How we validate our accounting data

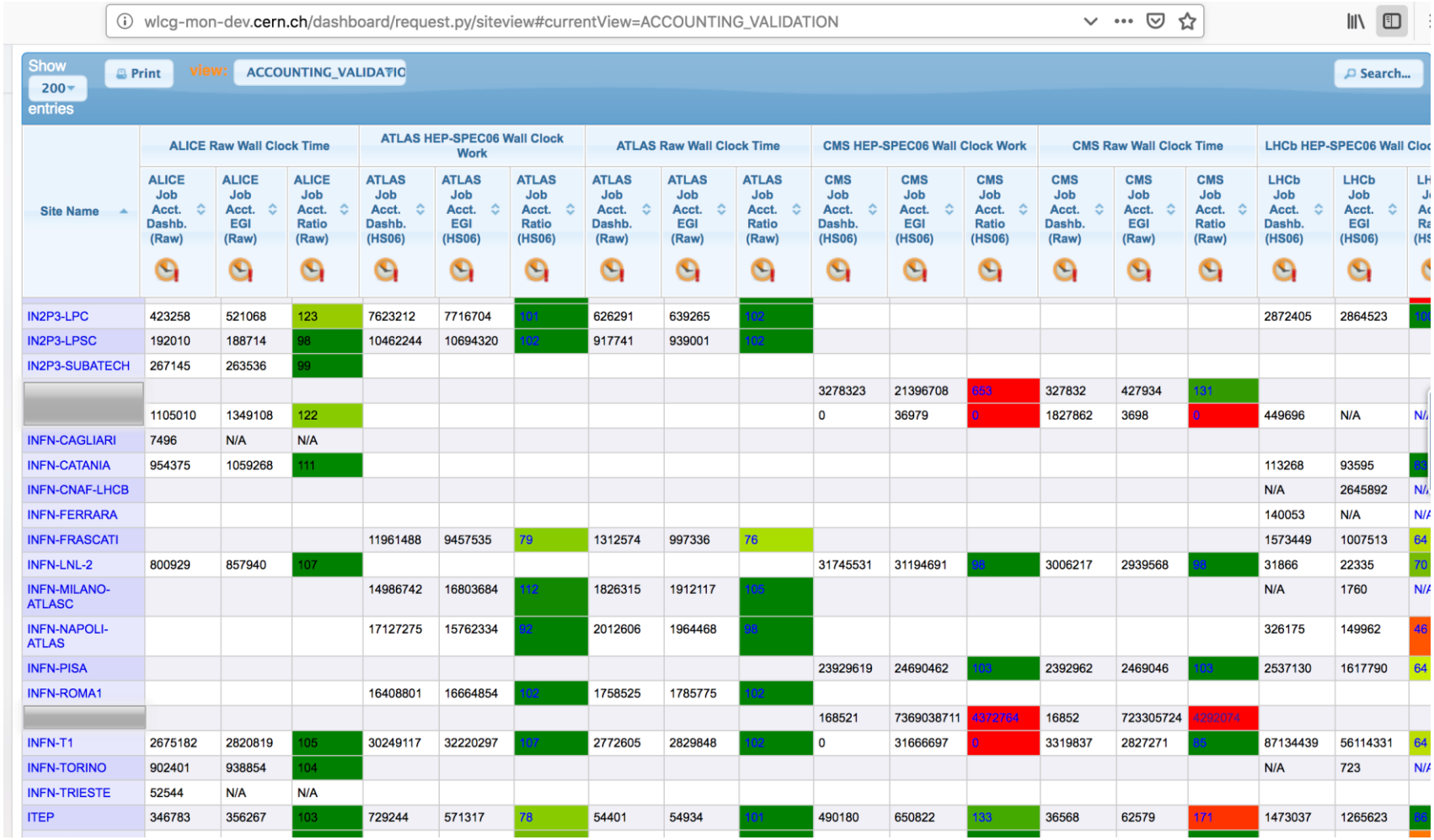
GDB 13.03.2019

Julia Andreeva , CERN

Cross checks with alternative data sources. CPU.

- For CPU accounting an automatic validation of the EGI accounting data is performed in the beginning of every month.
- Monthly EGI numbers are compared with the internal experiment accounting (wallclock time and wallclock work).
- Pilot consumption is compared with payload consumption, so 100% consistency is not expected. Moreover, the difference is defined by pilot efficiency
- Results are available in SSB

Cross checks with alternative data sources. CPU. Example (2)

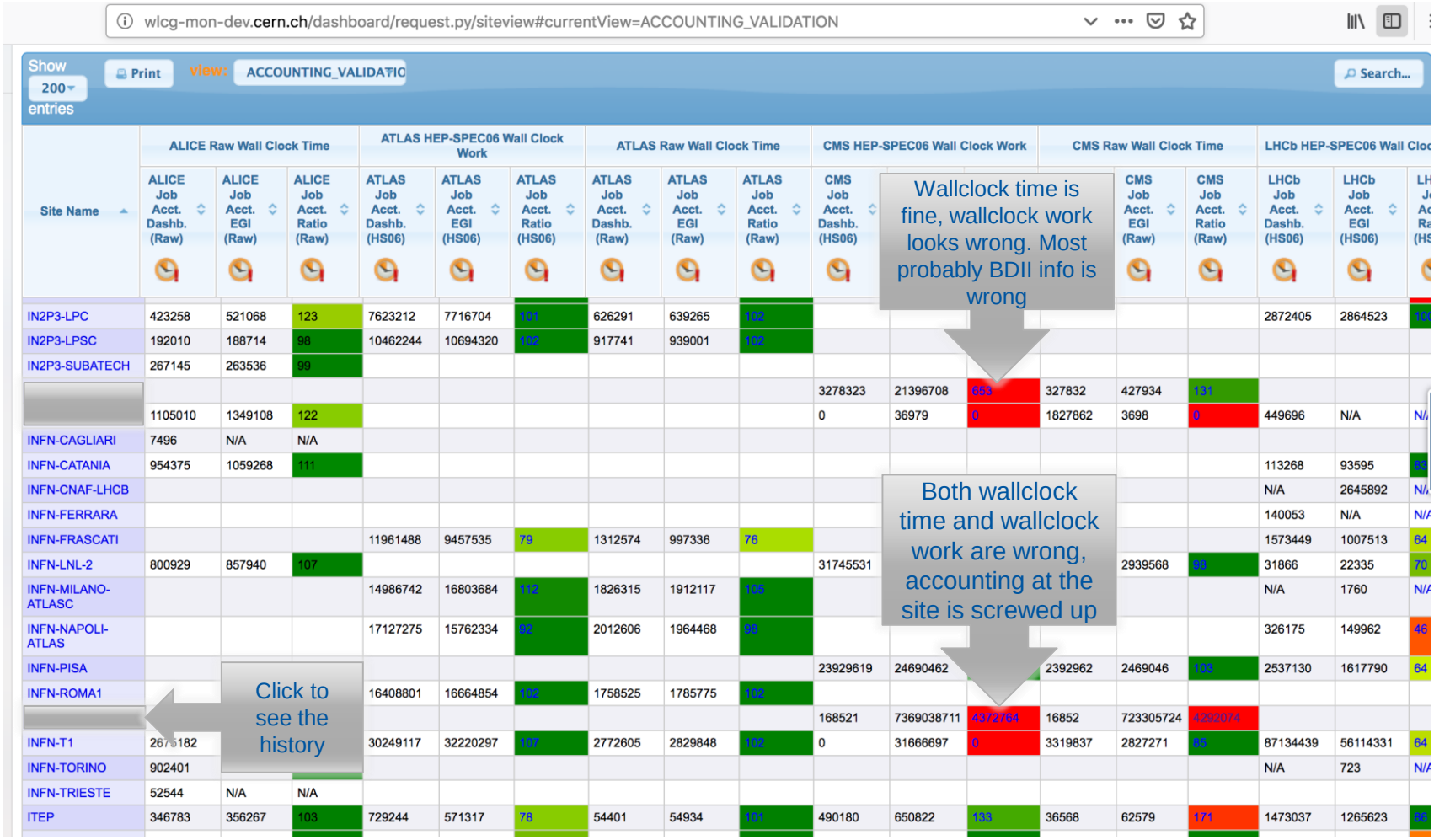


wlcg-mon-dev.cern.ch/dashboard/request.py/siteview#currentView=ACCOUNTING_VALIDATION

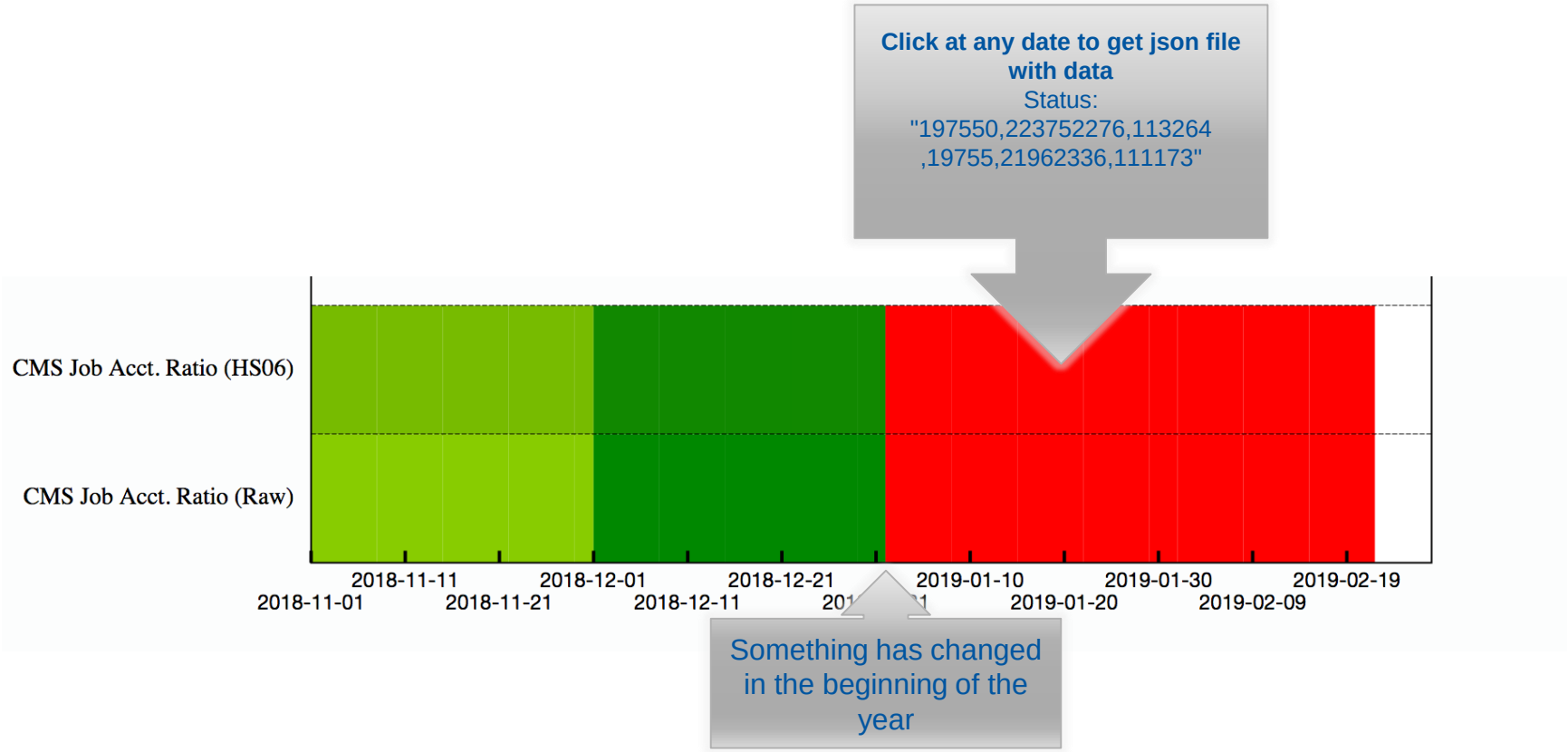
Show 200 entries | Print | view: ACCOUNTING_VALIDATIC | Search...

Site Name	ALICE Raw Wall Clock Time			ATLAS HEP-SPEC06 Wall Clock Work			ATLAS Raw Wall Clock Time			CMS HEP-SPEC06 Wall Clock Work			CMS Raw Wall Clock Time			LHCb HEP-SPEC06 Wall Cloc		
	ALICE Job Acct. Dashb. (Raw)	ALICE Job Acct. EGI (Raw)	ALICE Job Acct. Ratio (Raw)	ATLAS Job Acct. Dashb. (HS06)	ATLAS Job Acct. EGI (HS06)	ATLAS Job Acct. Ratio (HS06)	ATLAS Job Acct. Dashb. (Raw)	ATLAS Job Acct. EGI (Raw)	ATLAS Job Acct. Ratio (Raw)	CMS Job Acct. Dashb. (HS06)	CMS Job Acct. EGI (HS06)	CMS Job Acct. Ratio (HS06)	CMS Job Acct. Dashb. (Raw)	CMS Job Acct. EGI (Raw)	CMS Job Acct. Ratio (Raw)	LHCb Job Acct. Dashb. (HS06)	LHCb Job Acct. EGI (HS06)	LH Ac Ra (HS
IN2P3-LPC	423258	521068	123	7623212	7716704	101	626291	639265	102							2872405	2864523	10
IN2P3-LPSC	192010	188714	98	10462244	10694320	102	917741	939001	102									
IN2P3-SUBATECH	267145	263536	99															
										3278323	21396708	653	327832	427934	131			
	1105010	1349108	122							0	36979	0	1827862	3698	0	449696	N/A	N/
INFN-CAGLIARI	7496	N/A	N/A															
INFN-CATANIA	954375	1059268	111													113268	93595	83
INFN-CNAF-LHCB																N/A	2645892	N/
INFN-FERRARA																140053	N/A	N/A
INFN-FRASCATI				11961488	9457535	79	1312574	997336	76							1573449	1007513	64
INFN-LNL-2	800929	857940	107							31745531	31194691	98	3006217	2939568	98	31866	22335	70
INFN-MILANO-ATLASC				14986742	16803684	112	1826315	1912117	105							N/A	1760	N/A
INFN-NAPOLI-ATLAS				17127275	15762334	92	2012606	1964468	98							326175	149962	46
INFN-PISA										23929619	24690462	103	2392962	2469046	103	2537130	1617790	64
INFN-ROMA1				16408801	16664854	102	1758525	1785775	102									
										168521	7369038711	4372764	16852	723305724	4202074			
INFN-T1	2675182	2820819	105	30249117	32220297	107	2772605	2829848	102	0	31666697	0	3319837	2827271	85	87134439	56114331	64
INFN-TORINO	902401	938854	104													N/A	723	N/A
INFN-TRIESTE	52544	N/A	N/A															
ITEP	346783	356267	103	729244	571317	78	54401	54934	101	490180	650822	133	36568	62579	171	1473037	1265623	86

Cross checks with alternative data sources. CPU. Example (1)



Checking history. Example (3)



Cross checks with alternative data sources. Storage

- For WSSA we have two alternative data sources : experiment storage accounting systems and EGI storage accounting prototype
- No proper APIs are currently available for WSSA and EGI accounting prototypes, so crosschecks are performed manually by downloading monthly data
- Foresee to automate the procedure when APIs are available

Cross checks with alternative data sources. Storage. Example

	Grafana	EGI	EGI/GRAFANA
AGLT2	6.1E+15	-1	0
ANLASC	1.8E+12	-1	0
Australia-ATLAS	1.2E+15	-1	0
BEIJING-LCG2	3.085E+14	3.19684E+14	0.965017
BNL-ATLAS	5.99E+16	-1	0
CA-SFU-T2	2.1E+15	2.1555E+15	0.97425
CA-VICTORIA-WESTGRID-T2	6.912E+14	3.58026E+14	1.93058
CA-WATERLOO-T2	1.7E+15	-1	0
CERN-PROD	1.064E+17	-1	0
CSCS-LCG2	1.6E+15	-1	0
CYFRONET-LCG2	5.455E+14	5.50709E+14	0.990542
DESY-HH	1.3E+15	-1	0
DESY-ZN	1.3E+15	1.82107E+15	0.713866
EELA-UTFSM	3.091E+14	3.27407E+14	0.944084
FMPhI-UNIBA	4.989E+14	5.25995E+14	0.948488
FZK-LCG2	2.73E+16	2.78364E+16	0.980731
GRIF	3.2E+15	7.6557E+14	4.17989
GoeGrid	1.6E+15	-1	0
HK-LCG2	3.845E+14	3.84292E+14	1.00054
IEPSAS-Kosice	5.982E+14	6.53502E+14	0.915376
IFIC-LCG2	1.7E+15	-1	0
IL-TAU-HEP	4.278E+14	-1	0
IN2P3-CC	2.59E+16	2.64563E+16	0.978973
IN2P3-CPPM	1.4E+15	1.55122E+15	0.902516
IN2P3-LAPP	1.5E+15	1.51597E+15	0.989464
IN2P3-LPC	1E+15	1.07026E+15	0.934352
IN2P3-LPSC	7.303E+14	7.93972E+14	0.919805
INFN-BOLOGNA-T3	3E+12	-1	0
INFN-COSENZA	2.881E+14	3.10928E+14	0.92658
INFN-FRASCATI	5.788E+14	-1	0
INFN-GENOVA	2.3E+12	-1	0
INFN-LECCE	2.2E+12	-1	0
INFN-MILANO-ATLASC	9.075E+14	-1	0
INFN-NAPOLI-ATLAS	1.9E+15	1.91661E+15	0.991335
INFN-ROMA1	1.2E+15	-1	0
INFN-ROMA3	4.8E+12	-1	0
INFN-T1	1.94E+16	-1	0
JINR-LCG2	8.359E+14	8.55158E+14	0.97748
LRZ-LMU	1.3E+15	1.70614E+15	0.761953
MPPMU	1.3E+15	1.59816E+15	0.813436

Comparison of WSSA with the EGI Storage Accounting (for ATLAS).
For the sites which do have data in the EGI accounting, the agreement is pretty good
The excel comparison files for January for ATLAS are attached to the Accounting Task Force twiki page

WSSA monthly accounting reports

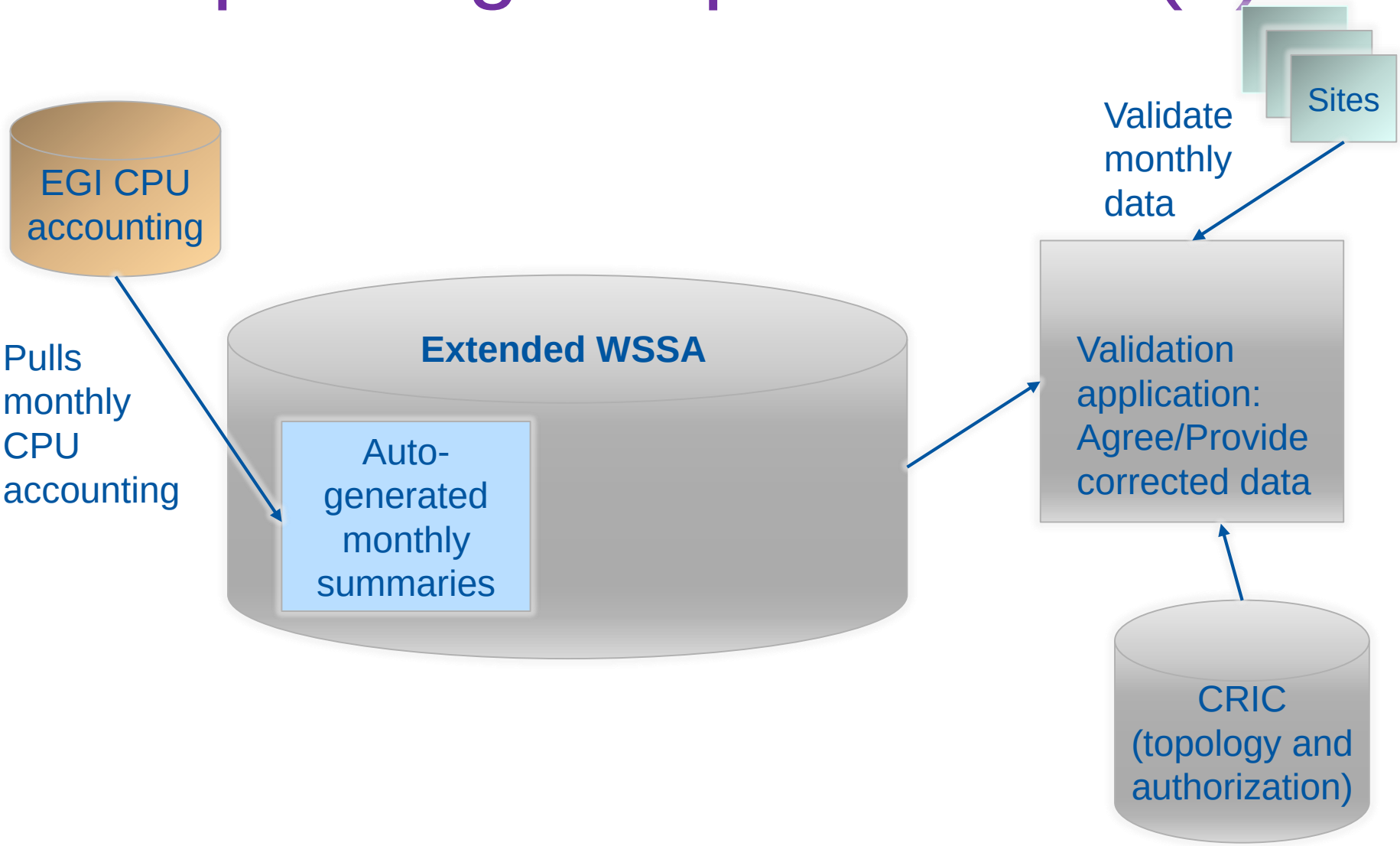
- WSSA accounting data is not yet integrated in the monthly accounting reports
- First we want numbers to be validated by site admins
- Non-official reports are sent monthly to the sites asking for feedback to

  wssa-notification@cern.ch
- Grateful to site admins who responded and helped us to detect and fix some issues

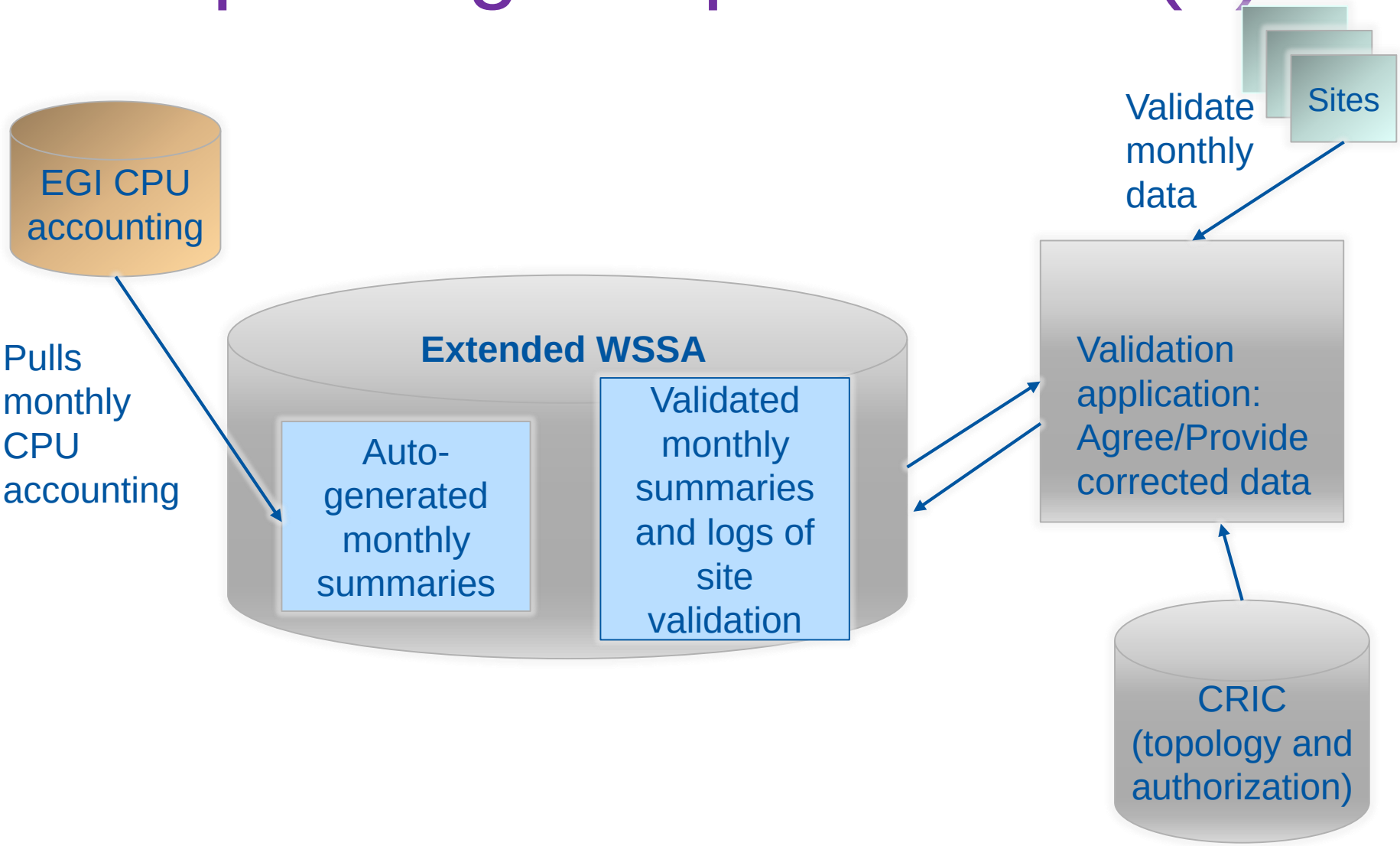
Validation applications won't help if sites are not checking them!

- Proper accounting information is site responsibility
- From the WLCG central operations side we put in place tools to detect wrong info, but we have no effort to follow up on every per-site inconsistency
- Checking two numbers in the beginning of each month is not a big deal, but we know **that it does not happen for all sites**

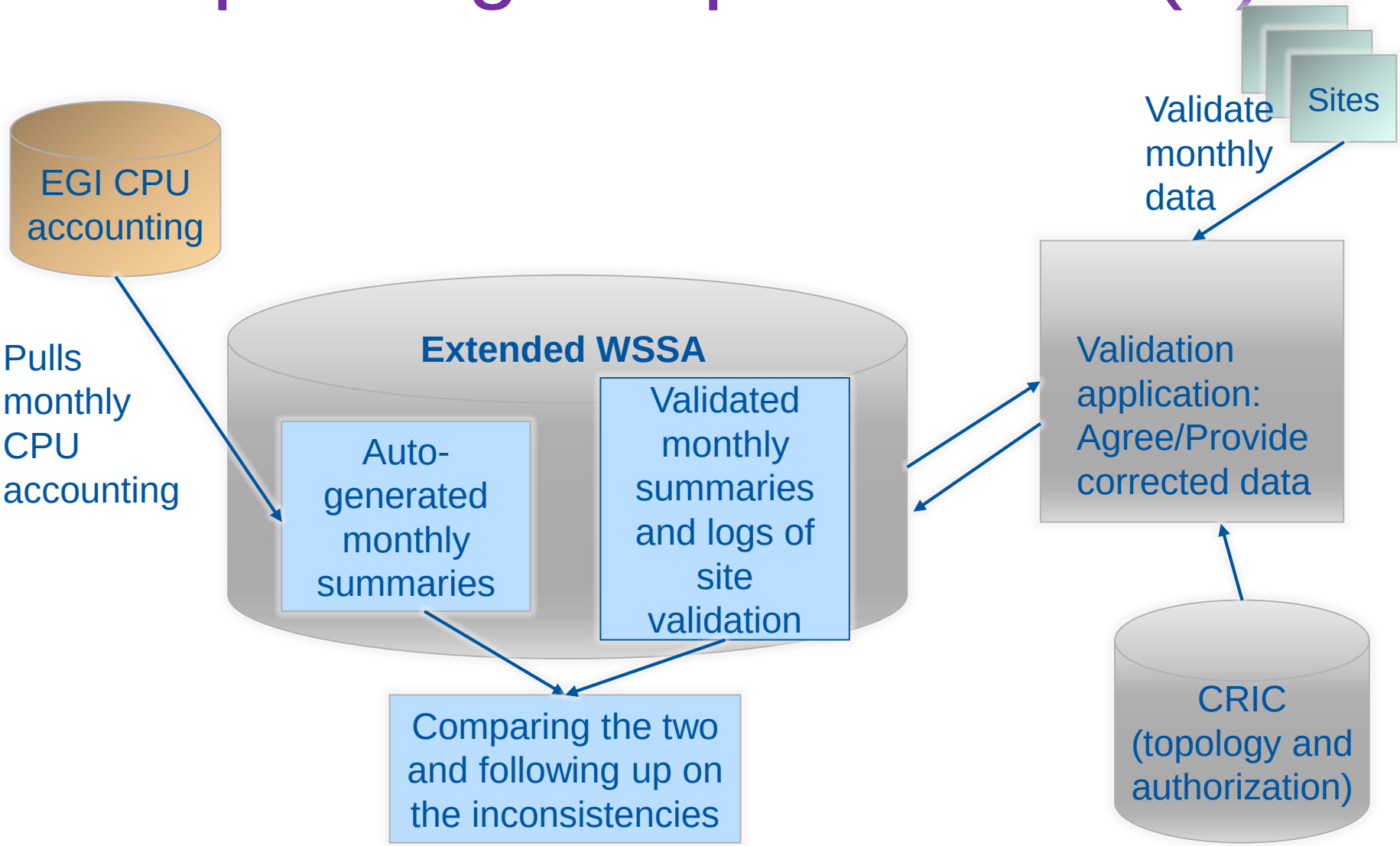
Improving the procedure (1)



Improving the procedure (2)



Improving the procedure (3)



Summary

- Setting up data validation procedures comparing data with alternative data sources allowed to detect and fix many issues on the sites
- However, in order to spot eventual problems, sites are required to check the UIs of the data validation application monthly as well as monthly accounting reports and follow up when issues are detected
- In order to improve further, we will provide the UI to validate/correct both storage and CPU auto-generated accounting data by site admins. Monthly accounting reports will be based on the approved data.
- The cases where auto-generated numbers are very different from the ones provided by site admins will be followed up