US CMS Tier-2 2012 Workshop

Introduction

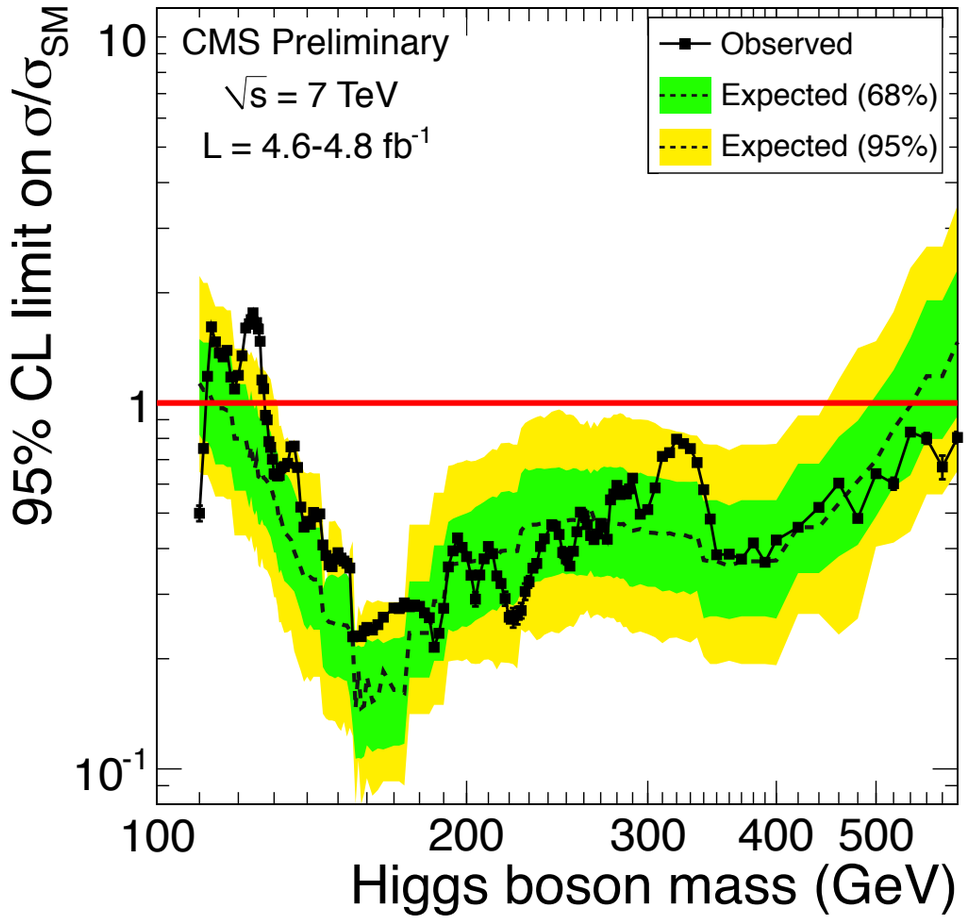
Ken Bloom

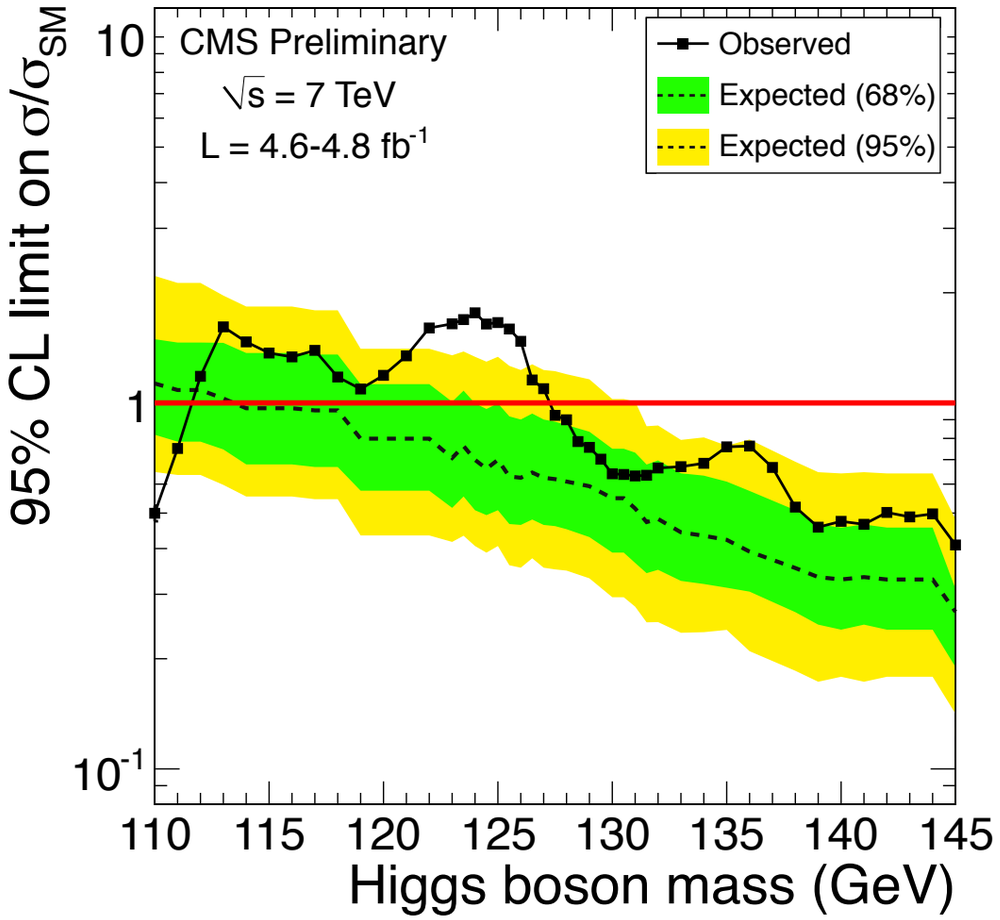
March 19, 2012

Welcome to Lincoln!

Another amazing year!

- CMS makes headlines around the world because of the physics that we do:

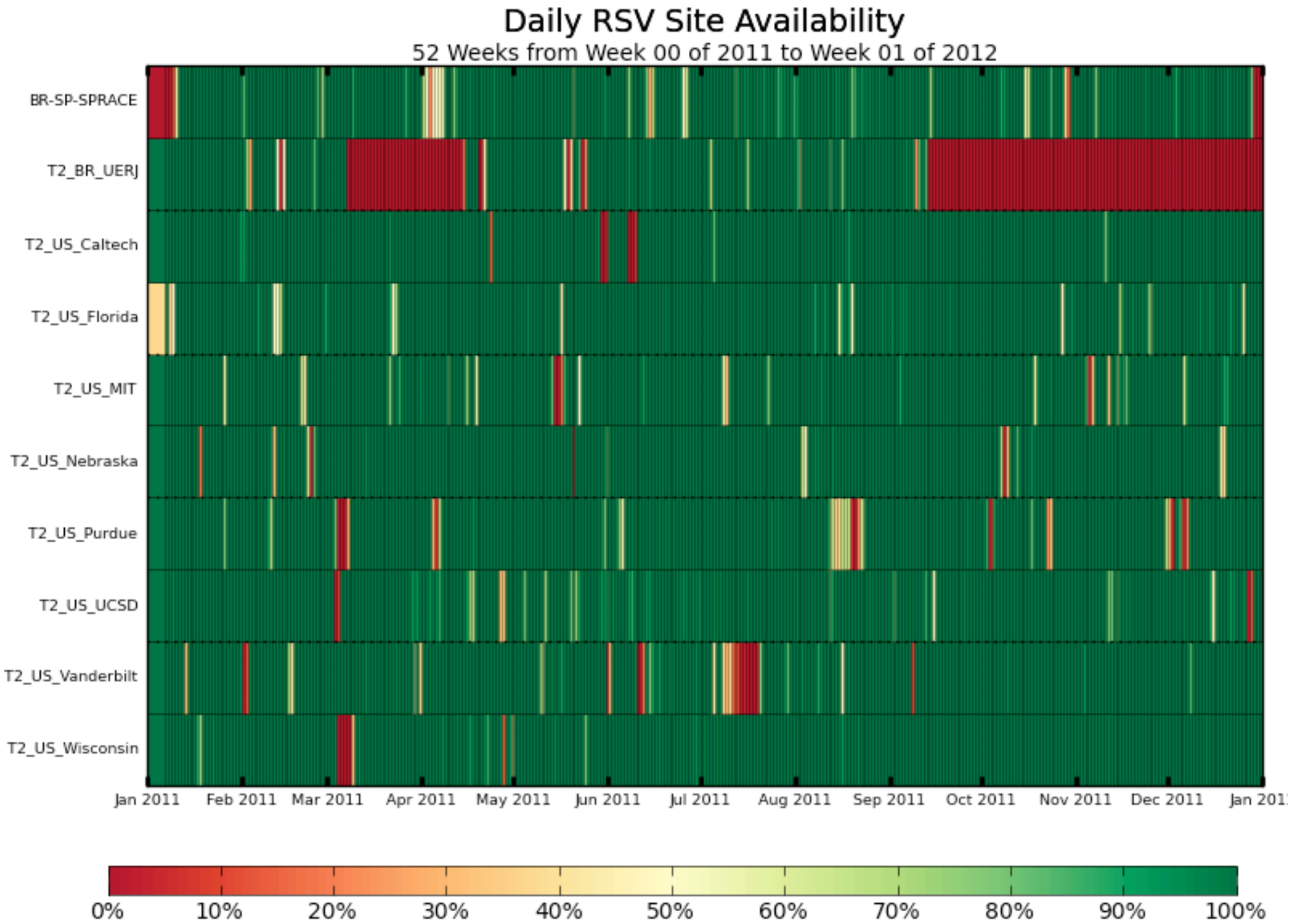




- But the physics output is predominantly driven by work done at the Tier-2 centers -- this would not be possible without you.
- And since the US has some of the biggest and best-run centers, we are particularly important for CMS.

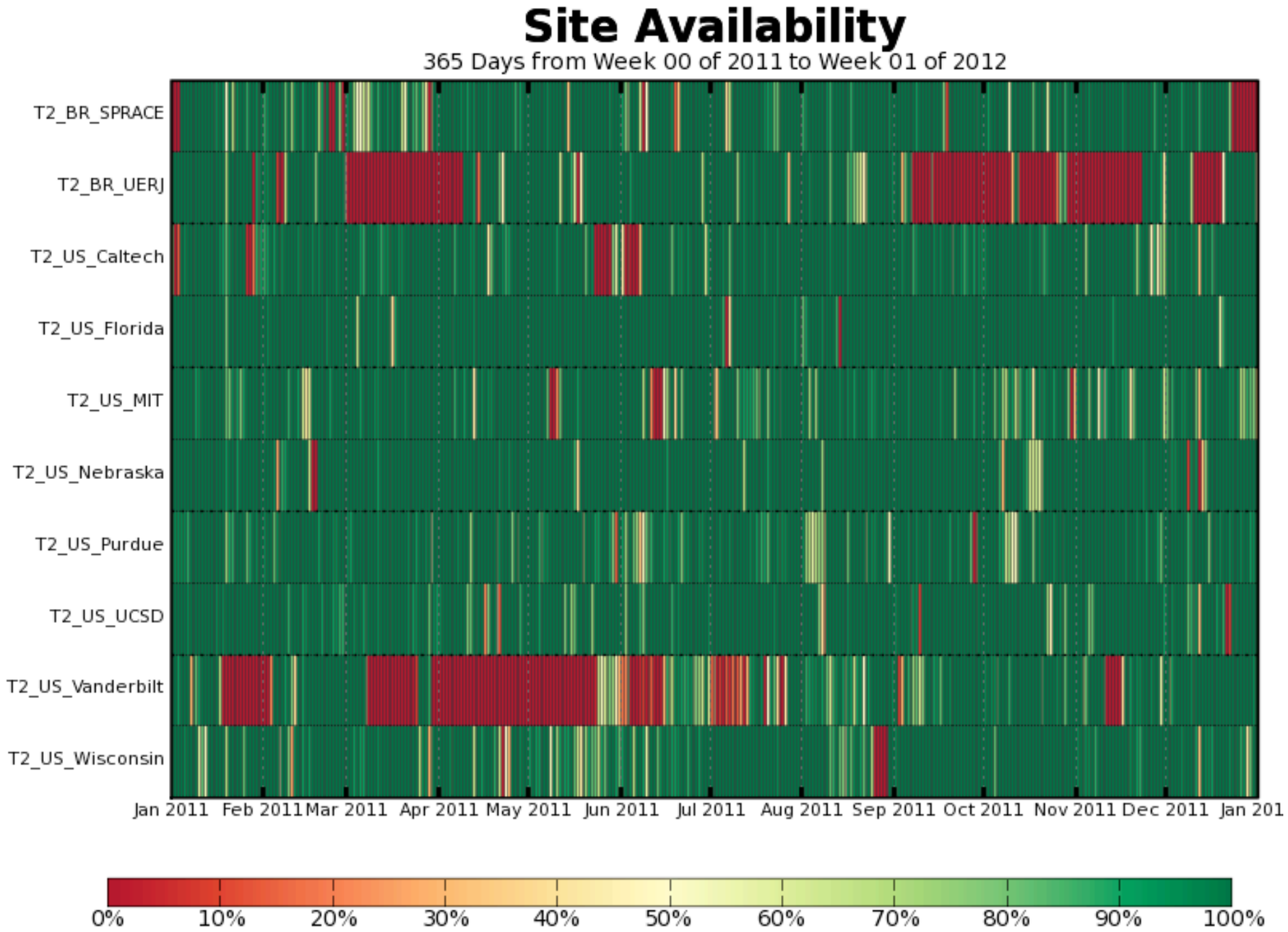
2011 in review: RSV

- The RSV results get turned into the WLCG availability/reliability numbers, which get reported to funding agencies.

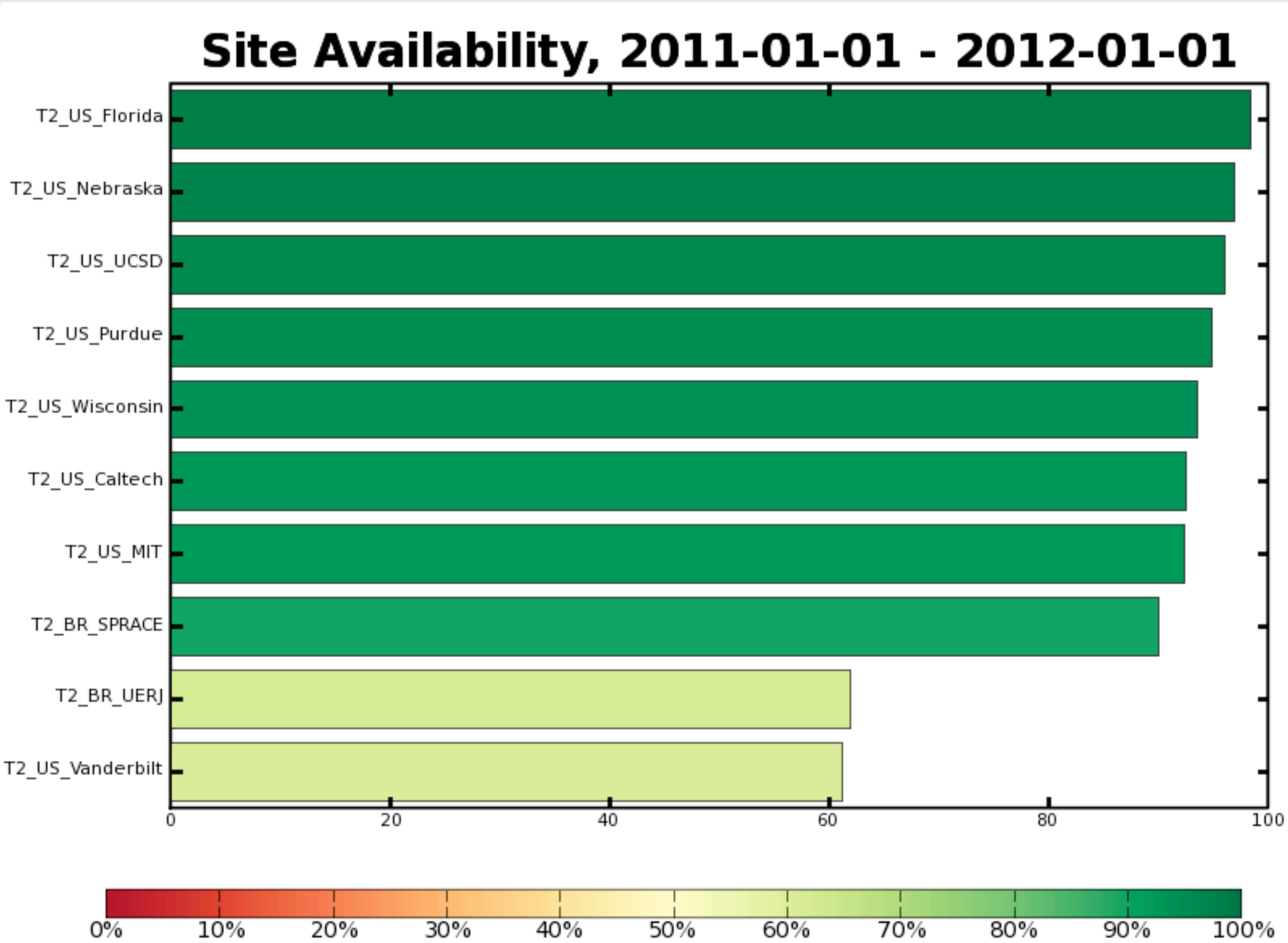


2011 in review: SAM

- SAM is now SUM!
  - I will concede that the rollout was not well publicized



2011 in review: SAM



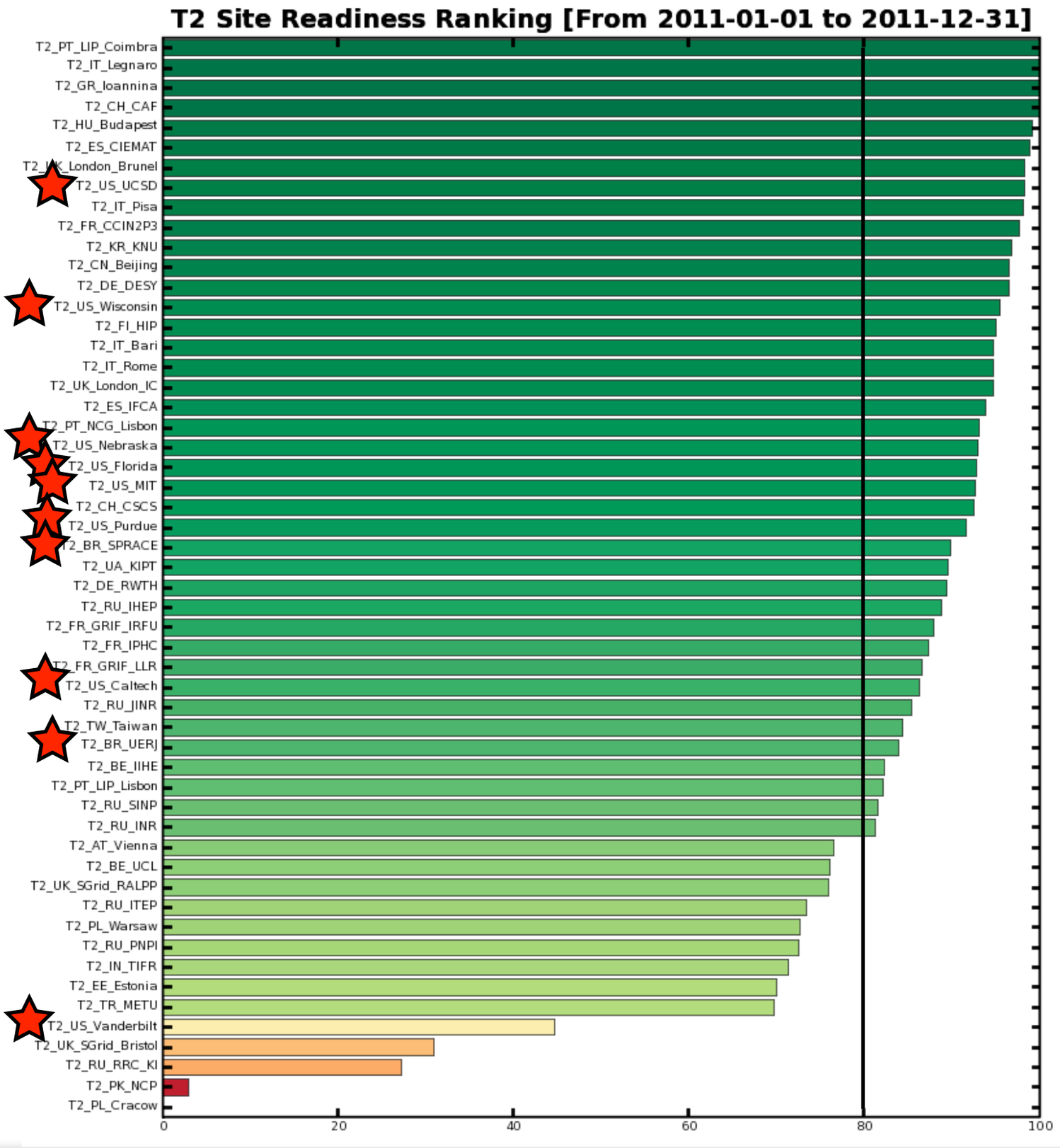
2011 in review: job robot

- The job robot, one of our oldest tools, is supposed to be replaced by Hammercloud "soon" -- the jobs are already running

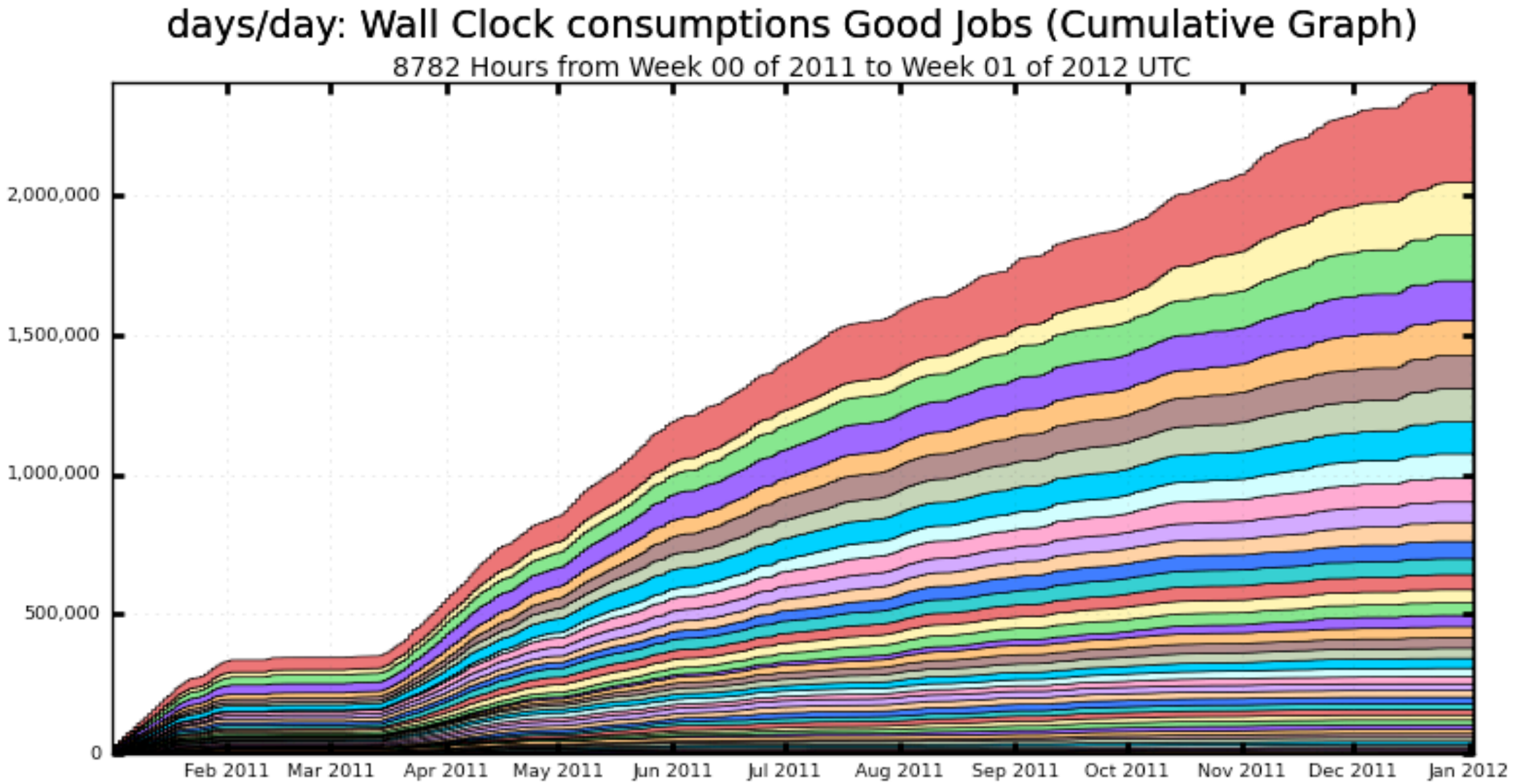
2011 in review: site readiness

- This is currently our gold standard for evaluating sites, although I think it's worth discussing what it does/doesn't capture

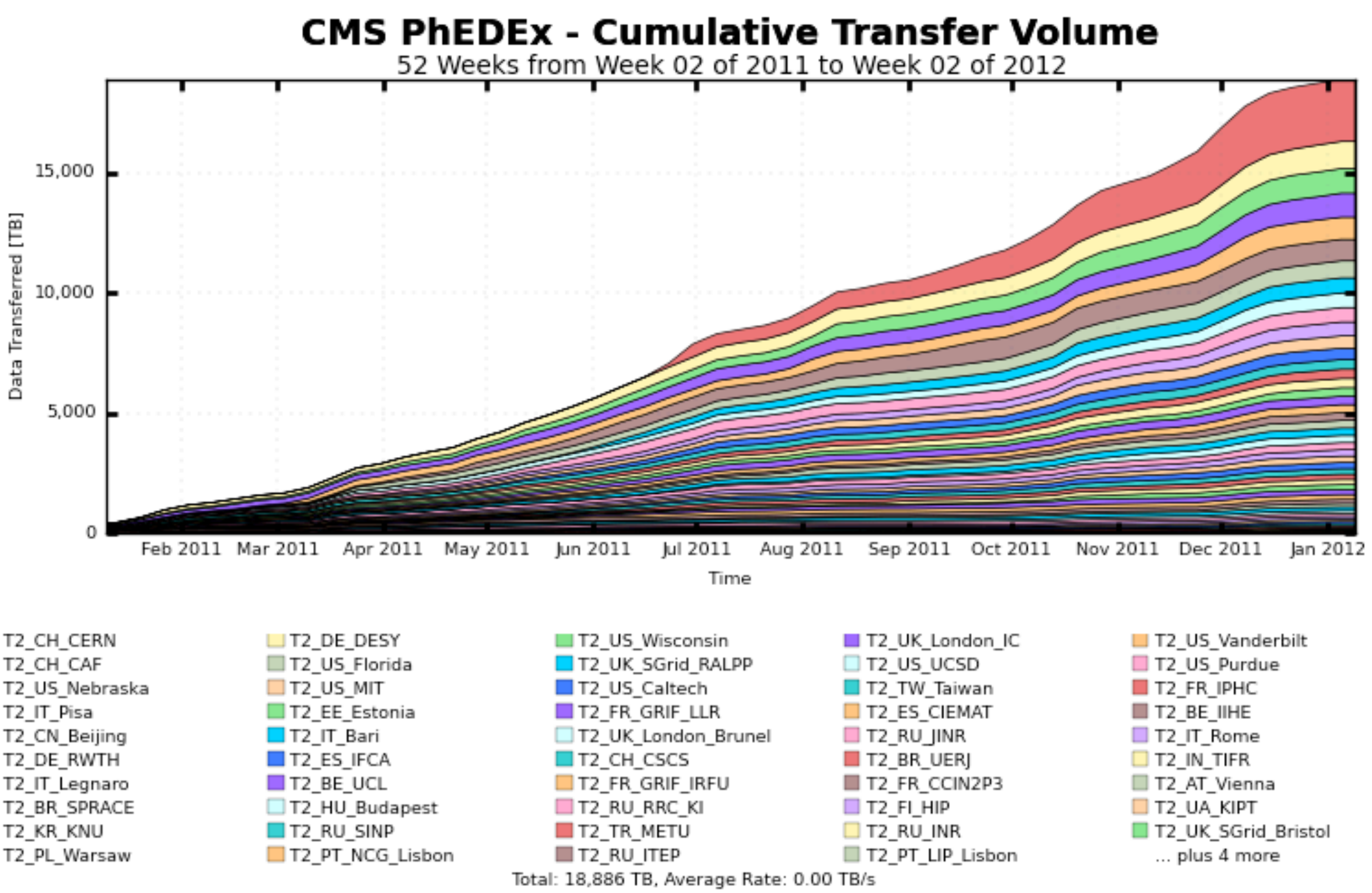
2011 in review: site readiness



2011 in review: production

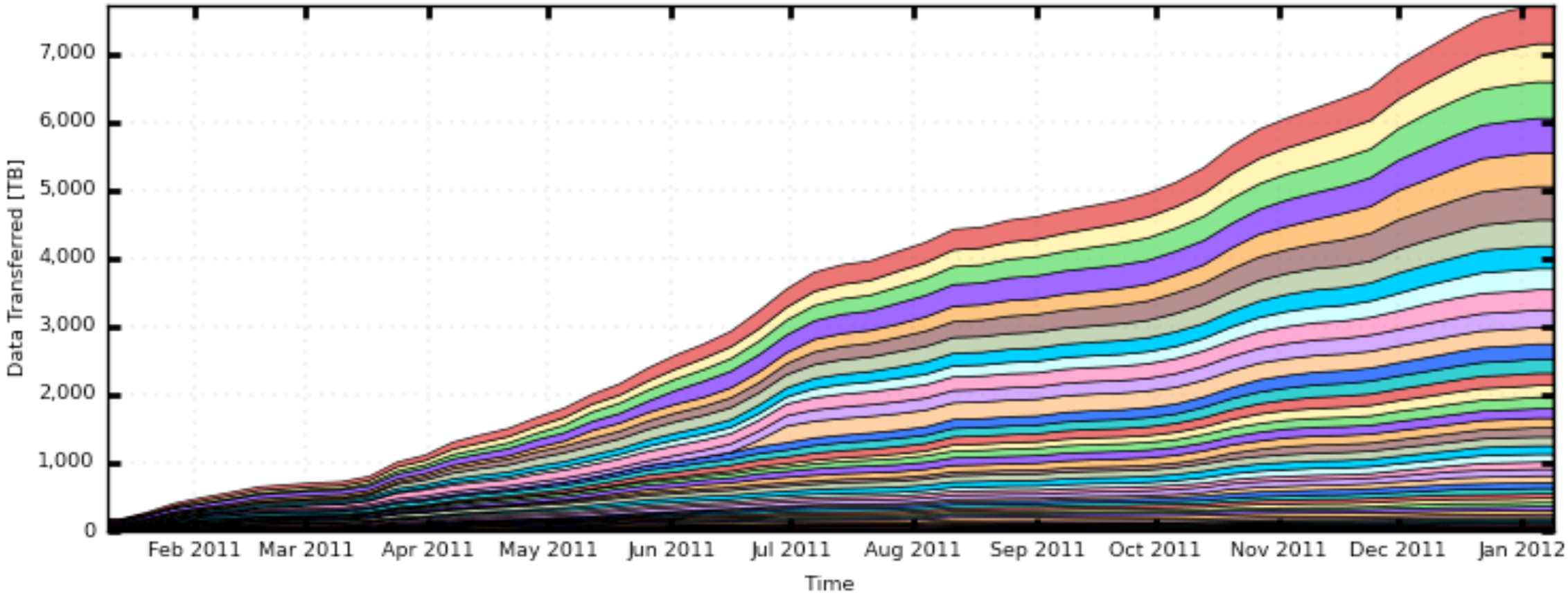


2011 in review: inbound transfers



- 4.9 of 18.9 TB was from other T2 sites

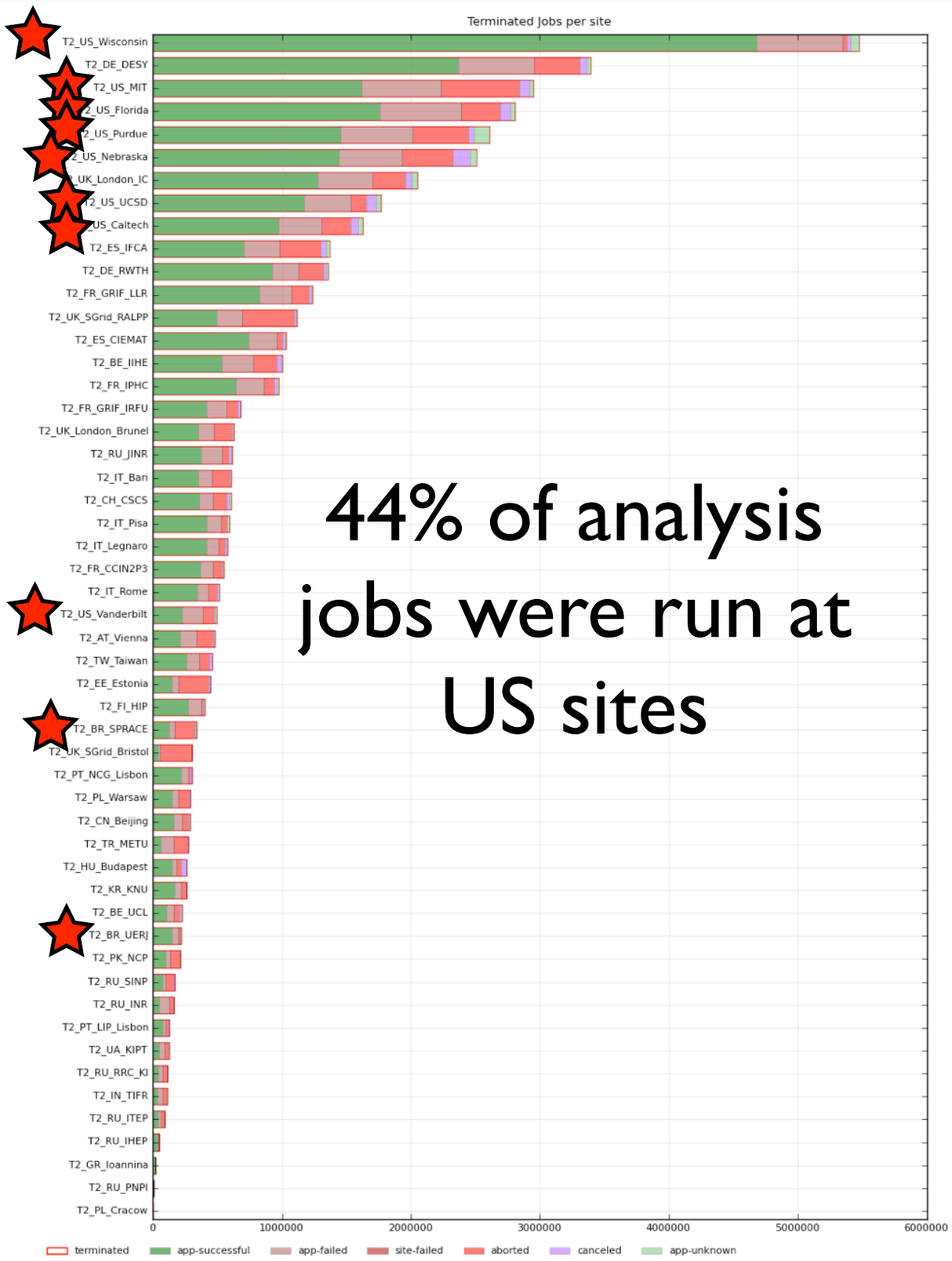
2011 in review: outbound transfers



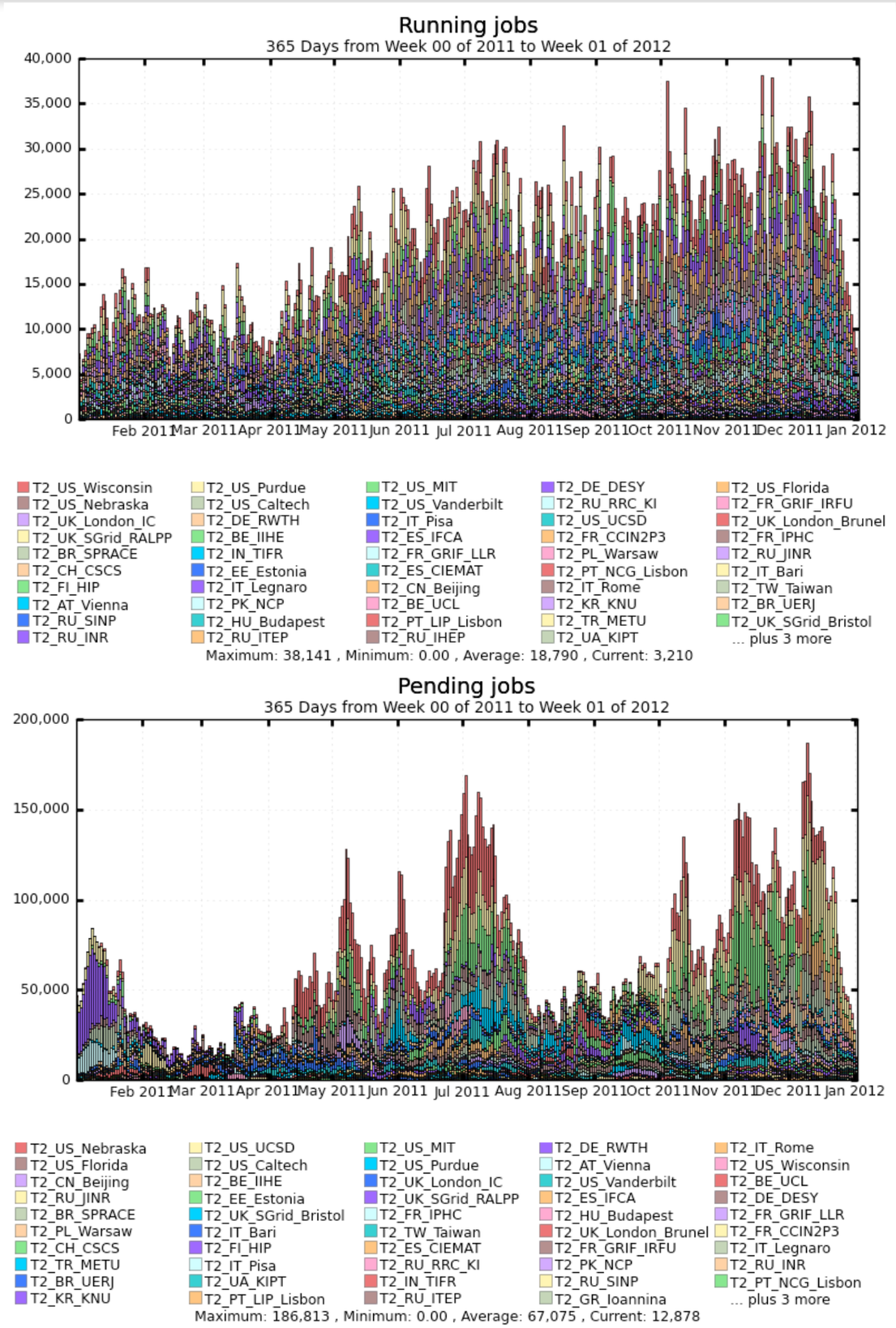
T2_US_Wisconsin, T2_DE_DESY, T2_US_Nebraska, T2_US_Purdue, T2_UK_London_IC
T2_US_UCSD, T2_US_Florida, T2_US_MIT, T2_US_Caltech, T2_DE_RWTH
T2_FR_GRIF_LLR, T2_CH_CAF, T2_ES_CIEMAT, T2_ES_IFCA, T2_FR_IPHC
T2_UK_SGrid_RALPP, T2_IT_Rome, T2_FR_CCIN2P3, T2_CH_CSCS, T2_BE_IIHE
T2_IT_Bari, T2_IT_Pisa, T2_IT_Legnaro, T2_FR_GRIF_IRFU, T2_RU_JINR
T2_UK_London_Brunel, T2_TW_Taiwan, T2_EE_Estonia, T2_BE_UCL, T2_FI_HIP
T2_US_Vanderbilt, T2_HU_Budapest, T2_CN_Beijing, T2_AT_Vienna, T2_KR_KNU
T2_CH_CERN, T2_BR_SPRACE, T2_PT_NCG_Lisbon, T2_TR_METU, T2_BR_UERJ
T2_IN_TIFR, T2_UA_KIPT, T2_RU_RRC_KI, T2_RU_INR, T2_PL_Warsaw
T2_RU_SINP, T2_RU_IHEP, T2_RU_ITEP, T2_PT_LIP_Lisbon, ... plus 3 more

Total: 7,713 TB, Average Rate: 0.00 TB/s

2011 in review: analysis job hosting

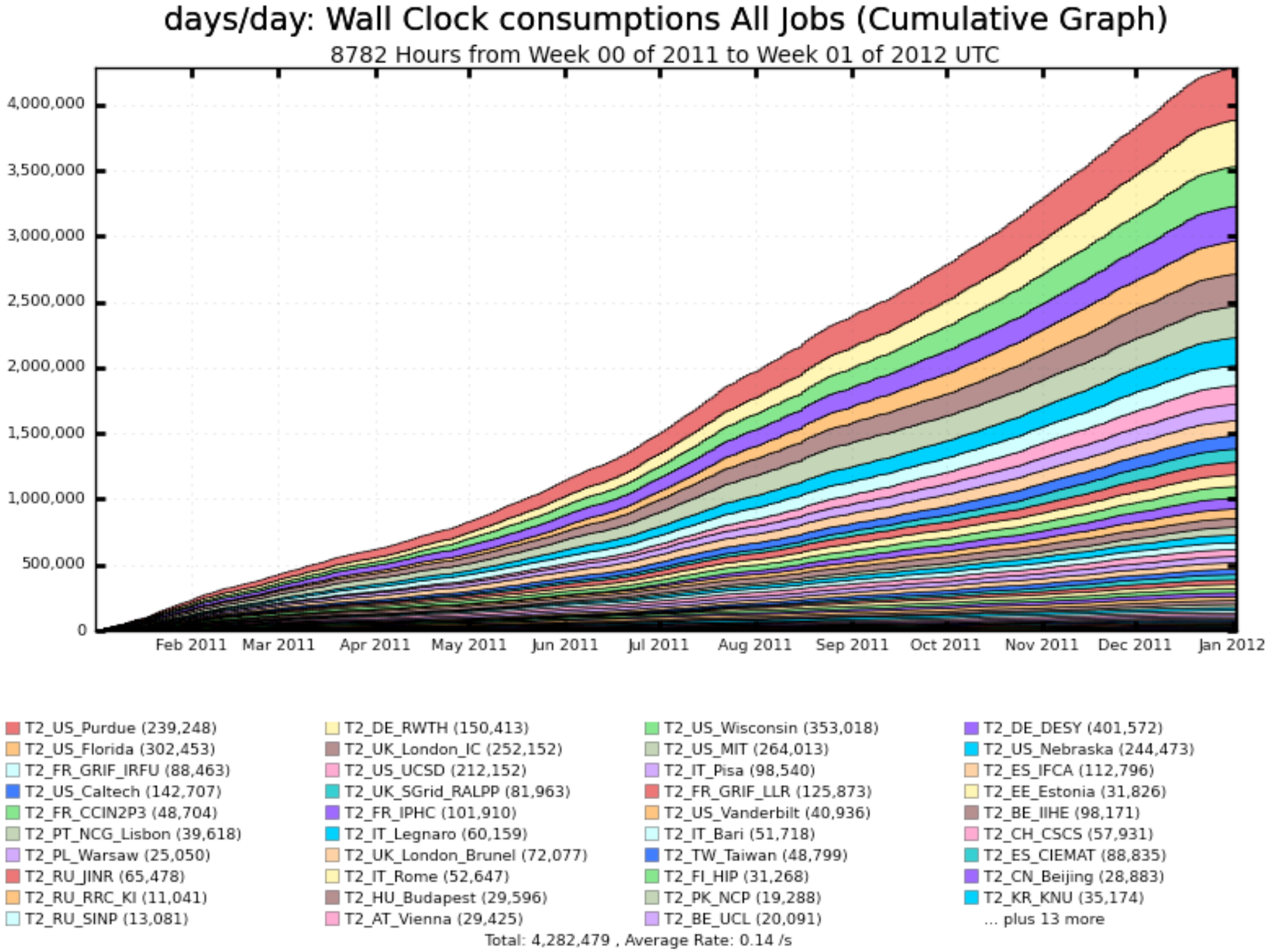


44% of analysis jobs were run at US sites

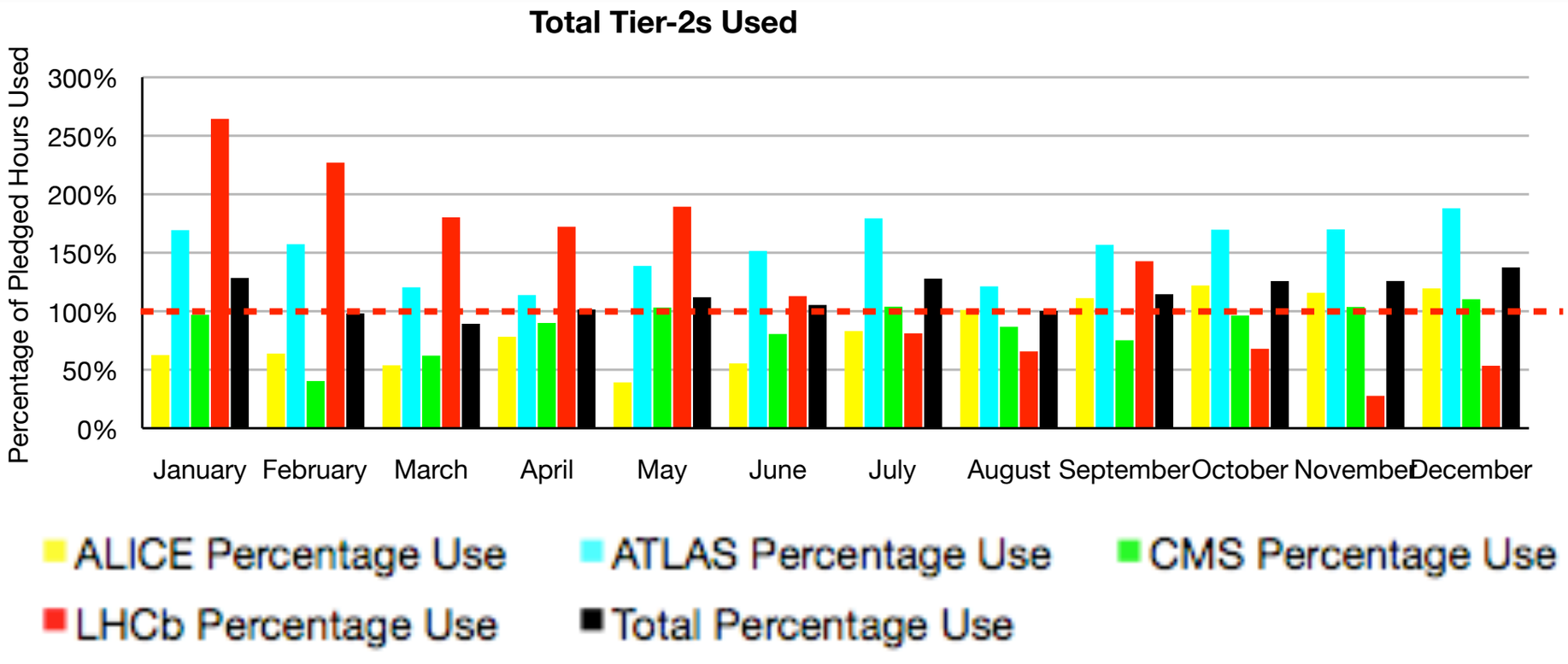


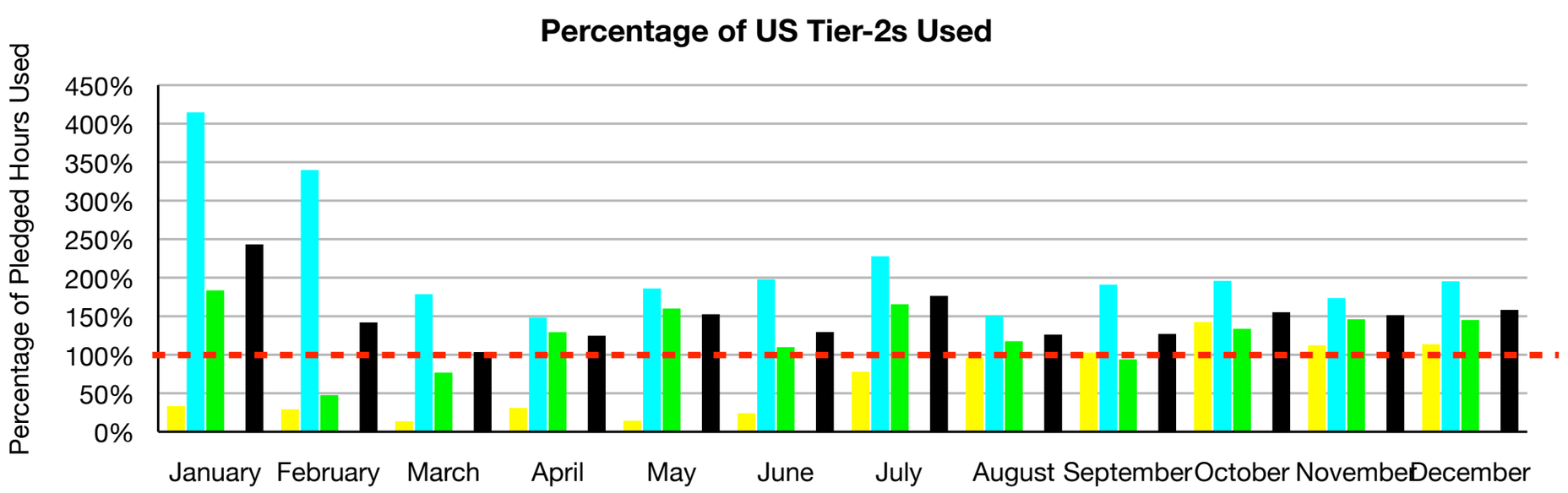
2011 in review: analysis job hosting

- I wonder if this is more meaningful than job count. Would be nicer as a pie chart, perhaps. But legend doesn't agree with plot....



2011 in review: broader context





- T2 CPU usage 88% of pledge for CMS overall, 126% of pledge for US T2 sites

Deployment

Site	CPU (HS06)	Batch Slots	Space for hosting (TB)	WAN (Gb/s)	Last update
T2_US_Caltech	16800	1662	937	20	12/17/11
T2_US_Florida	17140	2445	927	10	1/27/12
T2_US_MIT	22150	3200	855	10	4/26/11
T2_US_Nebraska	23250	3100	1180	10	2/2/12
T2_US_Purdue	46095	5200	960	10	9/20/10
T2_US_UCSD	20790	2544	971	2 x 10	3/09/12
T2_US_Wisconsin	29600	3300	1200	20	11/08/11
Total	175825	21451	7030		

- The US pledge for 2012 is 12500 HS06, 1000 TB/site
  - We have pledged a smaller fraction of CMS resources this year
  - Thus we met CPU pledge last year! And almost there on disk....
- Given that, I would like to "lap" ourselves this year and get to the 2013 pledge in this year if possible
  - Probably 14% more CPU, no increase in disk -- should be doable just with replacing machines at the end of their lifecycle

Deployment

Site	CPU (HS06)	Batch Slots	Space for hosting (TB)	WAN (Gb/s)	Last update
T2_US_Caltech	16800	1662	937	20	12/17/11
T2_US_Florida	17140	2445	927	10	1/27/12
T2_US_MIT	22150	3200	855	10	4/26/11
T2_US_Nebraska	23250	3100	1180	10	2/2/12
T2_US_Purdue	46095	5200	960	10	9/20/10
T2_US_UCSD	20790	2544	971	2 x 10	3/09/12
T2_US_Wisconsin	29600	3300	1200	20	11/08/11
Total	175825	21451	7030		

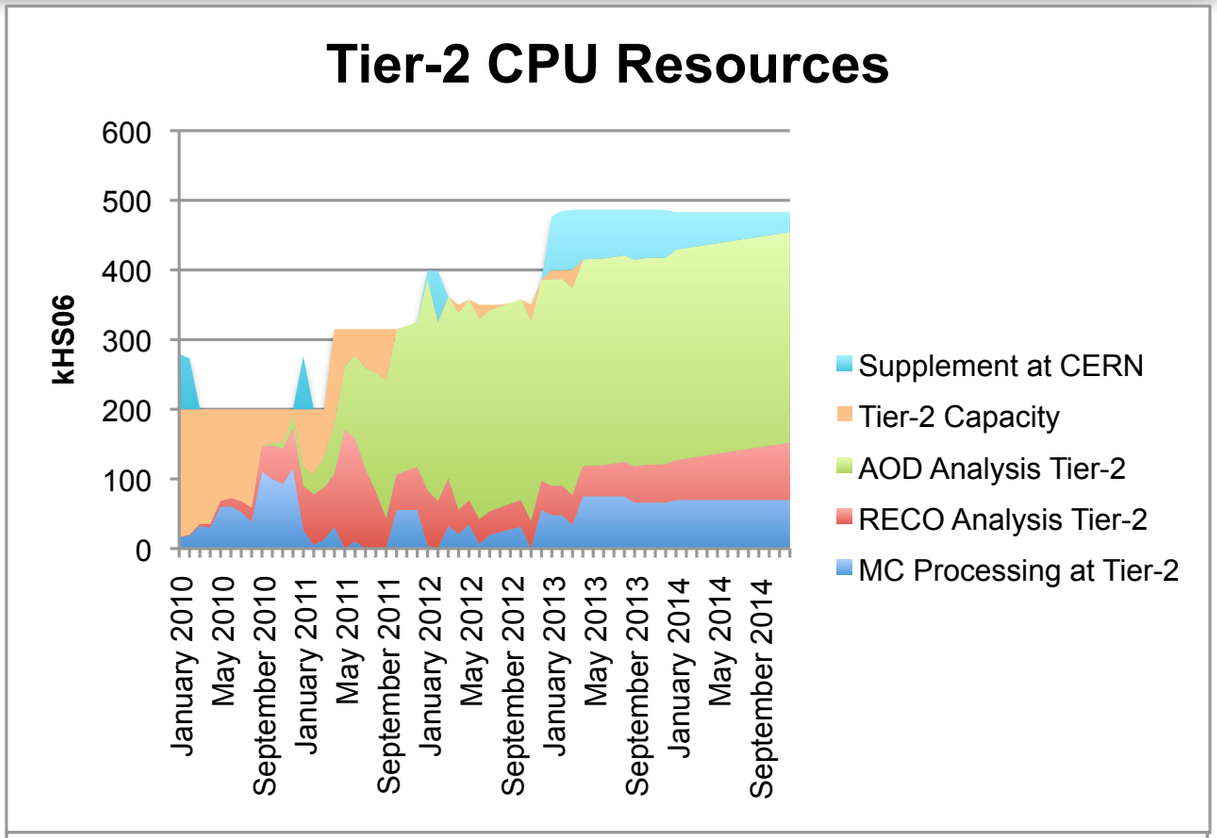
- Nevertheless, these are "build-to-cost" facilities, and thus will deploy as much hardware as we can afford
  - We need to evaluate the CPU/disk balance
- We are being asked: why aren't site deployments more equal?
  - Yes, the disk is fairly equal, but the CPU is not
- Studying the costs and benefits of each site is a priority for us this year, and could affect future funding distribution [Bauerdick]

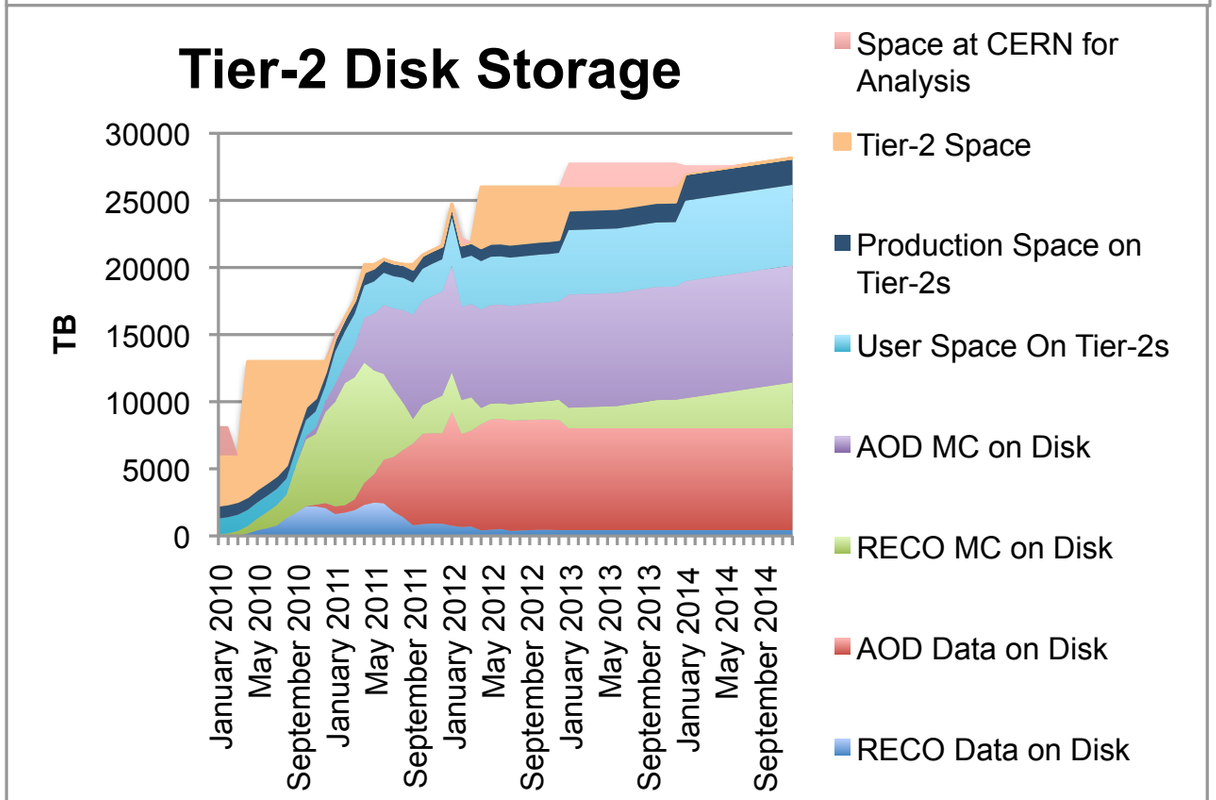
Looking ahead at CMS

- This year, more pile-up = more complex events = larger events, more time to process, harder data analysis
  - Strong efforts by CMS to ameliorate this in software
- After 2012, the LHC will be down for two years for upgrades that will allow a higher collision energy
  - No new data until 2015! Thus, a great thirst to record data at a higher rate than was originally envisioned
- Computing management has responded by proposing two new kinds of datasets:
  - Short-term parking: reconstruct at T1 instead of T0
  - Long-term parking: don't reconstruct until next year!
  - These together would be about as many events as the prompt reconstruction (if there is sufficient demand/physics case)

Impacts at Tier 2

- Extra data → more T2 analysis
  - Get to use T0 for analysis activities in 2013, but still need to increase T2
- T2 disk still constrained
  - Very successful move of analysis to AOD rather than RECO, assume 95% of analysis on AOD
  - Previously assumed all data equally interesting; revise model to reduce number of copies
- Can still improve T2 modeling
  - Does it scale with number of events or number of users?





Challenges and Evolution

- All that being said, I believe that we have demonstrated that we can scale up facilities at a rate that matches the resource needs
  - The exception might be in disk space -- if sites are constrained on this, we really must understand why
- I think the greater challenge is keeping up with an evolving operating environment, on many, many fronts:
  - CMS computing operations, e.g. CVMFS, data consistency [Gutsche]
  - Moving to SL6 [Lundstedt/Attebury]
  - Changes to OSG [Roy]
  - New/better job submission systems [Dost, Melo]
- And in general, we're always faced with the challenge of operating our sites as efficiently and successfully as possible

Notes and comment this month

- Speaking of evolving systems, this has been a tough few weeks for sites:
  - Upgrade to OSG 3, especially GRAM5, has exposed some unexpected bugs
  - The switch to EMI in Europe at the same time hasn't helped matters
  - Not to mention the switch from SAM to SUM
- Given all this, we need as much feedback from sites and users as possible to help us debug the newest tools
  - Preferably before CMS starts analyzing 2012 data in earnest....
- The new operations organization is becoming more interested in the use of disk space at the sites
  - We must make sure that local users are not taking up resources that are designated for general use by CMS.

This week's meetings

- We've tried to set up the workshop to allow you to have more engagement with OSG colleagues. Other sessions of interest:
  - Monday afternoon: Federated data, "greatest hits" from sites
  - Monday evening: Lincoln Community Concert Band, 7:30 PM, College View Church at 48th and Prescott
  - Tuesday morning: Campus grids and clouds
  - Tuesday afternoon: WLCG technology evaluation groups, CMS T3 (includes some things that didn't fit in this morning)
  - Wednesday: OSG plenaries, including digital humanities

Perspectives

- The US CMS Tier-2 program has now been in progress for nearly seven years
  - We didn't know it was going to be as successful as it has been!
- We held the second-ever workshop for sites in Lincoln in 2006
  - How many of you were here for that?
- I and the rest of CMS have really appreciated and respected all of the great work being done at the sites
  - Both in operations and in technology development that has really taken us in some unexpected directions
  - We believe that there is real value in the intellectual engagement that we get from the universities involved
- It continues to be a pleasure for me to work with all of you