Xrootd Monitoring & Data Popularity (the CMS experience)

Maria Girone, Domenico Giordano
(CERN IT-ES)

Creating Federated Data Stores For the LHC
22/11/2011

Introduction

We have developed the CMS Popularity Service that tracks over time the official data accessed by the users on WLCG

The CMS Coll. has expressed interest in monitoring also the popularity of the data accessed via Xrootd

Outline

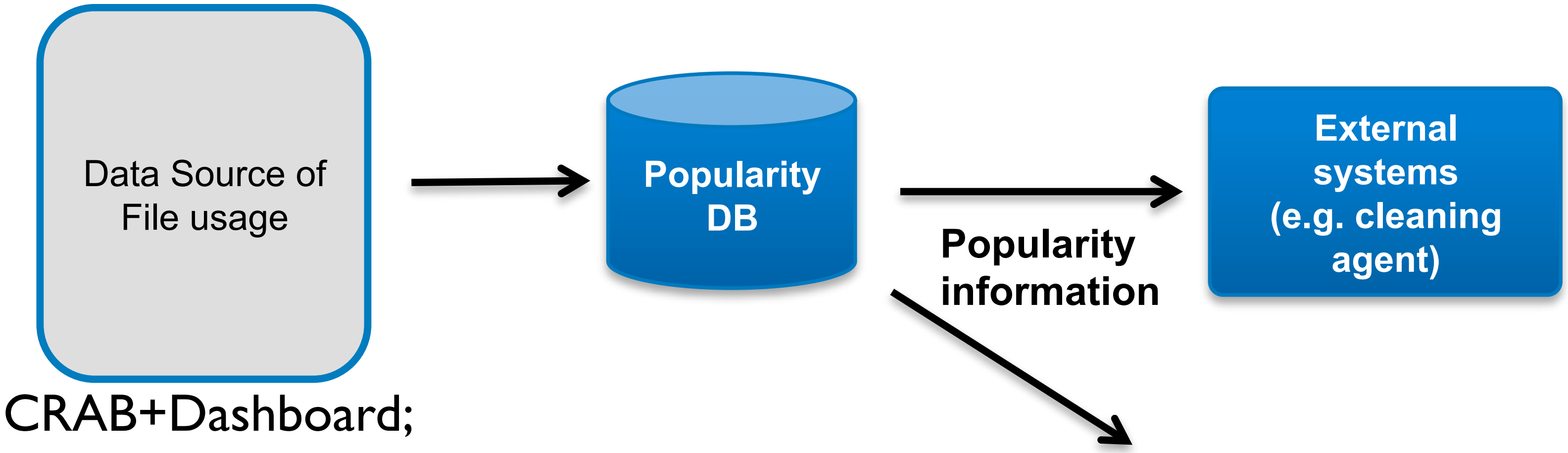
- The concept of the Data Usage Popularity
- The CMS experience based on the distributed analysis tools (CRAB)
  - Architecture, Operation Statistics, Metrics
- The Xrootd use case
  - Architecture, Strategies
- Conclusions

Definition

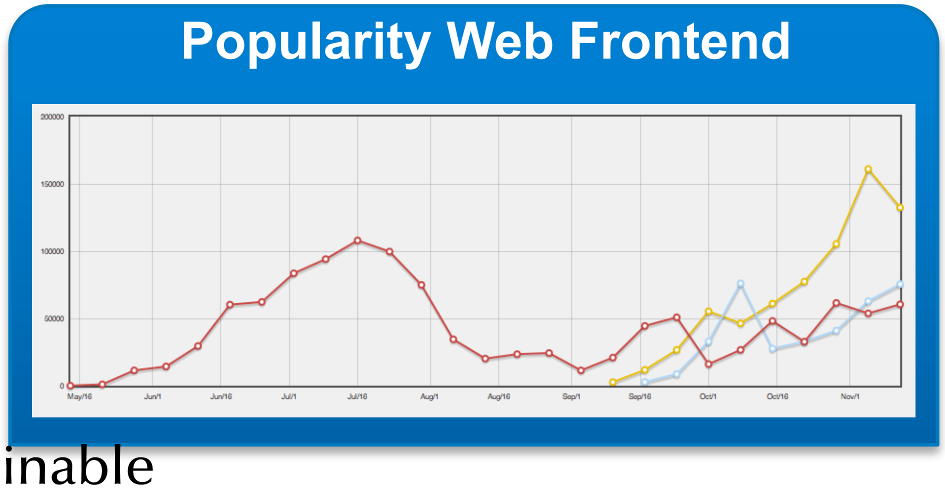
The purpose of a Data Popularity service is to provide

- Usage statistics along time about files / blocks / datasets accessed by the users
  - Information in terms of number of accesses, file access success/failure, CPU hours of processing, dataset name, number of users
  - Traces the evolution in time of the (un)needed data
- data service for further applications
  - eg. a site cleaning agent

Architecture


Data Source of File usage
Popularity DB
Popularity information
External systems (e.g. cleaning agent)

CRAB+Dashboard;
Xrootd;
others



Philosophy:

- do not duplicate existing data sources
- keep the overall design simple and maintainable

Components

- the heart of the system is the Popularity DB (PopDB)
  - Allows collection of the file based data extracted from datasource
  - Aggregate at the level of blocks/datasets, Correlate with other attributes (site, user, #users, # CPU hours)
- Lightweight web layer
  - implements multiple interfaces: json API, plots, tables

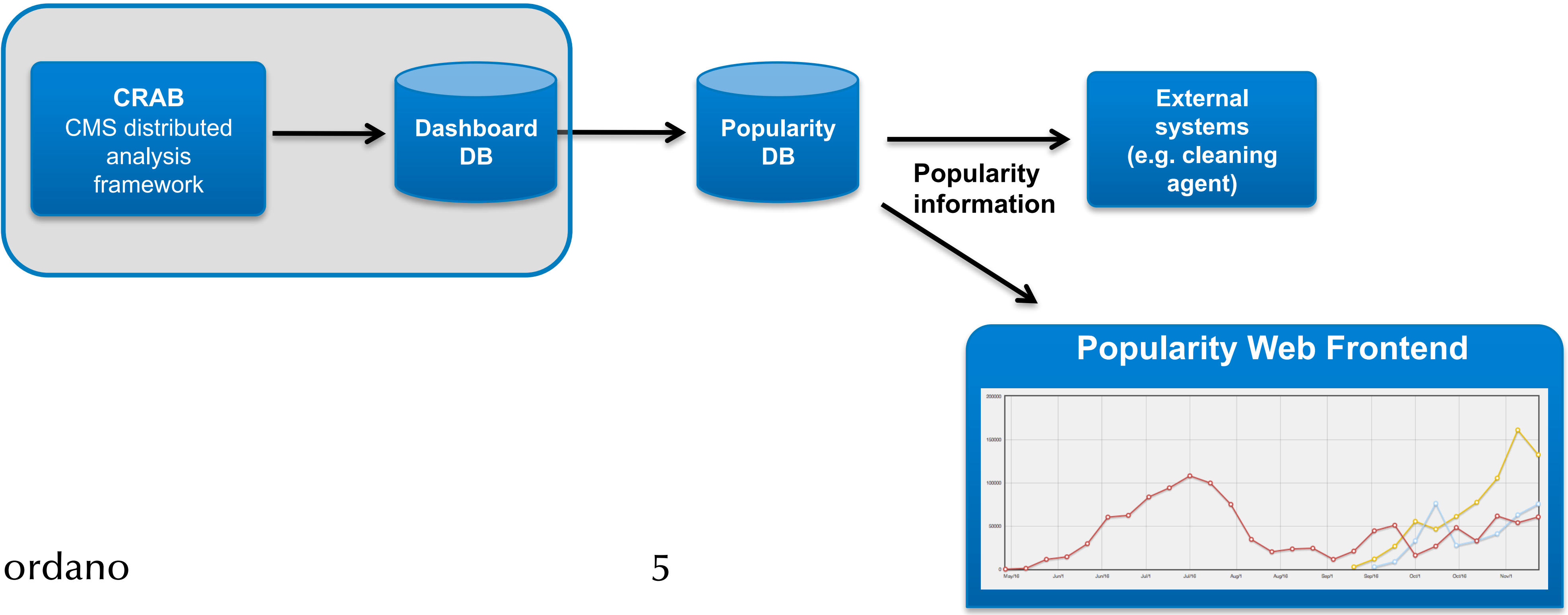
First implemented use case

Data Popularity Service based on distributed analysis tools

- Developed by IT/ES for CMS: based on CRAB + Dashboard
- Allows to define the Data Popularity based on the user activity on WLCG
  - # accessed input files, # files per job, CPU time per file, file access exit status, lumi sections
  - Distinct per site and datasets

Limitation: collects ONLY statistics from CRAB jobs

- No hint about activity on local batch systems or interactive jobs



Operation, Statistics & Throughput

The CRAB-based Popularity workflow is steadily collecting data since 6 months

- Amount of data uploaded:
  - O(90k) jobs/day , O(360k) files/day
    - NB: jobs accessing official files/dataset (user collections not incl.)
- Speed of the procedure
  - Raw-Table upload: O(50 min)/day , MV update: O(3min)/day
- Size of the Table
  - Raw-Table: O(30 GB), MVs: O(600 MB)

So far no problems/bottlenecks found in the procedure (CRAB → Dashboard → PopDB → MaterializedViews)

ES

DataSet Statistics

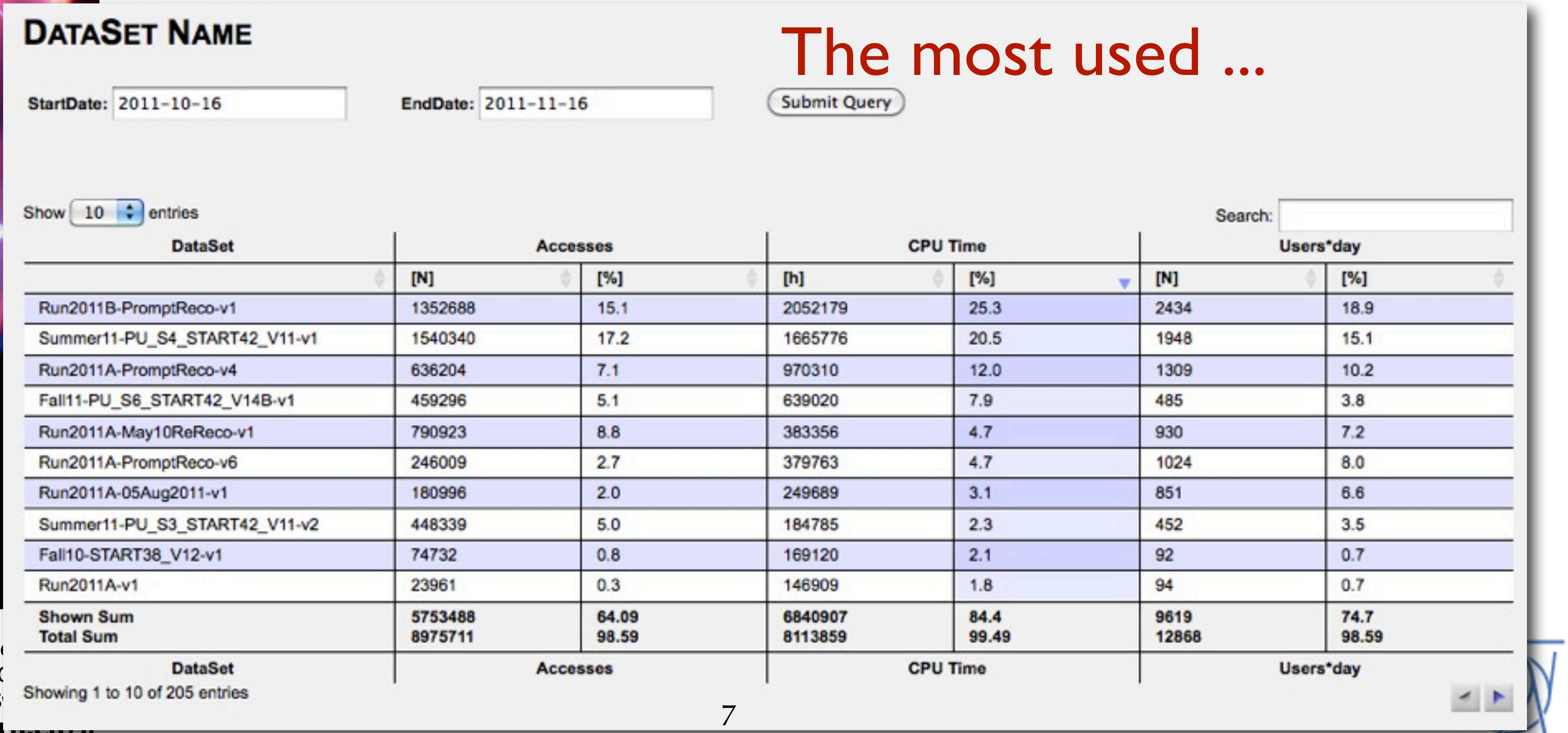
CERN IT Department

Absolute and relative metrics

- Configurable time window aggregation

Identify the most (and less) used DataSets

- Inclusively (integrate on all DataTier and PD)
- Exclusively, per each DataTier and PD



The most used ...

DataSet Name

StartDate: 2011-10-16 EndDate: 2011-11-16 Submit Query

Show 10 entries Search:

DataSet	Accesses		CPU Time		Users*day	
	[N]	[%]	[h]	[%]	[N]	[%]
Run2011B-PromptReco-v1	1352688	15.1	2052179	25.3	2434	18.9
Summer11-PU_S4_START42_V11-v1	1540340	17.2	1665776	20.5	1948	15.1
Run2011A-PromptReco-v4	636204	7.1	970310	12.0	1309	10.2
Fall11-PU_S6_START42_V14B-v1	459296	5.1	639020	7.9	485	3.8
Run2011A-May10ReReco-v1	790923	8.8	383356	4.7	930	7.2
Run2011A-PromptReco-v6	246009	2.7	379763	4.7	1024	8.0
Run2011A-05Aug2011-v1	180996	2.0	249689	3.1	851	6.6
Summer11-PU_S3_START42_V11-v2	448339	5.0	184785	2.3	452	3.5
Fall10-START38_V12-v1	74732	0.8	169120	2.1	92	0.7
Run2011A-v1	23961	0.3	146909	1.8	94	0.7
Shown Sum Total Sum	5753488 8975711	64.09 98.59	6840907 8113859	84.4 99.49	9619 12868	74.7 98.59
DataSet	Accesses		CPU Time		Users*day	

Showing 1 to 10 of 205 entries

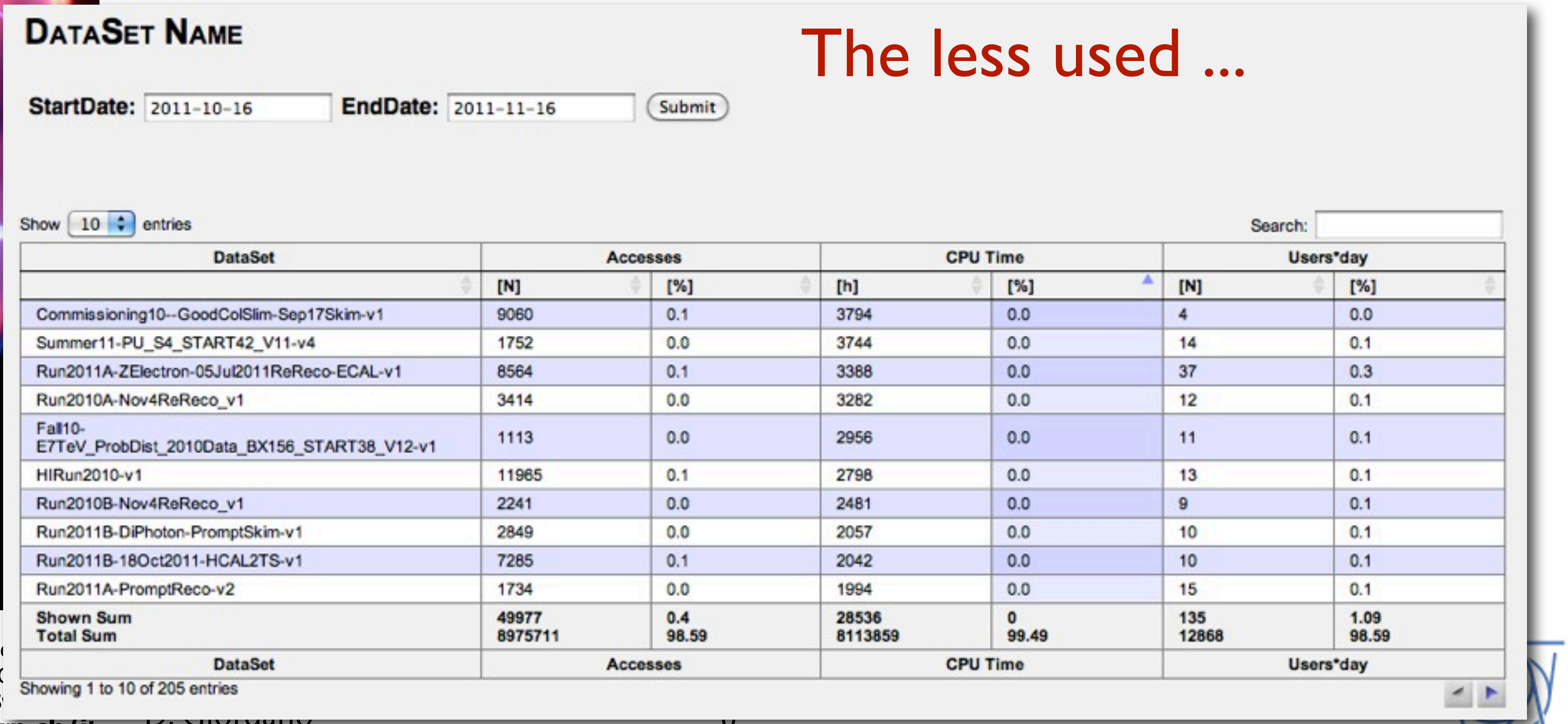
DataSet Statistics

Absolute and relative metrics

- Configurable time window aggregation

Identify the most (and less) used DataSets

- Inclusively (integrate on all DataTier and PD)
- Exclusively, per each DataTier and PD



The less used ...

DataSet Name

StartDate: 2011-10-16 EndDate: 2011-11-16 Submit

Show 10 entries

Search:

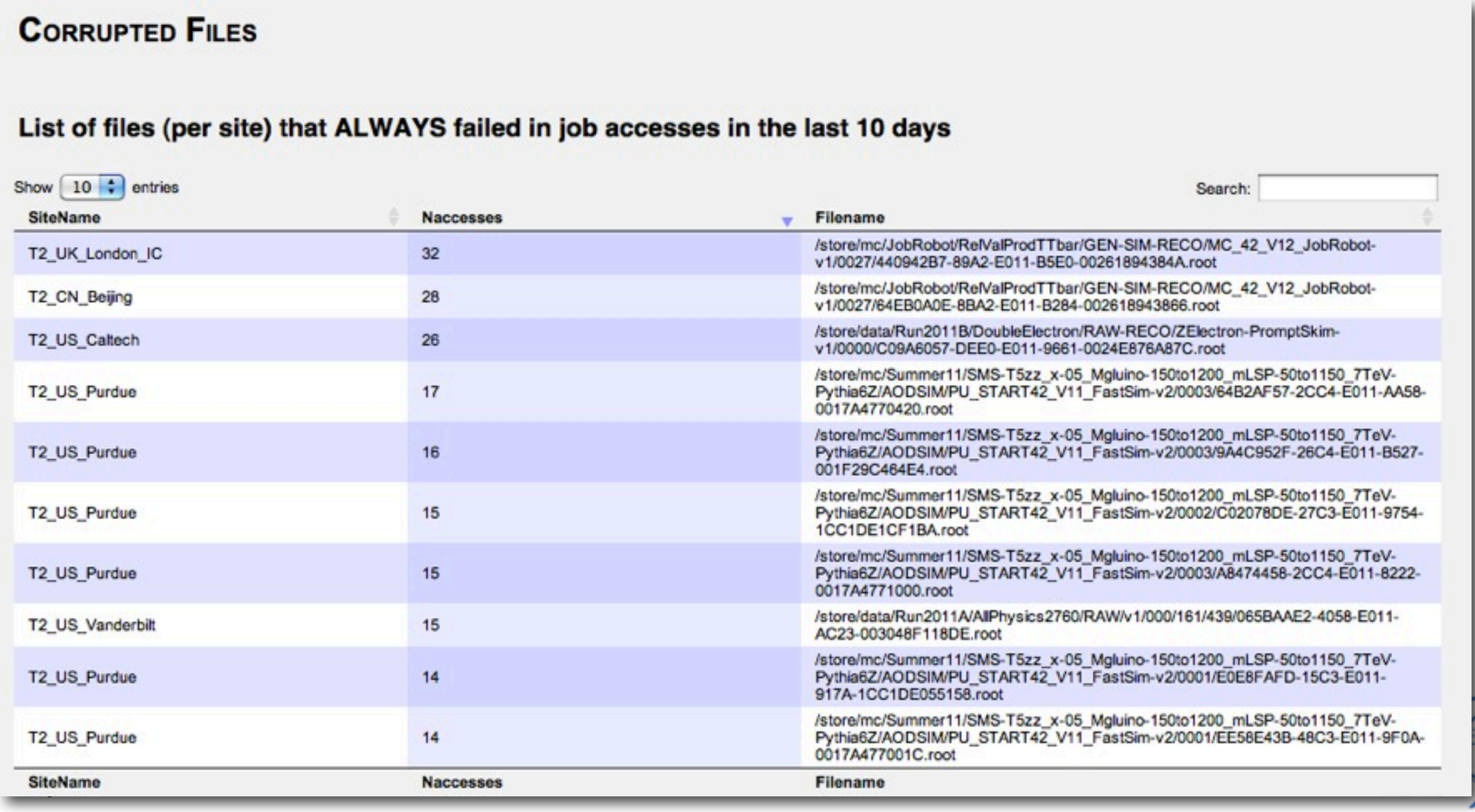
DataSet	Accesses		CPU Time		Users*day	
	[N]	[%]	[h]	[%]	[N]	[%]
Commissioning10--GoodColSlim-Sep17Skim-v1	9060	0.1	3794	0.0	4	0.0
Summer11-PU_S4_START42_V11-v4	1752	0.0	3744	0.0	14	0.1
Run2011A-ZElectron-05Jul2011ReReco-ECAL-v1	8564	0.1	3388	0.0	37	0.3
Run2010A-Nov4ReReco_v1	3414	0.0	3282	0.0	12	0.1
Fall10-E7TeV_ProbDist_2010Data_BX156_START38_V12-v1	1113	0.0	2956	0.0	11	0.1
HIRun2010-v1	11965	0.1	2798	0.0	13	0.1
Run2010B-Nov4ReReco_v1	2241	0.0	2481	0.0	9	0.1
Run2011B-DiPhoton-PromptSkim-v1	2849	0.0	2057	0.0	10	0.1
Run2011B-18Oct2011-HCAL2TS-v1	7285	0.1	2042	0.0	10	0.1
Run2011A-PromptReco-v2	1734	0.0	1994	0.0	15	0.1
Shown Sum **Total Sum**	**49977** **8975711**	**0.4** **98.59**	**28536** **8113859**	**0** **99.49**	**135** **12868**	**1.09** **98.59**
DataSet	Accesses		CPU Time		Users*day	

Showing 1 to 10 of 205 entries

Identification of corrupted files

Identify the failed accesses in a specific site

- Account for ~3% of the CRAB job failures
- Cause the users move away from submitting jobs on a site, black listing it



Corrupted Files

List of files (per site) that ALWAYS failed in job accesses in the last 10 days

Show 10 entries

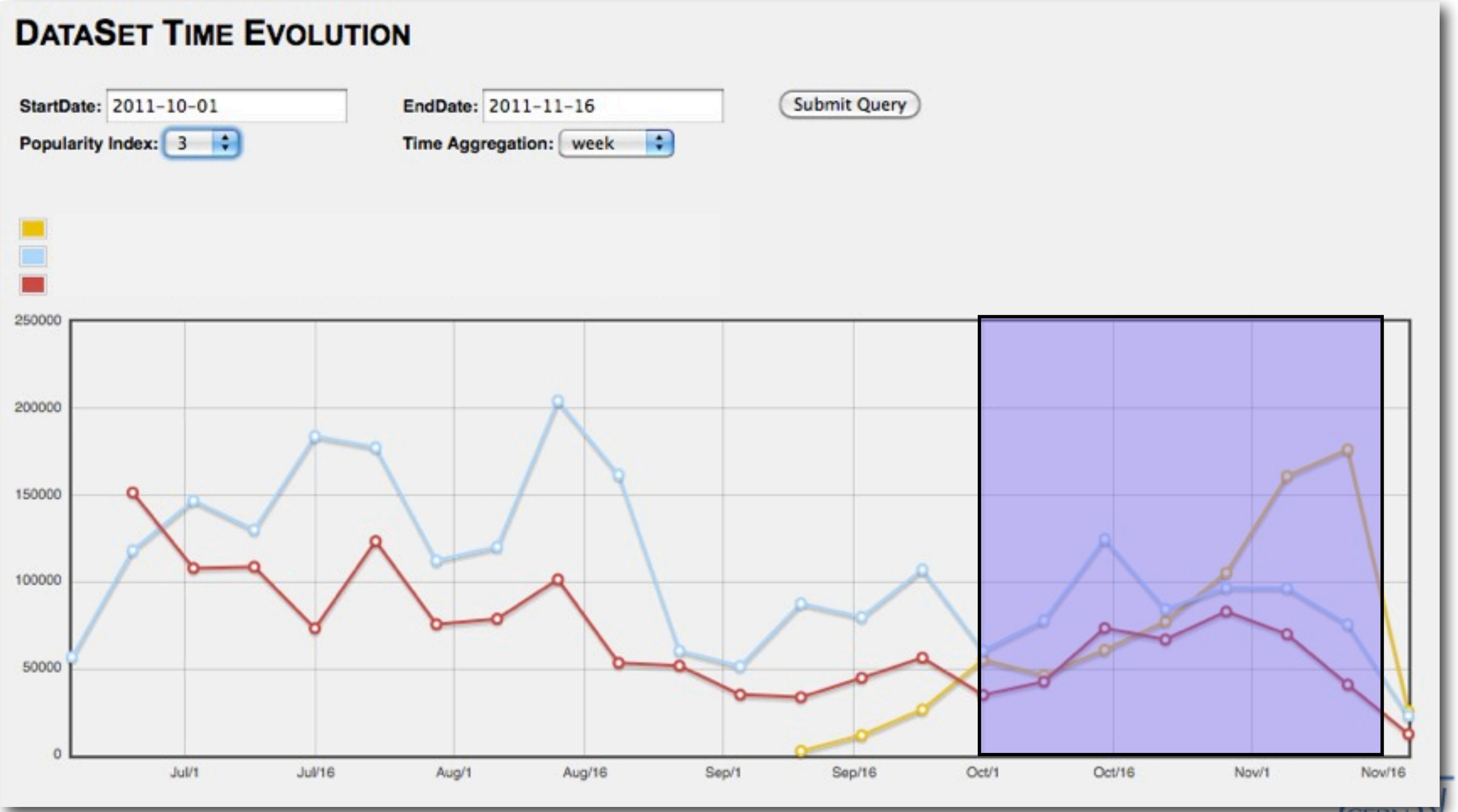
Search:

SiteName	Naccesses	Filename
T2_UK_London_IC	32	/store/mc/JobRobot/RelValProdTTbar/GEN-SIM-RECO/MC_42_V12_JobRobot-v1/0027/440942B7-89A2-E011-B5E0-00261894384A.root
T2_CN_Beijing	28	/store/mc/JobRobot/RelValProdTTbar/GEN-SIM-RECO/MC_42_V12_JobRobot-v1/0027/64EB0A0E-8BA2-E011-B284-002618943866.root
T2_US_Caltech	26	/store/data/Run2011B/DoubleElectron/RAW-RECO/ZElectron-PromptSkim-v1/0000/C09A6057-DEE0-E011-9661-0024E876A87C.root
T2_US_Purdue	17	/store/mc/Summer11/SMS-T5zz_x-05_Mgluino-150to1200_mLSP-50to1150_7TeV-Pythia6Z/AODSIM/PU_START42_V11_FastSim-v2/0003/64B2AF57-2CC4-E011-AA58-0017A4770420.root
T2_US_Purdue	16	/store/mc/Summer11/SMS-T5zz_x-05_Mgluino-150to1200_mLSP-50to1150_7TeV-Pythia6Z/AODSIM/PU_START42_V11_FastSim-v2/0003/9A4C952F-26C4-E011-B527-001F29C464E4.root
T2_US_Purdue	15	/store/mc/Summer11/SMS-T5zz_x-05_Mgluino-150to1200_mLSP-50to1150_7TeV-Pythia6Z/AODSIM/PU_START42_V11_FastSim-v2/0002/C02078DE-27C3-E011-9754-1CC1DE1CF1BA.root
T2_US_Purdue	15	/store/mc/Summer11/SMS-T5zz_x-05_Mgluino-150to1200_mLSP-50to1150_7TeV-Pythia6Z/AODSIM/PU_START42_V11_FastSim-v2/0003/A8474458-2CC4-E011-8222-0017A4771000.root
T2_US_Vanderbilt	15	/store/data/Run2011A/AllPhysics2760/RAW/v1/000/161/439/065BAAE2-4058-E011-AC23-003048F118DE.root
T2_US_Purdue	14	/store/mc/Summer11/SMS-T5zz_x-05_Mgluino-150to1200_mLSP-50to1150_7TeV-Pythia6Z/AODSIM/PU_START42_V11_FastSim-v2/0001/E0E8FAFD-15C3-E011-917A-1CC1DE055158.root
T2_US_Purdue	14	/store/mc/Summer11/SMS-T5zz_x-05_Mgluino-150to1200_mLSP-50to1150_7TeV-Pythia6Z/AODSIM/PU_START42_V11_FastSim-v2/0001/EE58E43B-48C3-E011-9F0A-0017A477001C.root
SiteName	Naccesses	Filename

Plots: DataSet Time Evolution

Study the DataSets lifetime and popularity evolution

- Configurable popularity window and time aggregation



DataSet Time Evolution

StartDate: 2011-10-01 EndDate: 2011-11-16

Popularity Index: 3 Time Aggregation: week

Popularity of Xrootd accessed files

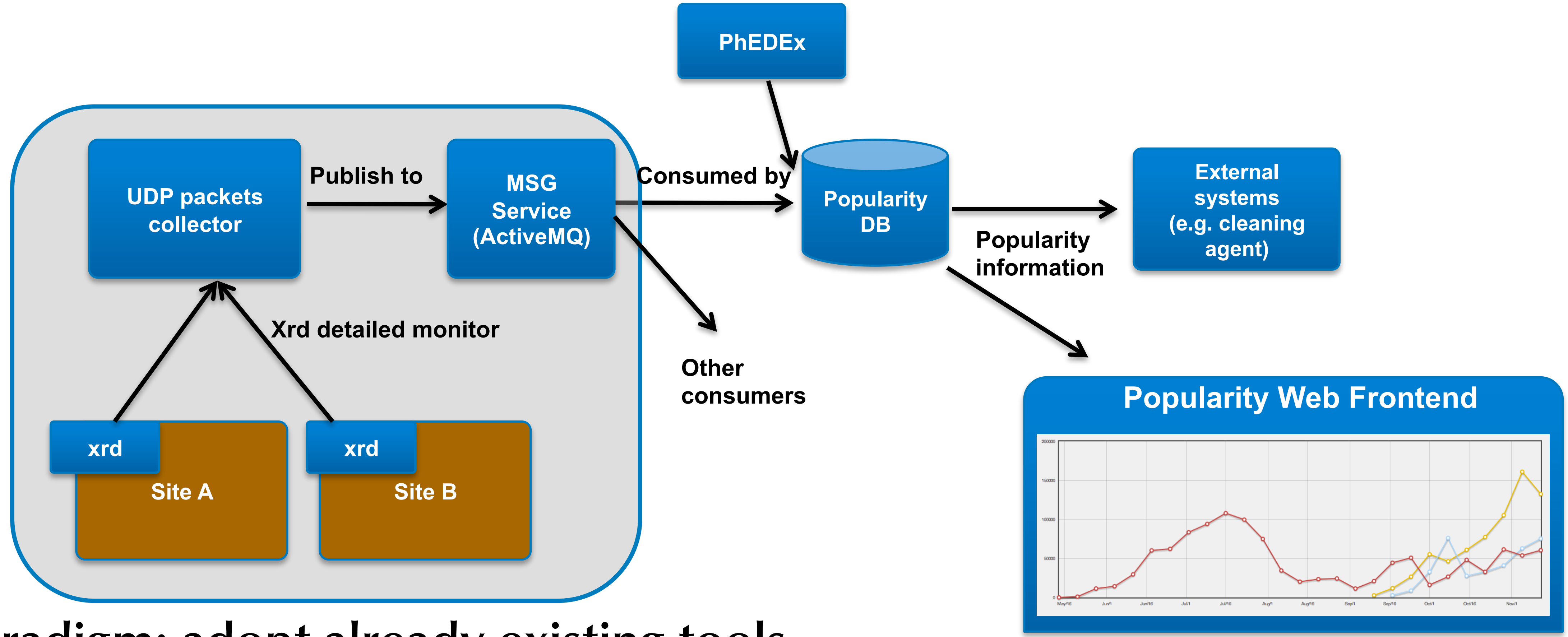
CMS expressed interest in monitoring the popularity of the data accessed via Xrootd

- First use case: popularity of the files accessed at CERN from the CMS-EOS DataSvc

Advantages:

- Not based on CRAB+dashboard, allows to define a data popularity also for batch/interactive job submission
- Can help in managing the user space:
  - Providing feedback about not only the popularity of the official datasets, but also of the user data

Architecture



Paradigm: adopt already existing tools

- Collector of the Xrootd detailed monitoring data
  - Based on UDP packets
  - Described in two talks of this section [Tadel, Oleynik]
- Messaging System for Grid (MSG)
  - Publish-subscribe model
    - Reduce the number of services collecting the UDP packets
    - Several consumers can access the MSG Broker

Xrootd-Data Popularity Metrics

For the Xrootd use case the popularity metrics are limited to the storage oriented information

- File accesses, opening/closing time, user, client & served domain
- No straightforward job-related metrics:
  - CPU time
    - could be inferred by the file Close-Open timestamp
  - file exist status
  - lumi sections, etc

A lighter DB schema (respect to the CRAB-based popularity) will be tailored on the available infos

Status

All the elements of the architecture have been separately tested

- Collector, MSG queue, PopDB
- Work ongoing to put them together in a single workflow, and to access the monitoring messages for the CMSEOS instance

Key point:

- Association LFN to a DataSet/Block
  - Need a sustainable/optimized query to the PhEDEx DataSvc

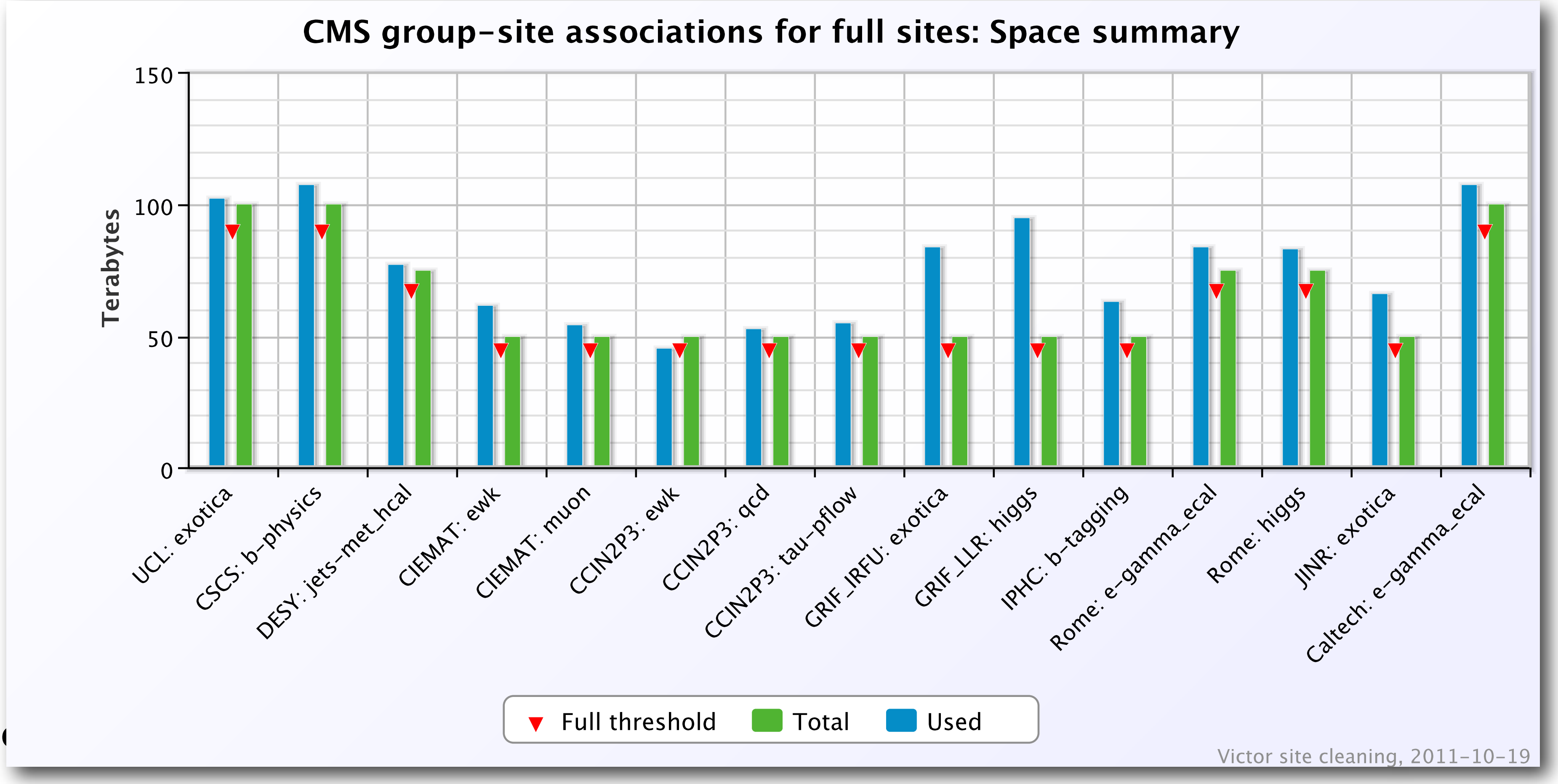
Site Cleaning Agent

Scan Tier2 sites reaching their space quota and suggest <u>obsolete, unused data</u> that can be safely deleted

- Popularity Web API, in conjunction with PhEDEx (the CMS data placement and file transfer system) information

Advantages

- Optimal handling of the storage assigned to all physics- groups
- Monitoring of the evolution in time of the used/pledged space and of the removed datasets
- Can be extended to the Xrootd-based Data Popularity



Victor site cleaning, 2011–10–19

D. Gior

Conclusions

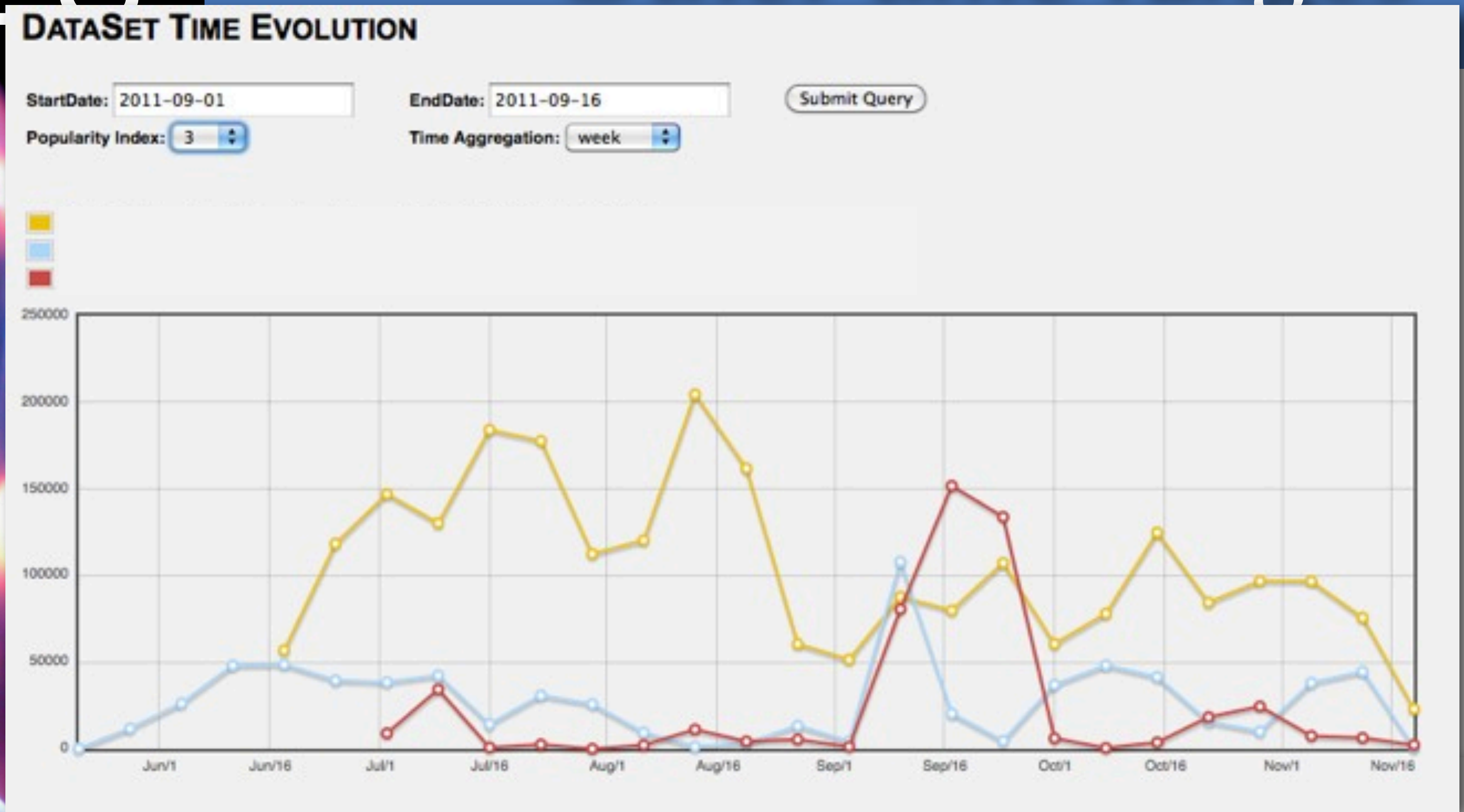
- We have developed the CMS Popularity Service that tracks over time file accesses and user activity on the WLCG

- A popularity-based Site Cleaning Agent has been developed
  - implements a strategy to free up space at Tier2 sites

- Most of these functionalities can be extended to the Xrootd-accessed files
  - Proposed infrastructure for the collection of Xrootd monitoring data
    - UDP packets Collector + MSG system + PopDB

backup

Other time ranges

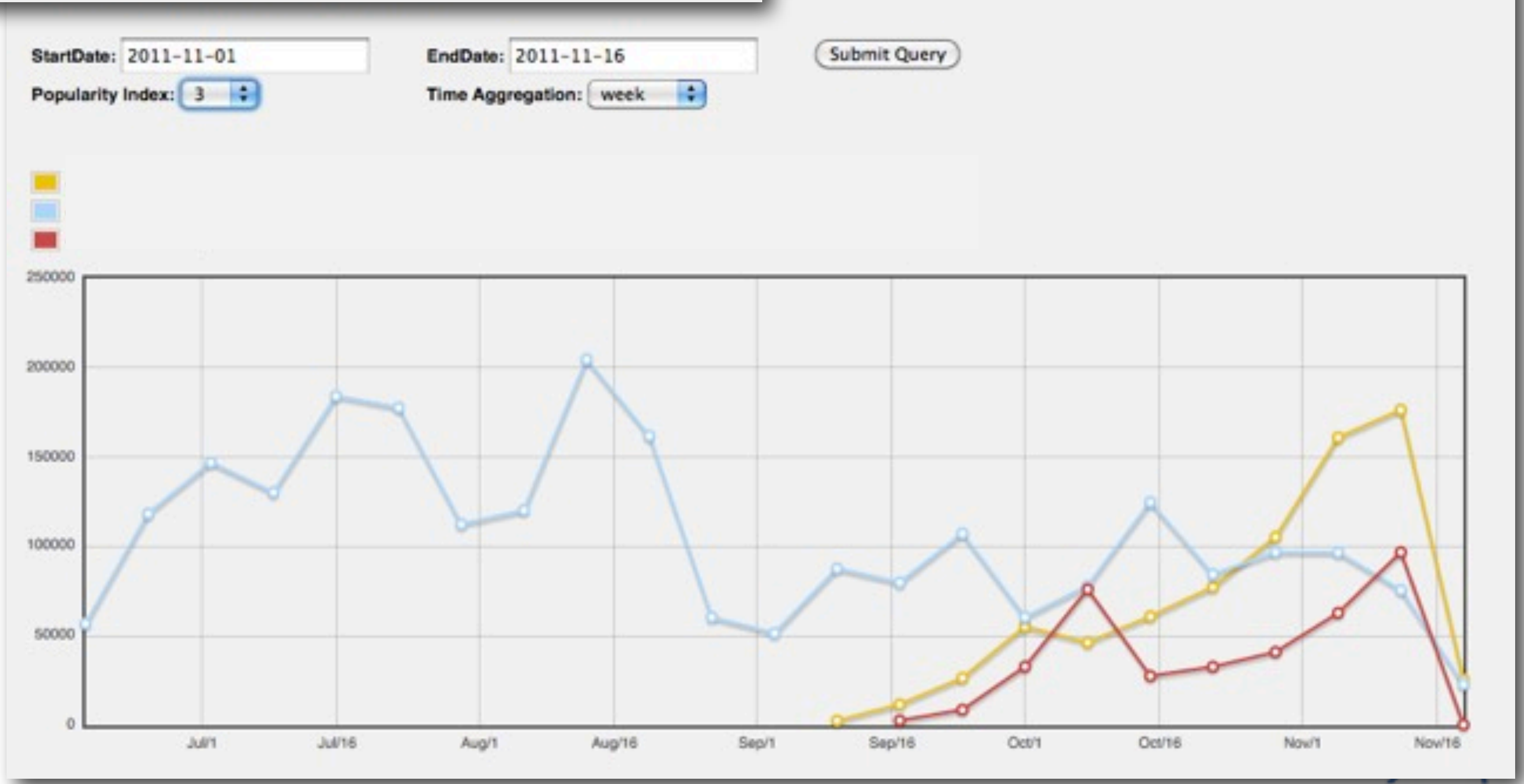


DATASET TIME EVOLUTION

StartDate: 2011-09-01 EndDate: 2011-09-16 Submit Query

Popularity Index: 3 Time Aggregation: week

🟨
🟦
🟥



StartDate: 2011-11-01 EndDate: 2011-11-16 Submit Query

Popularity Index: 3 Time Aggregation: week

🟨
🟦
🟥