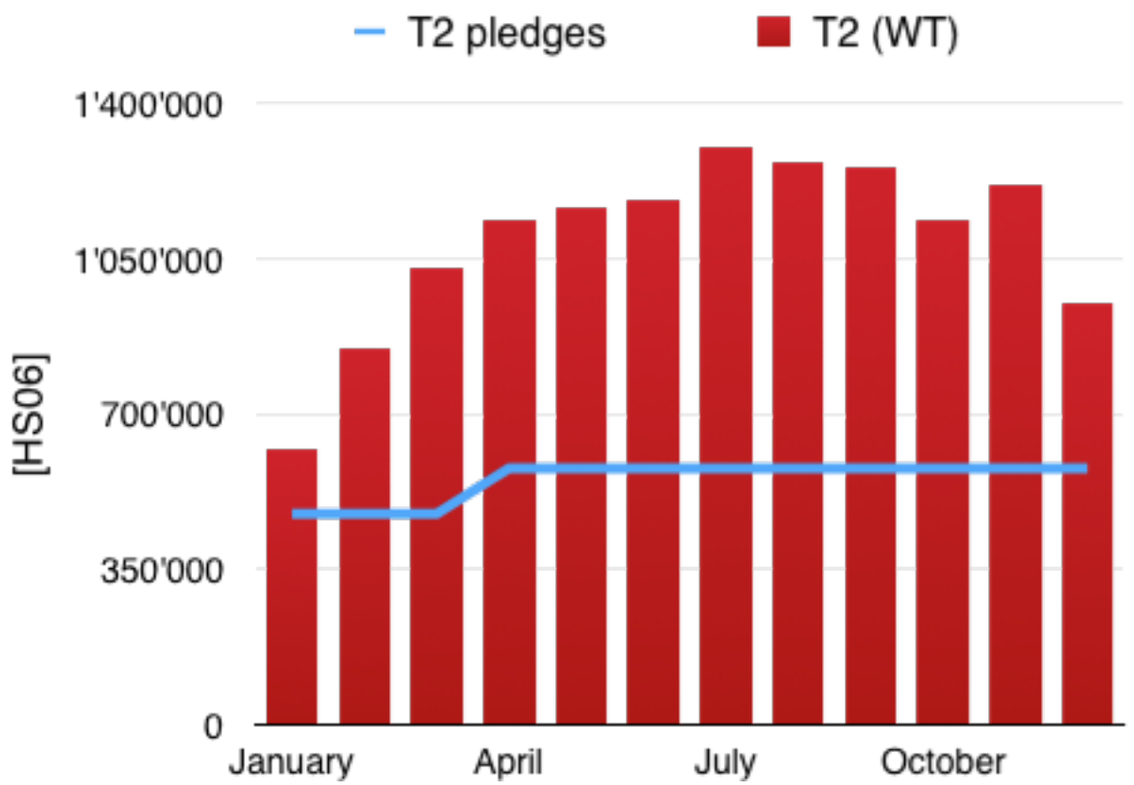
Number of used cores



MC simulation, MC reconstruction
Reprocessing, User analysis

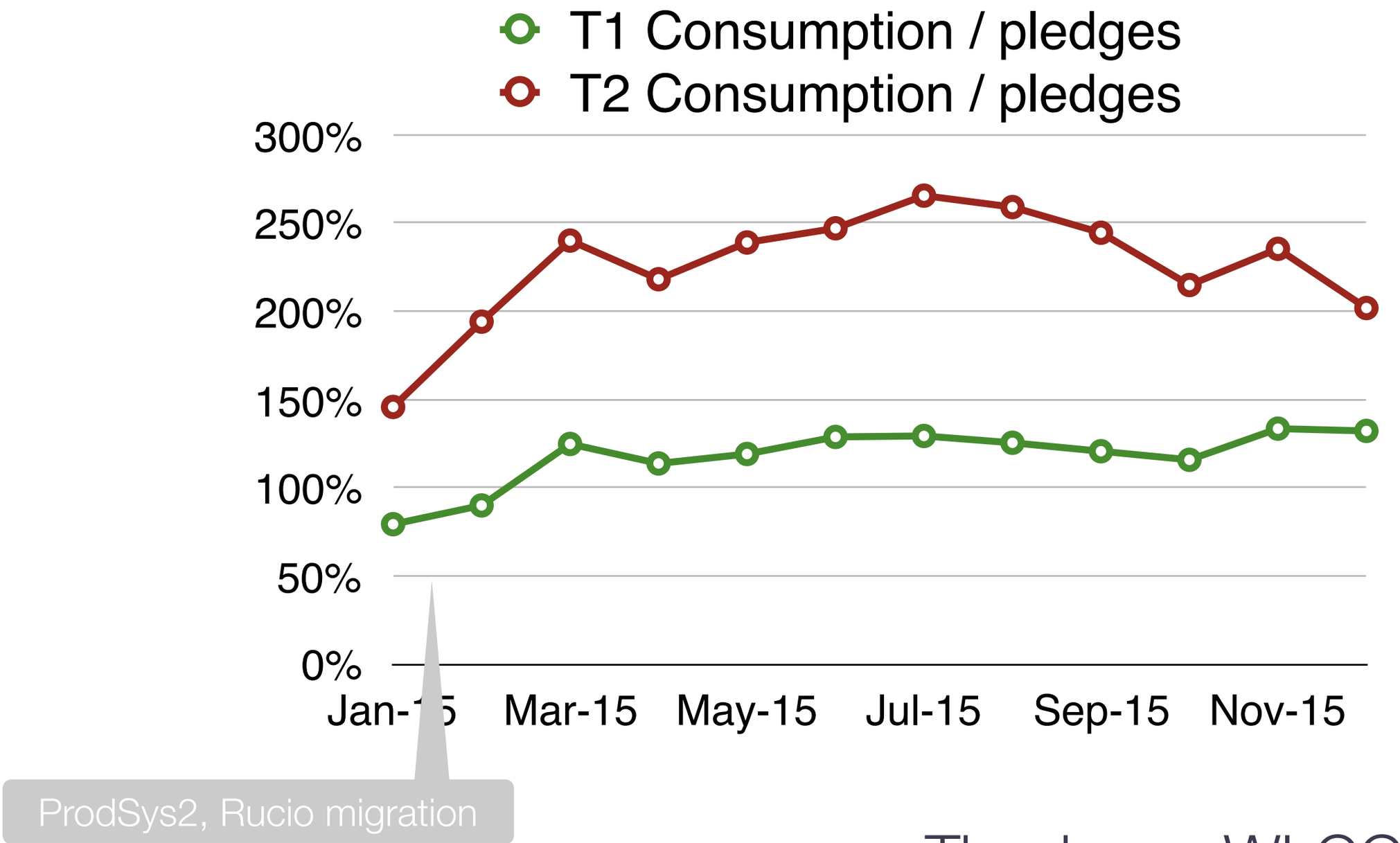
CPU consumption in 2015





Consumption continues to be above pledges at T1s and T2s

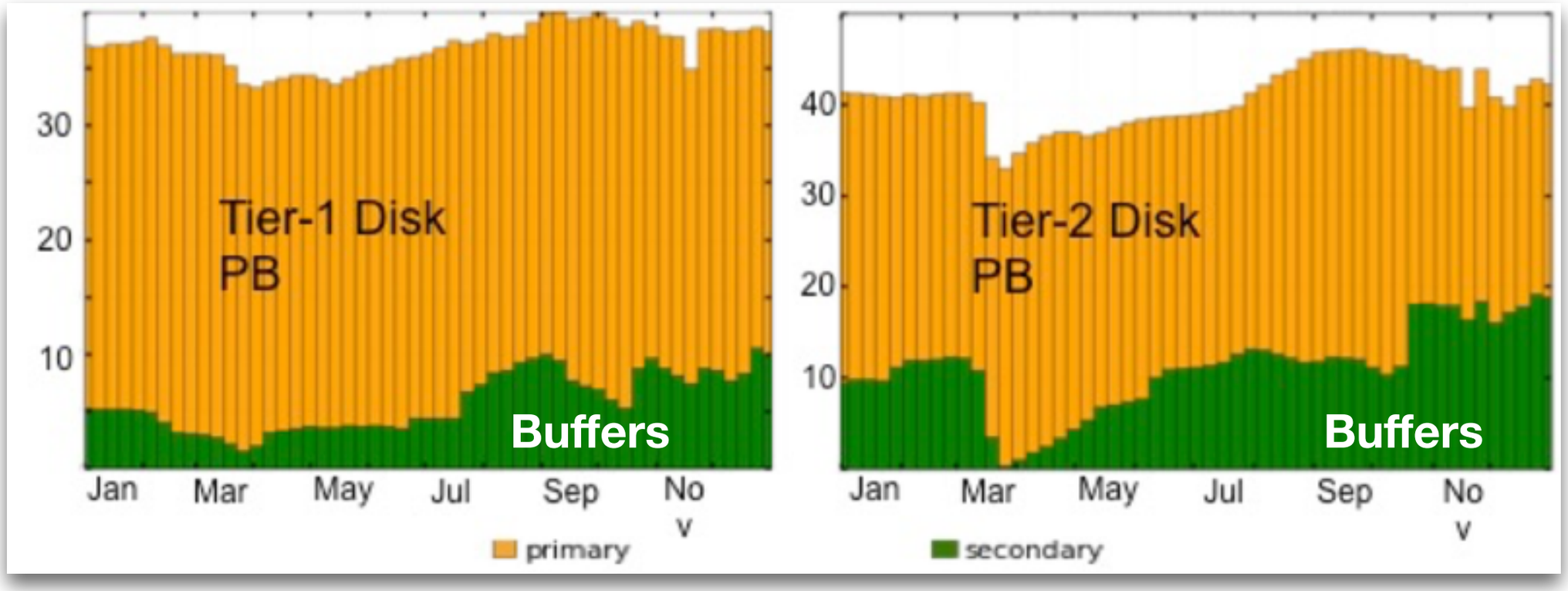
Consumption / pledges



Thank you WLCG sites!

Disk occupancy and availability

Over 90% occupancy **Around 80% occupancy**

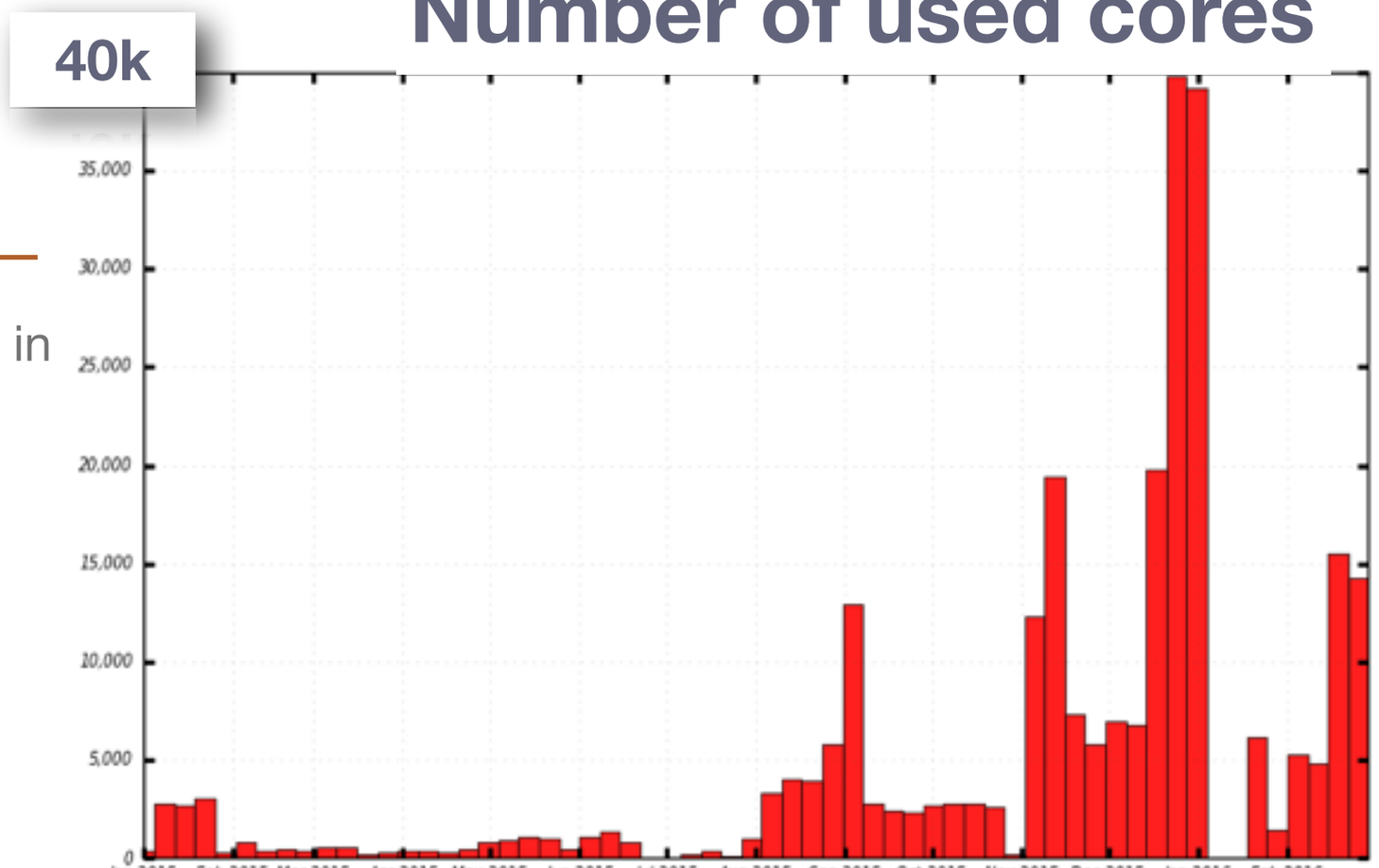


2015 [PB]	Tier-1	Tier-2
Available	47	60
Used	41	47
Used/Available	88%	78%

+ 5-10% cannot be used (tape buffers)

HLT farm

- Used for simulation (SimP1) mainly the in second part 2015
  - during LHC technical stops
  - when no ongoing work on trigger
  - record processing over end of year break



- **For 2016: ATLAS does not foresee to use the HLT farm for simulation for declared LHC between-fill time shorter than 4 hours**
  - Development and testing of automated switching between DAQ and SimP1 modes well advanced
  - Allows switching of the farm by shifter

Map of TDAQ TPU racks assignment, as host-status from TDAQ SysAdmins' ConfDB.

Rack Mode	01 TDAQ	02 TDAQ	03 TDAQ	04 TDAQ	05 TDAQ	06 TDAQ	07 TDAQ	08 TDAQ	09 TDAQ	10 TDAQ
Rack Mode	11 TDAQ	12 TDAQ	13 TDAQ							
Rack Mode										
Rack Mode										
Rack Mode				44 Sim@P1	45 Sim@P1	46 Sim@P1	47 Sim@P1	48 Sim@P1	49 Sim@P1	50 Sim@P1
Rack Mode	51 Sim@P1	52 Sim@P1	53 Sim@P1	54 Sim@P1	55 Sim@P1	56 Sim@P1	57 Sim@P1	58 Sim@P1	59 Sim@P1	60 Sim@P1
Rack Mode	61 Sim@P1	62 Sim@P1	63 Sim@P1	64 Sim@P1	65 Sim@P1	66 Sim@P1	67 Sim@P1	68 Sim@P1	69 Sim@P1	70 TDAQ
Rack Mode	71 TDAQ	72 TDAQ	73 TDAQ	74 Maint					79 Maint	80 TDAQ
Rack Mode	81 TDAQ	82 TDAQ	83 TDAQ	84 TDAQ	85 TDAQ	86 TDAQ	87 TDAQ	88 Maint	89 Maint	90 Maint
Rack Mode			93 Sim@P1	94 Sim@P1	95 Sim@P1					

See ticket #2180 [PERMANENT] time slots & resources for Sim@P1 operations

Software

AthenaMT

- Multi Threaded version of Athena based on Gaudi Hive
- Migration of all software to MT for Run 3
- Join effort between ATLAS offline and trigger communities
- Good progresses, on track wrt initial planing

Date	Framework
2015	Event Store Access via Data Handles; Event View Design Completed; Updated Configuration Design; Re-integration of Hive features into Gaudi trunk
Q2 2016	Event Views Implemented; IO Layer Redesigned Core Gaudi service migration starts
Q4 2016	Parallel algorithm support; Detector Store Reimplemented; Schedulable Incidents; Main athena development branch moved to Gaudi trunk

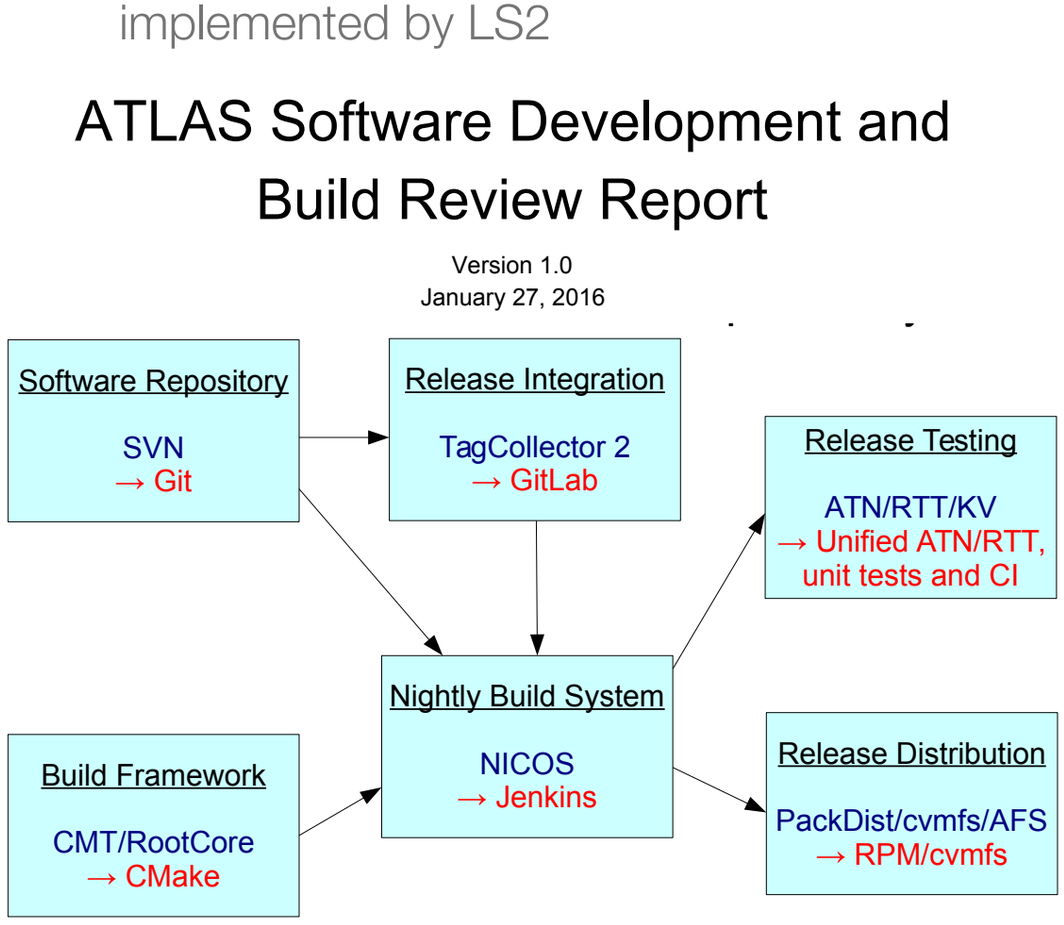
Gaudi Hive

Collaboration between ATLAS, LHCb, FCC experiments/projects

Would benefit from long term support and commitment from CERN

Software developments

- **Strong engagement in software quality**
  - systematic use of static and dynamic code checkers
  - enforcement of programming rules
  - implementation of code reviews
- **Simulation software**
  - Effort for easy harvest of opportunistic resources (HPC, cloud, volunteer)
  - Refactoring of ATLAS software to allow processing simulation without installing the full ATLAS software stack
- Good progresses on **condition data** for Run 3 project
  - BELLE II recently joined discussions between ATLAS & CMS, within the framework of the HSF
- **Software infrastructure and development review**
  - Document published within ATLAS
  - including 24 recommendations to be implemented by LS2



Move away from custom tools

Experience from first year of Run 2 and forward looking

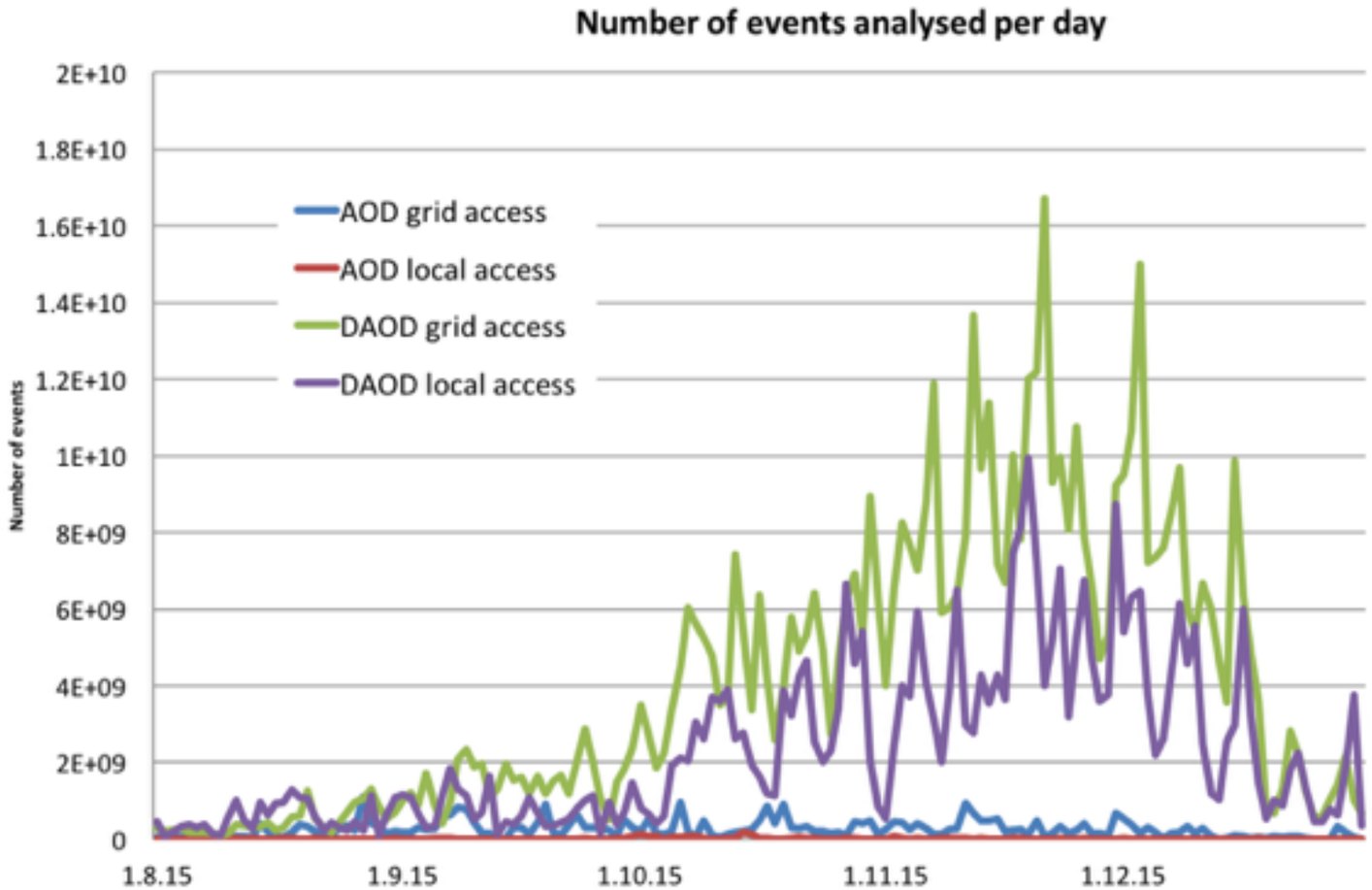
Run 2 experience

- ATLAS consumes more CPU than pledges
- New data management: dataset lifetime, increased tape usage
- New analysis model : DAODs
- Fast simulation not ramping up
- Higher usage of more sophisticated event generators
- Tier hierarchy flattening
- Analysis the driver for storage and network

…….+ many others

AOD vs DAOD

- Success of **Derivation Framework** and train model
- Fraction of CPU used by analysis on Grid dropped from ~25% to **10%** in last 4 months of 2015
- **95%** of analysed events through DOAD format
- **35%** of events analysed locally
  - without grid tools
  - but inputs can be located on grid storage (local facility,…)



Evolution of computing model

Planning for LHC Run 3 horizon
Today’s decisions, effect in 4-5 years (lifetime of equipments)

Data processing and storage

(re)Processings		
	Year N	Year N-x
Data / MC	2	1
Derivations	6	2

N-2 policy

Disk Storage		
(D)AOD	Year N	Year N-x
Versions	2	1
Copies	2	1

DAODs are not stored on tape

+ dynamic increase of copies based on popularity

(D)AOD data placement

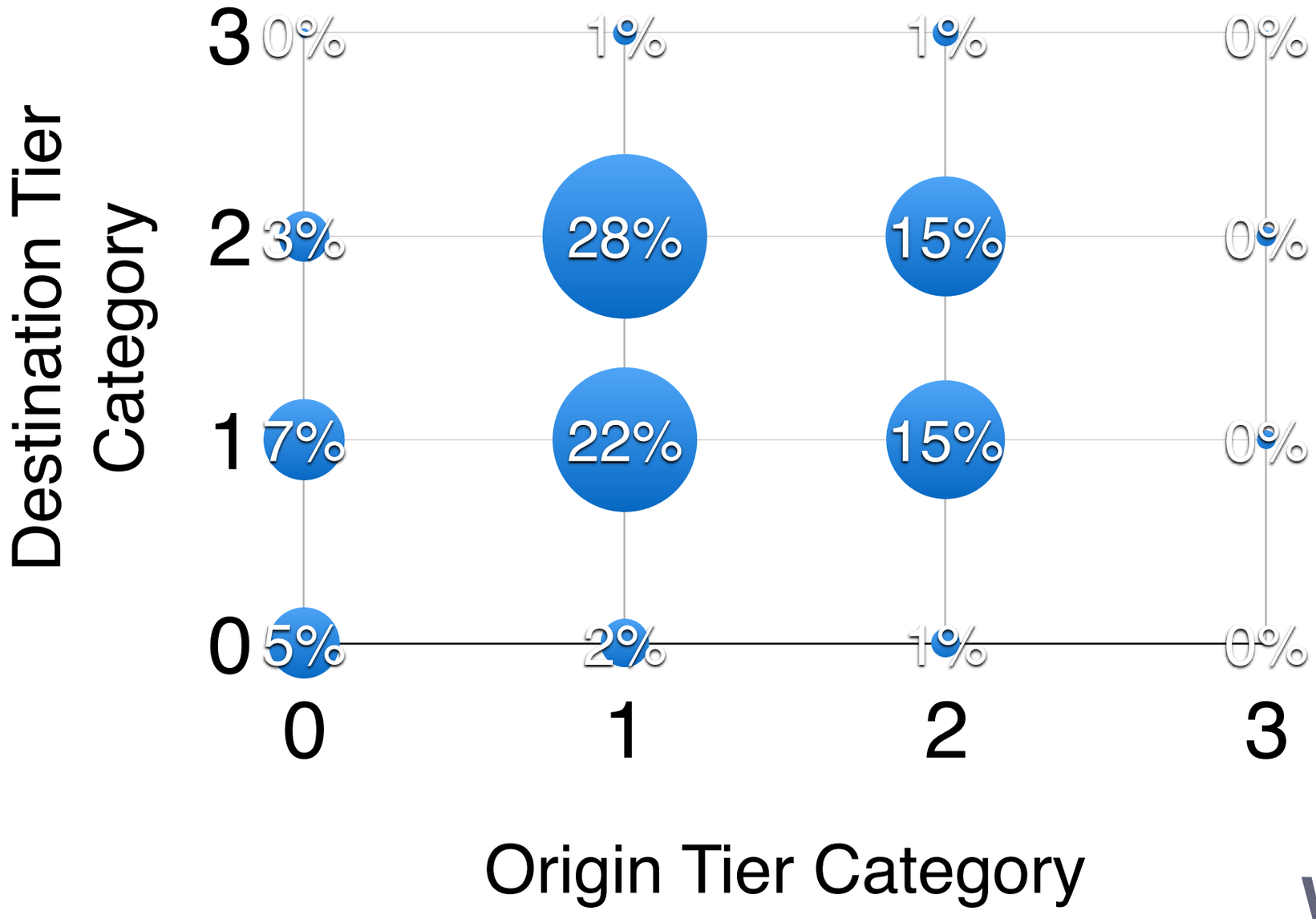
- **Run 1 start up:**
  - **>30 copies** of a given dataset (x per cloud)
  - Storage at every site - costly!
- **Run 2 :**
  - **2 copies** guaranteed
  - + dynamic (popularity) data placement - Network !
  - Increase of remote access and caching

ATLAS sites

(Too) many storage and computing end-points

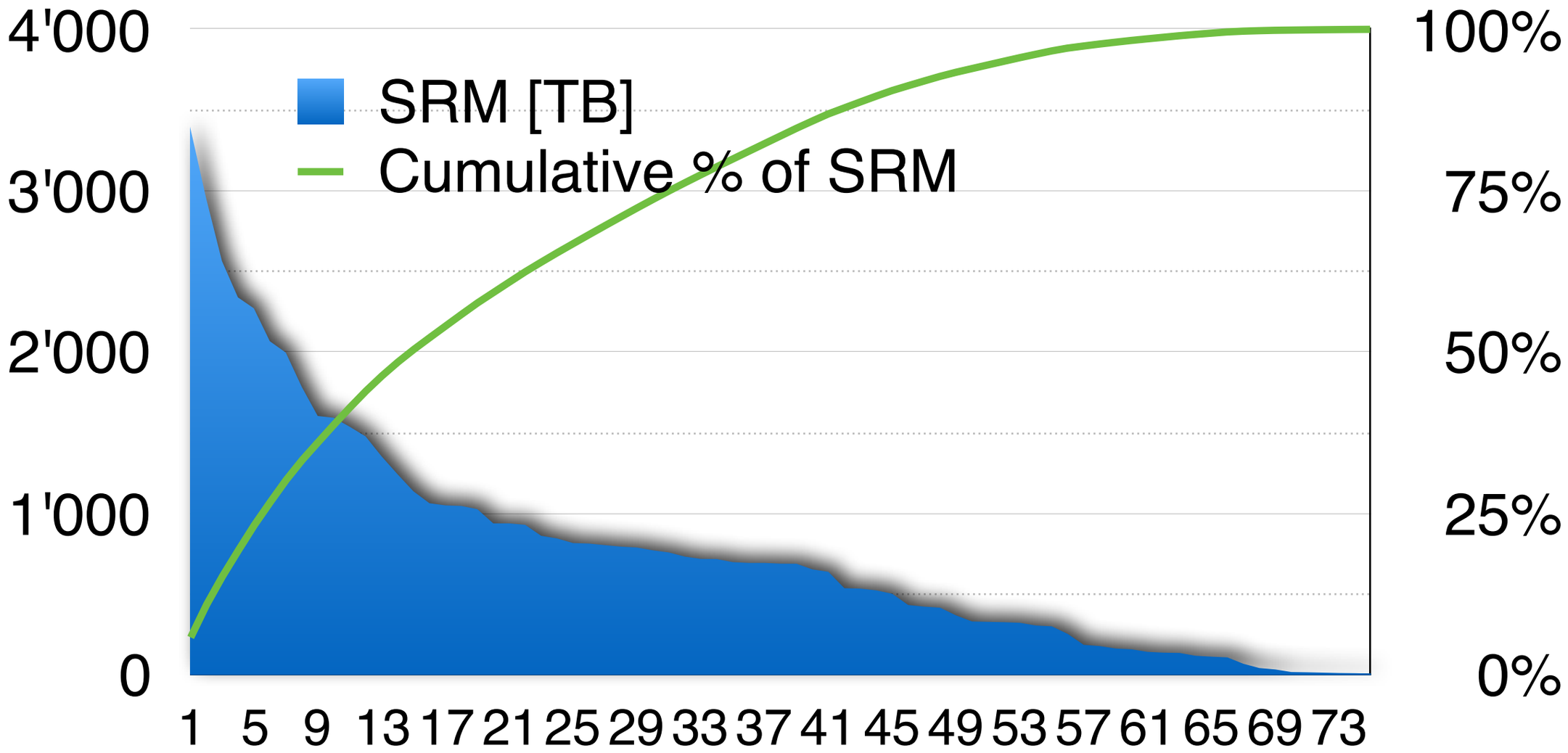
Transfer matrix (data volume)

The hierarchical historical model is gone



WAN access will increase!

Available storage at Tier 2 sites



More efficient to have larger and fewer storage end-points
2 possible categories : 'Cache based' & 'large' Tier 2s

Numbers in next slides have **not been reviewed** by CRSG

Preliminaries

- Model does not differentiate T1s & T2s, table on next slides shows 2017(8) requests for T1s & T2s individually based on cost model (flat budget) and some history
- Balance of resources between T1s & T2s for disk & CPU:
  - Funding Agency decision (provided reliability of sites is excellent)
    - To be discussed with ATLAS
- T1s are the backbone of the infrastructure
  - Tapes (usage will grow)
  - High quality of services (storage, availability, support…)
  - Flexibility of workflows

ATLAS requests for 2017 & 2018

	2016 C-RSG CERN-RRB-2015-014	2017 ATLAS October 2015 (Preliminary)	2017 ATLAS March 2016 (Final)	2018 ATLAS March 2016 (Preliminary)
T0 CPU	257	270	383	389
T1 CPU	520	662	703	763
T2 CPU	566	702	846	946
T0 Disk	17	18	20	20
T1 Disk	47	54	57	60
T2 Disk	72	91	78	84
T0 Tape	42	51	53	67
T1 Tape	116	185	173	257

[kHS06]

[PB]

Within ‘flat’ budget

2017 cost ~ 90% of 2016

new 2017 request ~5% lower than previous one

T1 request

	2016 C-RSG CERN-RRB-2015-014	2017 ATLAS October 2015	2017 ATLAS March 2016	2018 ATLAS March 2016
T0 CPU	257	270	383	389
T1 CPU	520	662	703	763
T2 CPU	566	702	846	946
T0 Disk	17	18	20	20
T1 Disk	47	54	57	60
T2 Disk	72	91	78	84
T0 Tape	42	51	53	67
T1 Tape	116	185	173	257

Reduced tape request, marginal increase of disk & CPU

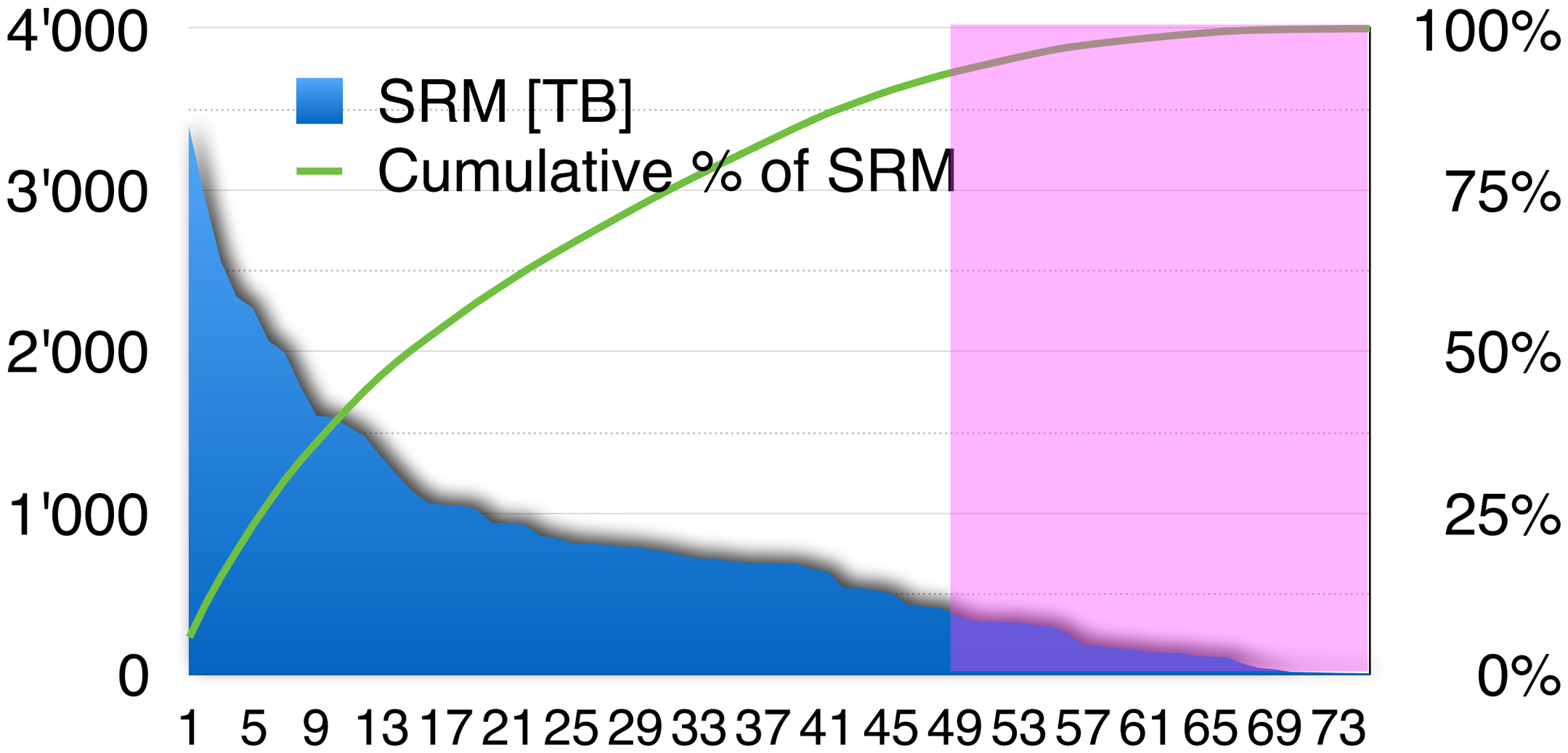
T2 request

	2016 C-RSG CERN-RRB-2015-014	2017 ATLAS October 2015	2017 ATLAS March 2016	2018 ATLAS March 2016
T0 CPU	257	270	383	389
T1 CPU	520	662	703	763
T2 CPU	566	702	846	946
T0 Disk	17	18	20	20
T1 Disk	47	54	57	60
T2 Disk	72	91	78	84
T0 Tape	42	51	53	67
T1 Tape	116	185	173	257

Trade disk for CPU resources at T2s

Disk request can further be reduced to ~73 PB by optimised placement of storage; to be discussed between FAs, WLCG & ATLAS

Storage at Tier 2 sites



Storage end points below **400 TB** should either :

- not invest in, nor renew, storage or
- aggregate with other end point(s)

Beyond pledged CPU resources still needed!

	2016 ATLAS	2017 ATLAS	2017 ATLAS	2018 ATLAS
Additional Goal		**October 2015**	**March 2016**	**March 2016**
CPU [kHS06]	800	800	696	728

Real needs ~150% of request submitted to CRSG

Possible evolutions of computing model

Possible evolution

- **Sizeable (TBD) regional centres**
  - True federations of distributed resources (Network!)
  - One entry point by centre
  - National / trans-national regional centres to match the scale
  - Technical solution to be worked out within WLCG
- **Cache based sites for those not part of a regional center**

DPM need to be supported for still many years

Summary

- LS1 efforts in evolving computing and analysis models, software, distributed computing and facilities showed a strong payoff in Run 2 Year 1

- Submitted 2017 resource request to CRSG on time, supported by quantitative metrics original to ATLAS

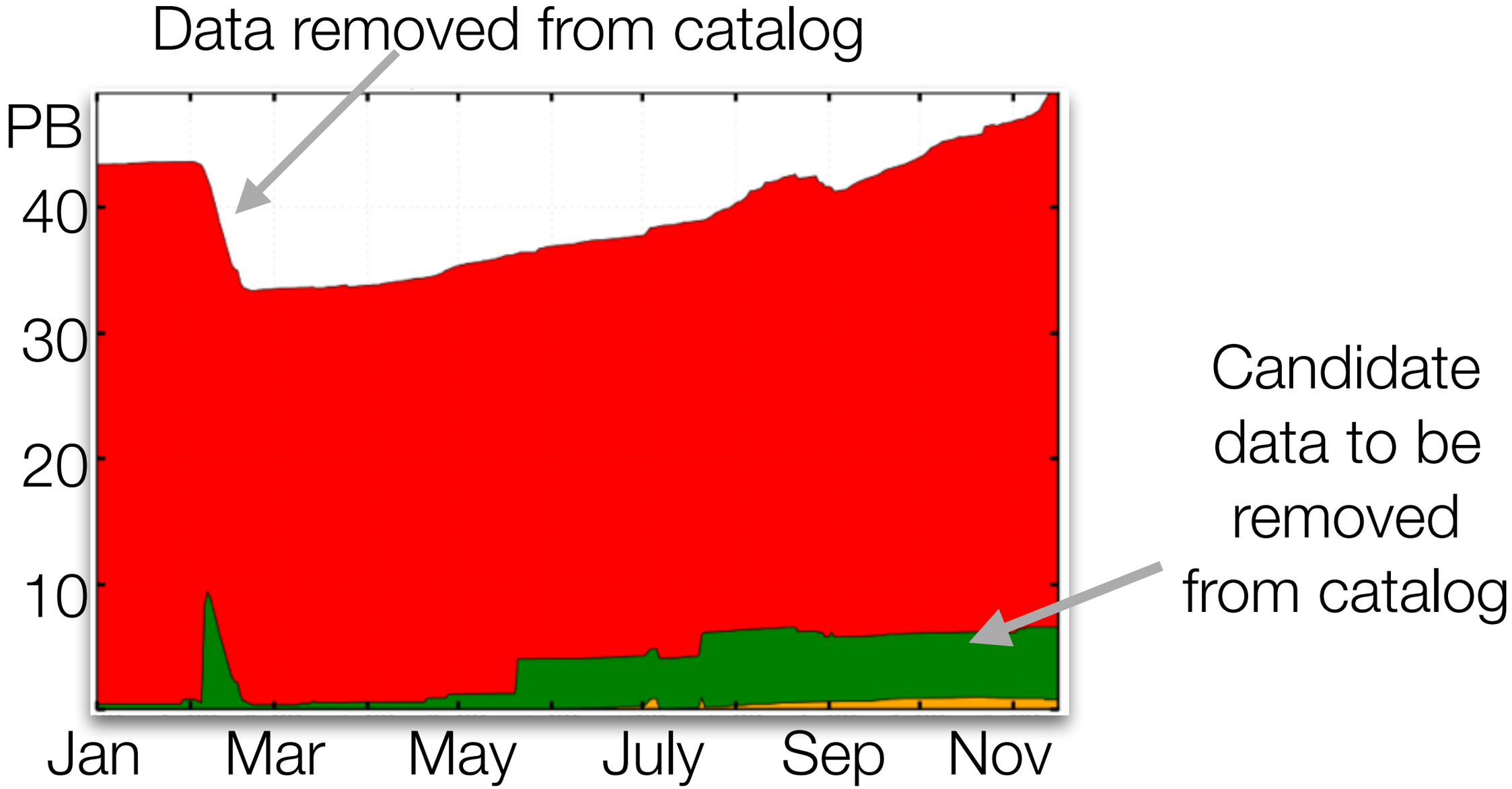
- Looking towards Run 3/4 needs with new rounds of development and optimization with potentially large payoffs, such as increased storage efficiency through federations and further leveraging the network

- Future evolutions of computing model will be discussed and iterated with Funding Agencies

Backups

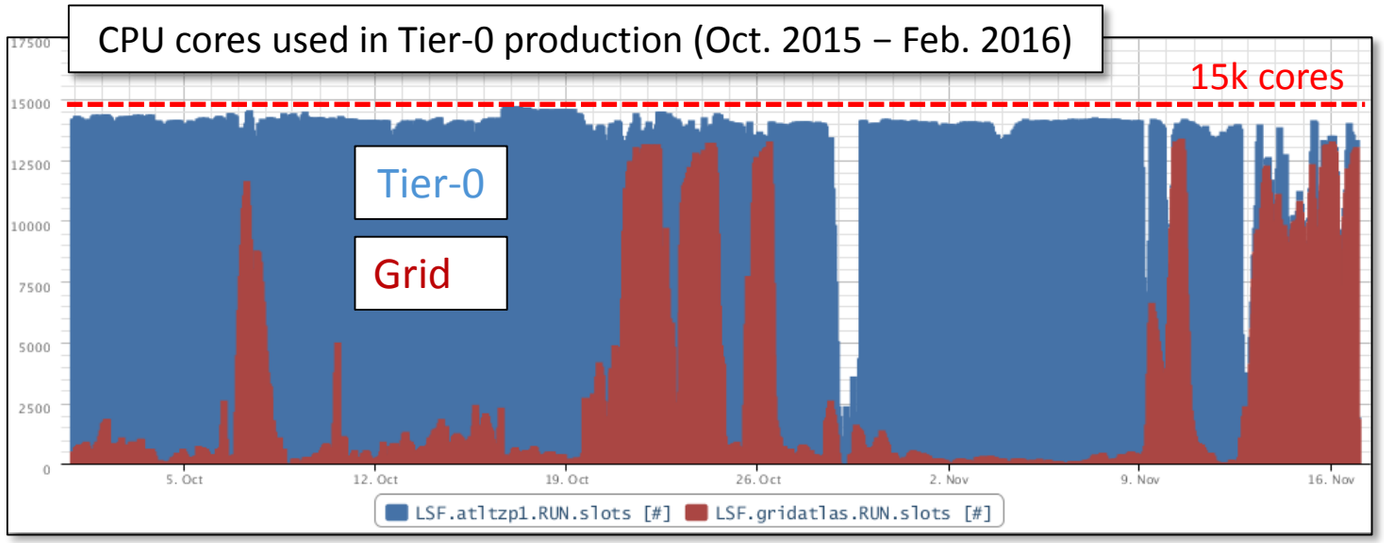
Dynamic usage of tape resources

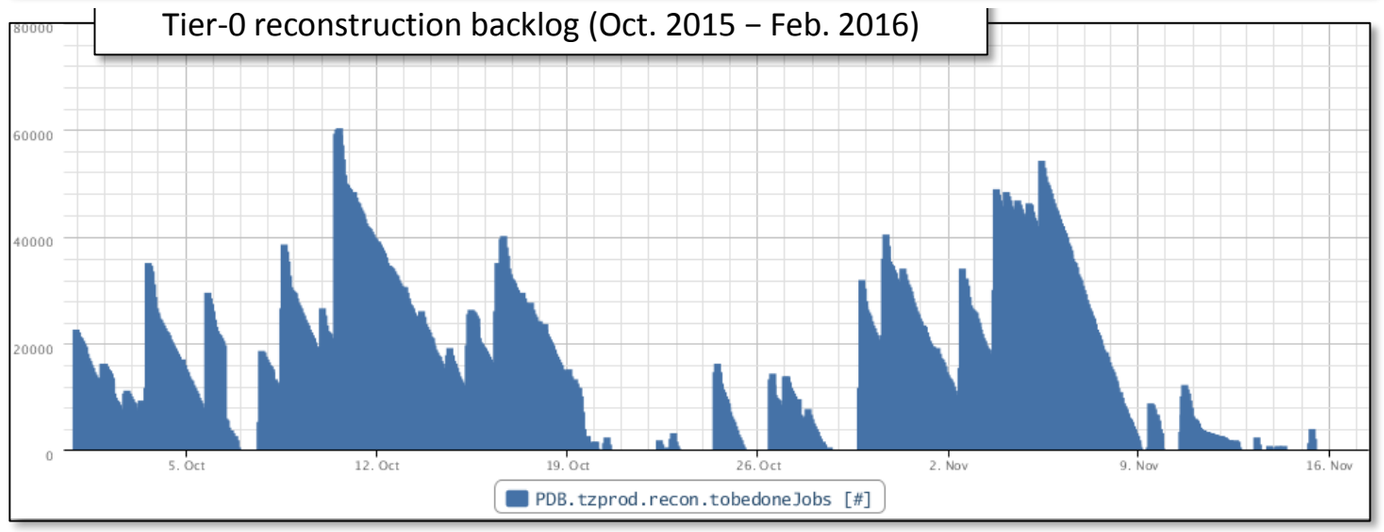
Tape usage at T1s



Tier-0 CPU request update for 2016 & 2017

- Tier 0 ressources saturating end of 2015 pp period for high LHC luminosity/ efficiency:
  - AthenaMP: lower CPU efficiency
  - Extra activities beyond bulk reconstruction
- Reevaluation of needs for 2016 and 2017, hypothesis : no backlog for
  - 1kHz,
  - 20/24 LHC running
- Expected improvements in workflows (~20%) included in updated requirements
- Tests being conducted to optimise hardware setup and workflows





kHS06	2016	2017
Old	233	274
New	341	383

Additional needs for 2016 under negotiation with CERN

Resources for upgrade studies

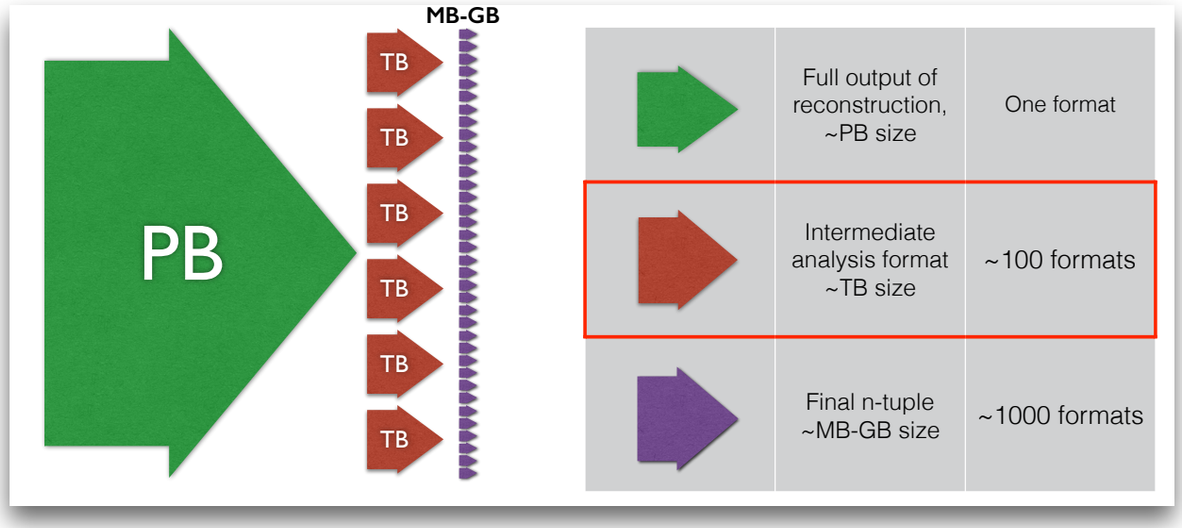
Table 5: the main parameters used to evaluate the resource needs for upgrade studies in 2017

	Simulation	Reconstruction				
		μ=60 (Run-3)	μ=0	μ=80	μ=140	μ=200
HS06/event	5170	825	110	1210	3025	7040
M events	147	23	59	59	59	59
kHS06 years	24.0	0.6	0.2	2.3	5.7	13.2

Derivation framework

- New analysis model for Run 2: group data format **DAOD** made using a train model
- Production of 84+ **DAOD** species by 19 trains on the grid
  - 24h after data reconstruction at Tier-0
  - Working!
  - Vital for quick turn around and robustness of analyses
- 2015 ATLAS results based on **DAODs**!

From PB to GB



AOD **DAOD** **n-tuple**

European sites

Possible evolutions of computing model