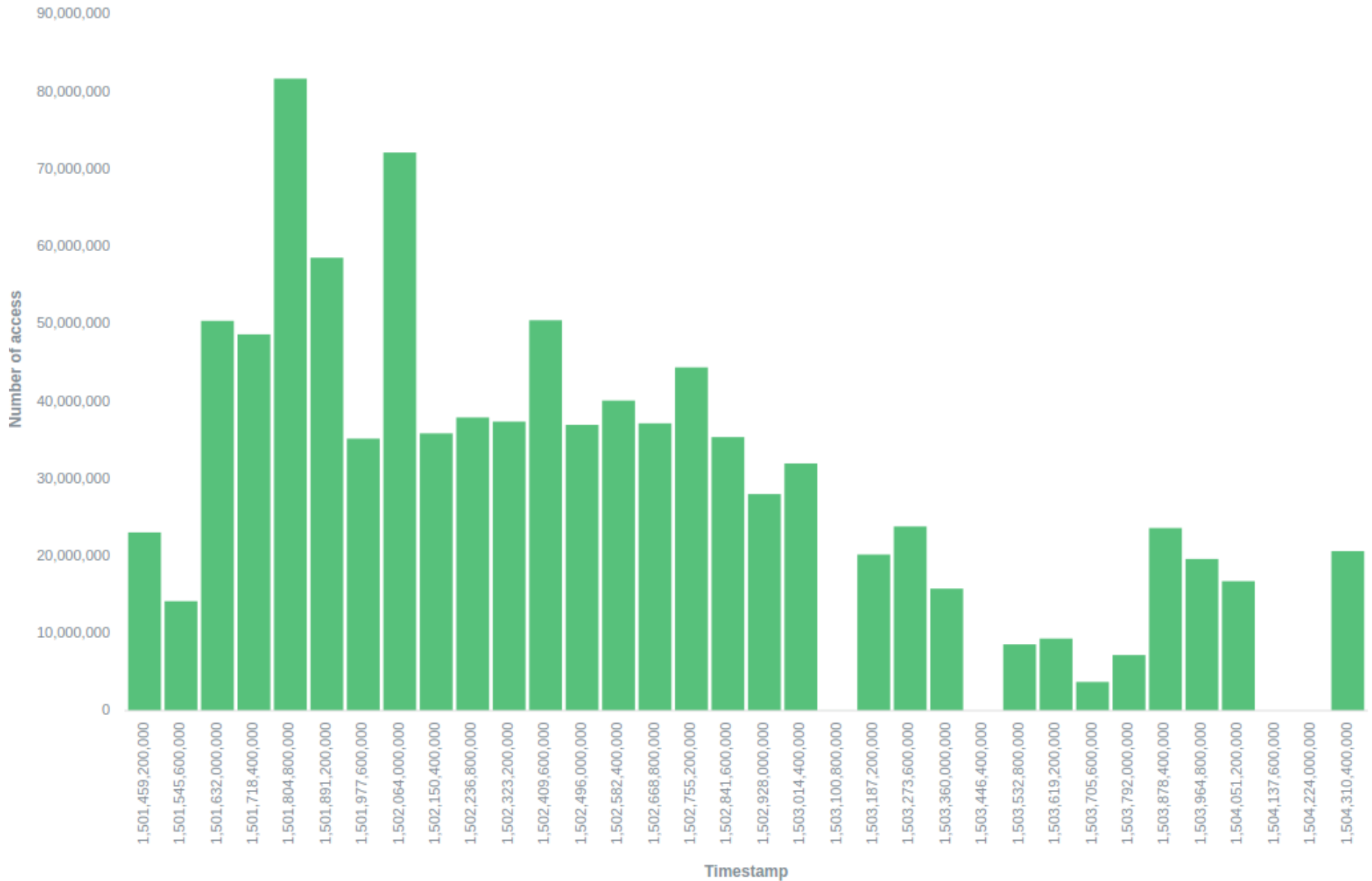# Popularity of CMS datasets

Presenter: Justinas Rumševičius
Supervisors: Valentin Kuznetsov
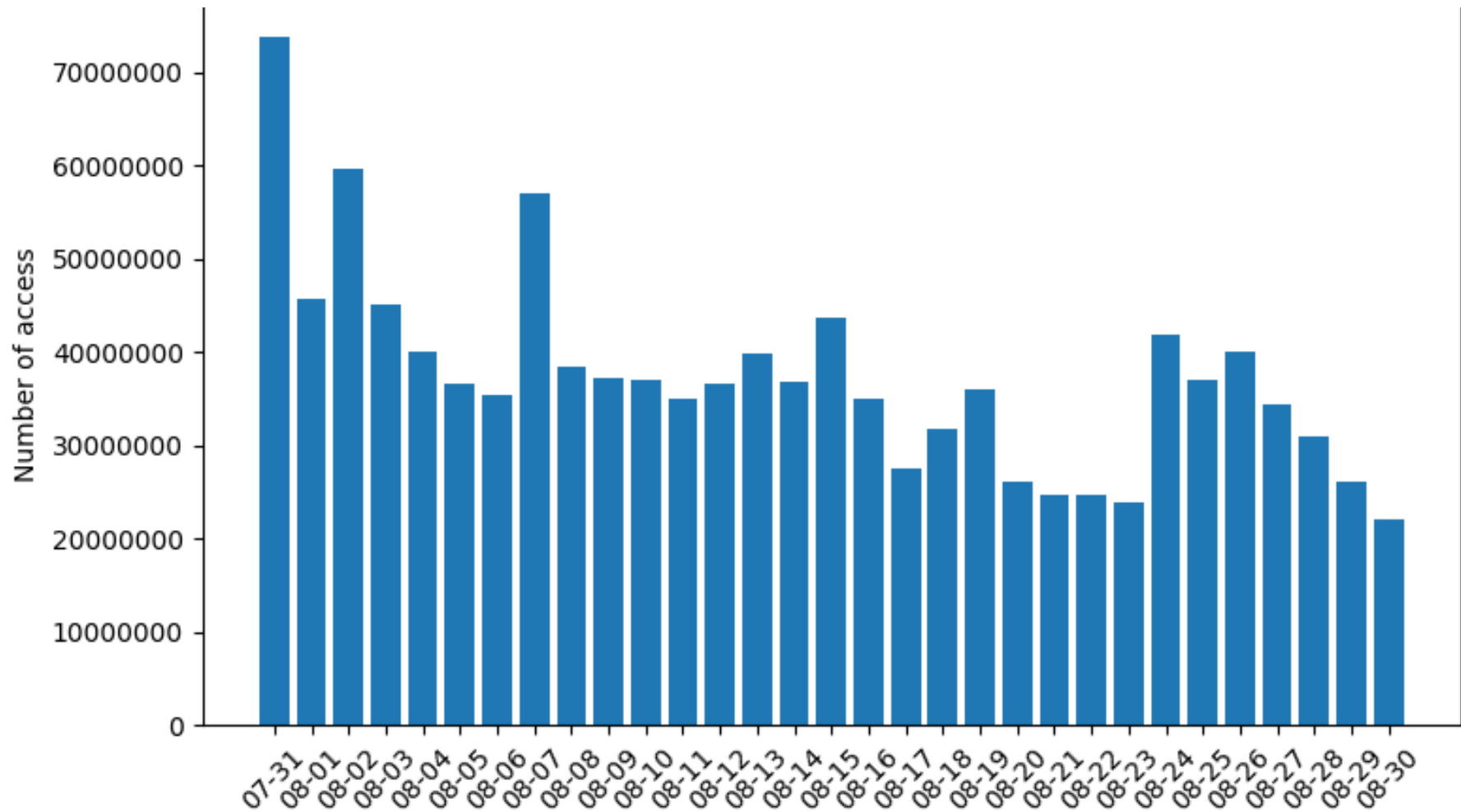Carl Vuosalo
Jean-Roch Vlimant

CERN
2017

Current progress

- Issues with Kibana timestamp

- Made some plots on my machine with matplotlib
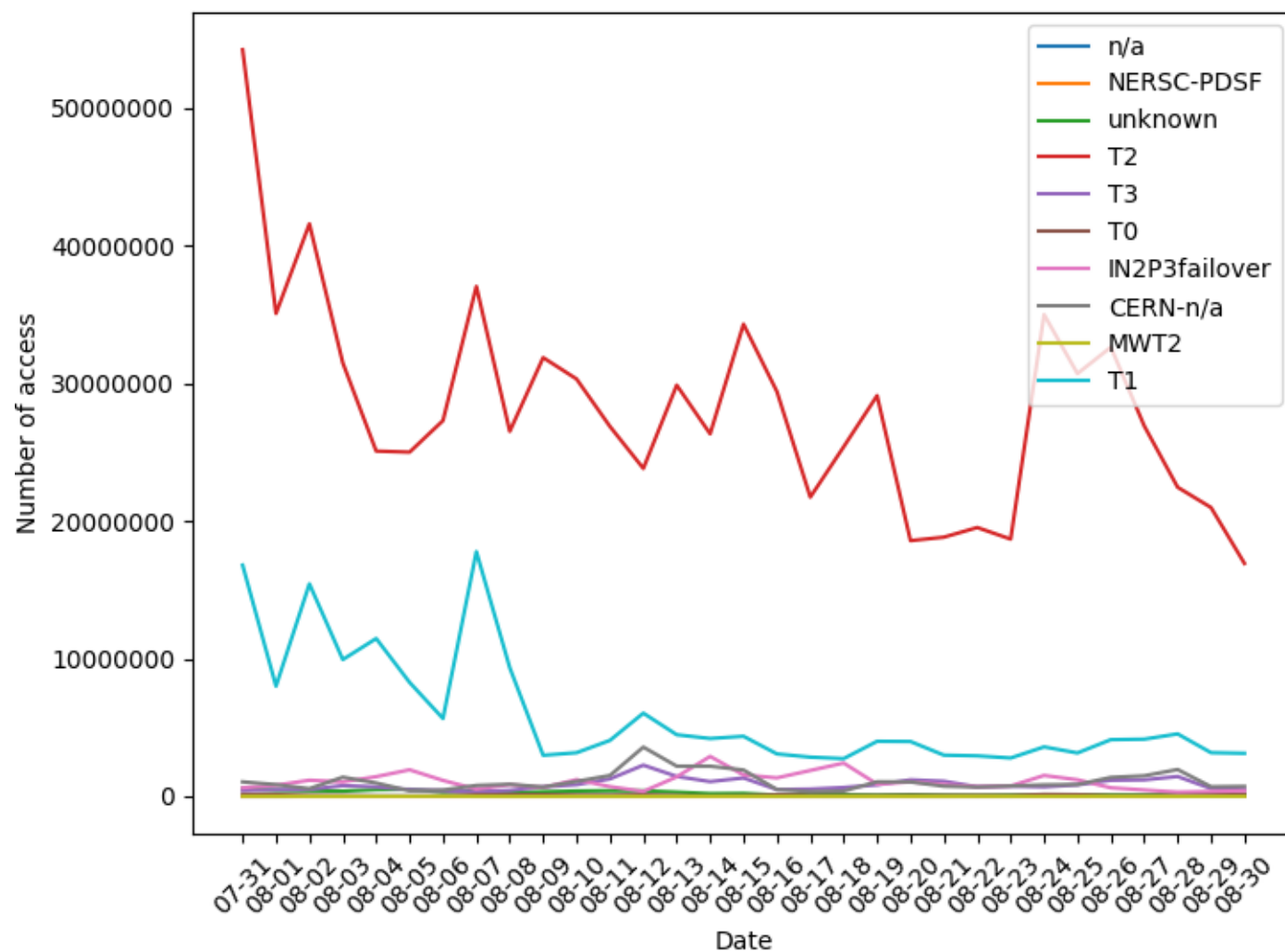
- Kibana and matplotlib plots are different
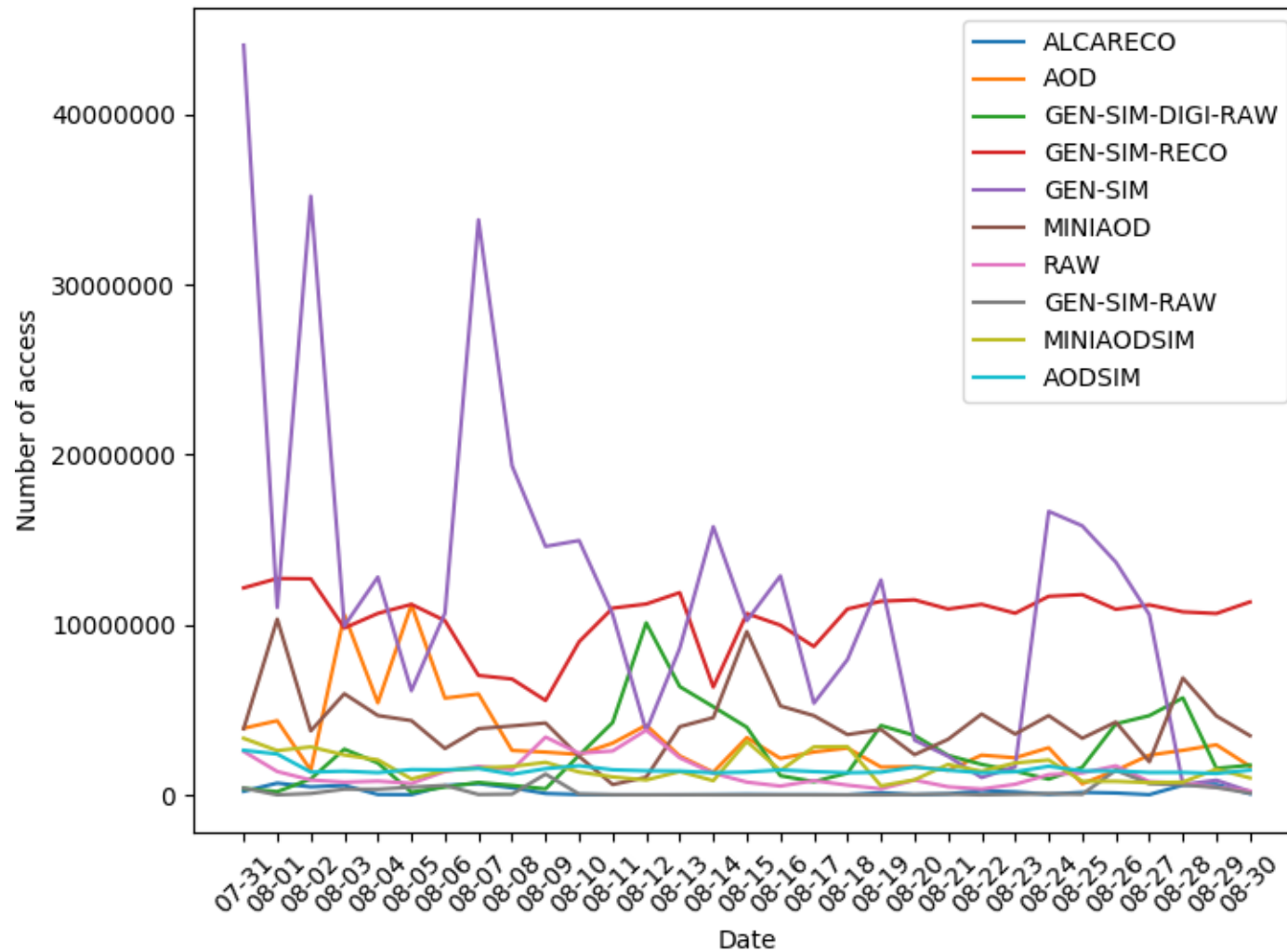
Plot – Number of accesses (Kibana)

Plot – Number of accesses (matplotlib)

Plot – Grouped by site tier

Plot – Grouped by data tier

Tables – grouped by data tier and site tier

Tier	Number of accesses	Percentage of all
GEN-SIM	366886628	31.6313%
GEN-SIM-RECO	322453427	27.8005%
MINIAOD	130272268	11.2315%
AOD	101181034	8.7234%
GEN-SIM-DIGI-RAW	77131946	6.6500%
MINIAODSIM	50456617	4.3501%
AODSIM	45984728	3.9646%
RAW	38896908	3.3535%
GEN-SIM-RAW	6947298	0.5990%
ALCARECO	5927966	0.5111%
FEVT	4551118	0.3924%
GEN-SIM-RECODEBUG	3742856	0.3227%
LHE	1837493	0.1584%
RECO	1480637	0.1277%
RAW-RECO	1117820	0.0964%
ALCAPROMPT	810765	0.0699%
RAWAODSIM	95022	0.0082%
GEN-SIM-DIGI-RAW-HLTDEBUG	63127	0.0054%
FEVTHLTALL	10712	0.0009%
GEN-SIM-DIGI-RECO	10561	0.0009%
Sum	1159858931	99.9977%

Site tier	Number of accesses	Percentage of all
T2	872958955	75.2625%
T1	183162782	15.7915%
CERN-n/a	34007785	2.9320%
IN2P3failover	33167070	2.8595%
T3	26136719	2.2534%
unknown	7620707	0.6570%
T0	2429082	0.2094%
n/a	224118	0.0193%
MWT2	128342	0.0111%
NERSC-PDSF	26154	0.0023%
UKI-LT2-RHUL	11353	0.0010%
XT2	6112	0.0005%
XT3	3220	0.0003%
Roaming	2480	0.0002%
X3	303	0.0000%
datagrid.cea.fr	30	0.0000%
ru-PNPI	14	0.0000%
Sum	1159885226	100.0000%

Next steps

- Timestamp is off by a couple of hours
- Make separate plots for different streams
- Filter out site tiers other than T0, T1, T2, T3
- Adjust automated scripts
- Put everything in Jupyter Notebook

My github repositories

- Github account:
    - https://github.com/justinasr

- Forked CMSSpark ("development" branch):
    - https://github.com/justinasr/CMSSpark
    - https://github.com/justinasr/CMSSpark/tree/development

- Wiki:
    - https://github.com/justinasr/CMSSpark/wiki

- Weekly reports:
    - https://github.com/justinasr/CERNReports