

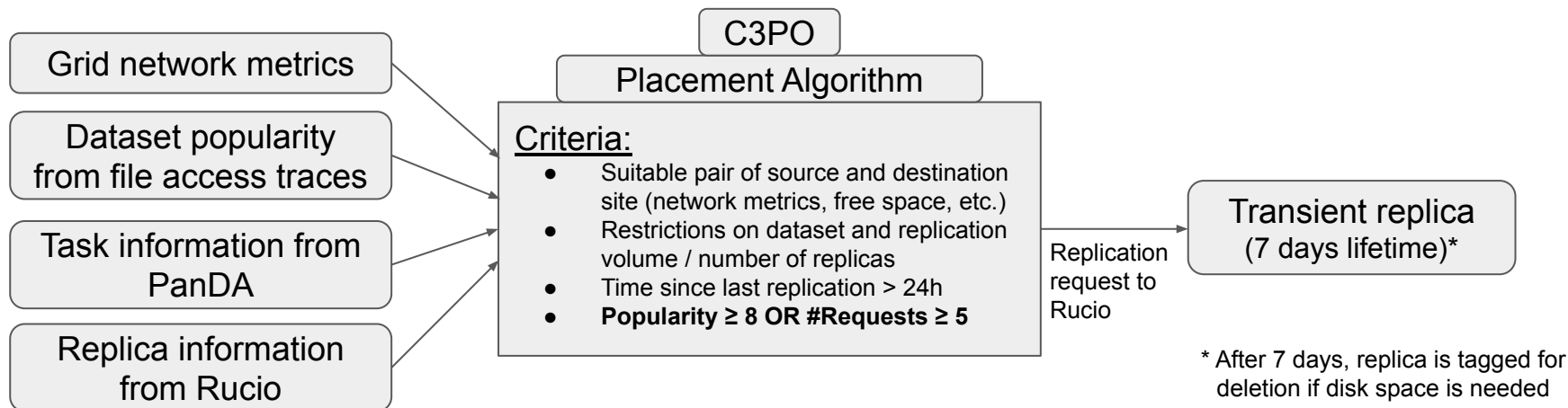
Dynamic Data Placement in Rucio

Thomas Beermann

Overview

- C3PO was meant as a PD2P successor as a Rucio daemon.
- Development started in 2015.
- It first ran in a dry-run mode only computing possible new replicas without actual replication.
- In 2016/17 it ran in an A/B testing mode with a low volume of new replicas.
- Extensive analysis of the performance was done by Thomas Maier and ultimately presented at CHEP 2018.
- Most of the information in this talk comes from the CHEP presentation. For more information you can find the proceedings [here](#).
- After that it was decided that major redevelopment is needed and the project is on hold since then.

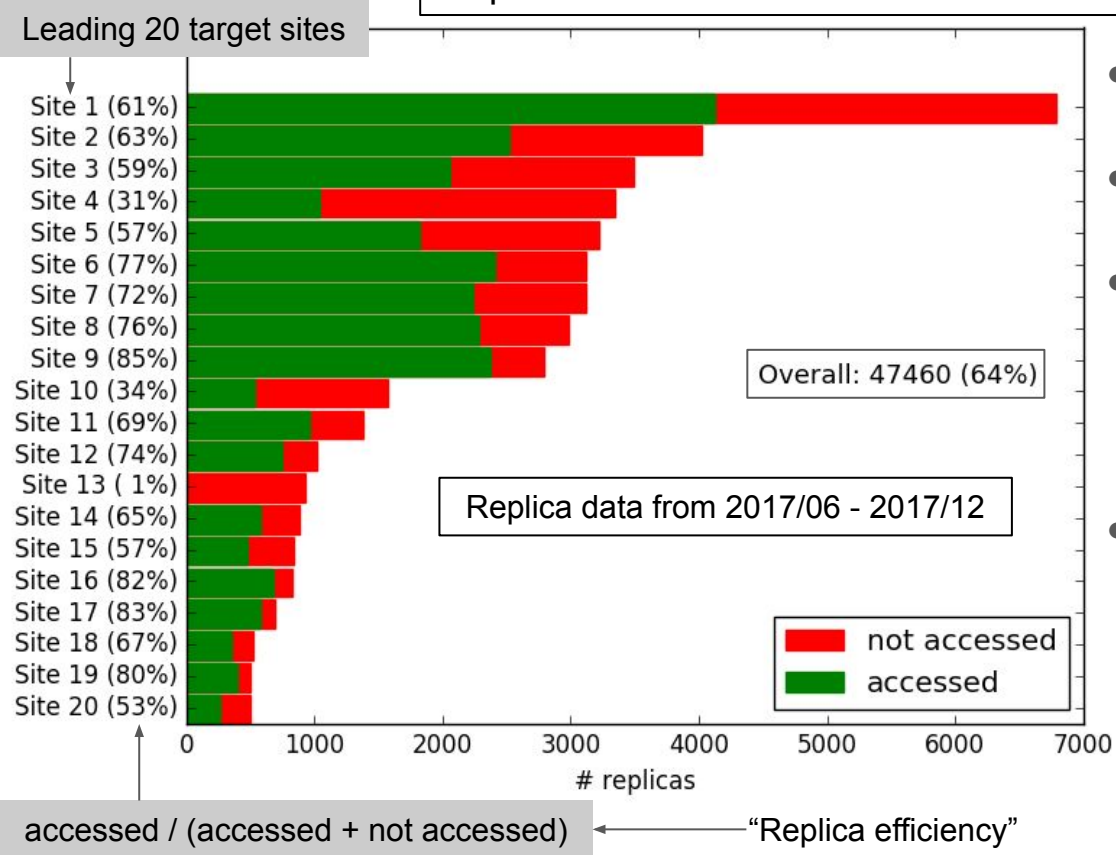
C3PO - Dynamic data placement agent in ATLAS



- Popularity: number of dataset accesses in the past 7 days
- #Requests: number of user analysis tasks that use the dataset as input, submitted in the past 24h

Replica usage after creation - Sites

Replicas which were **accessed** or **not accessed** after creation

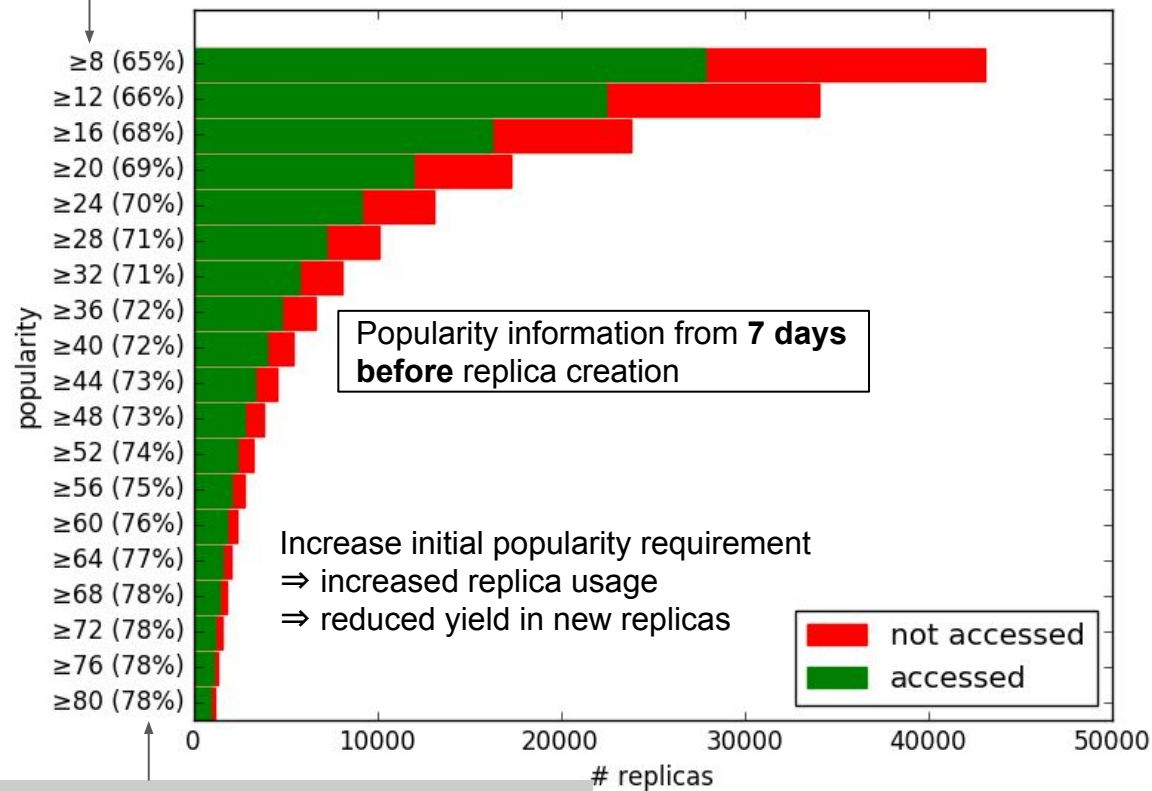


- File access information from Rucio traces ⇒ replicas were used or not?
- Overall, 64% of replicas were used after their creation
- Target site selection priority is weighted to avoid uneven distribution of replicas
 - Still clustering at a few sites
 - Possibly periods of high disk space availability
- Replica efficiency can strongly depend on where it was put
 - **Placement algorithm doesn't take into account computing resources at target site**
 - Some correlations with data type or format can be seen as well

Replica usage after creation - Popularity

Popularity $\geq X$

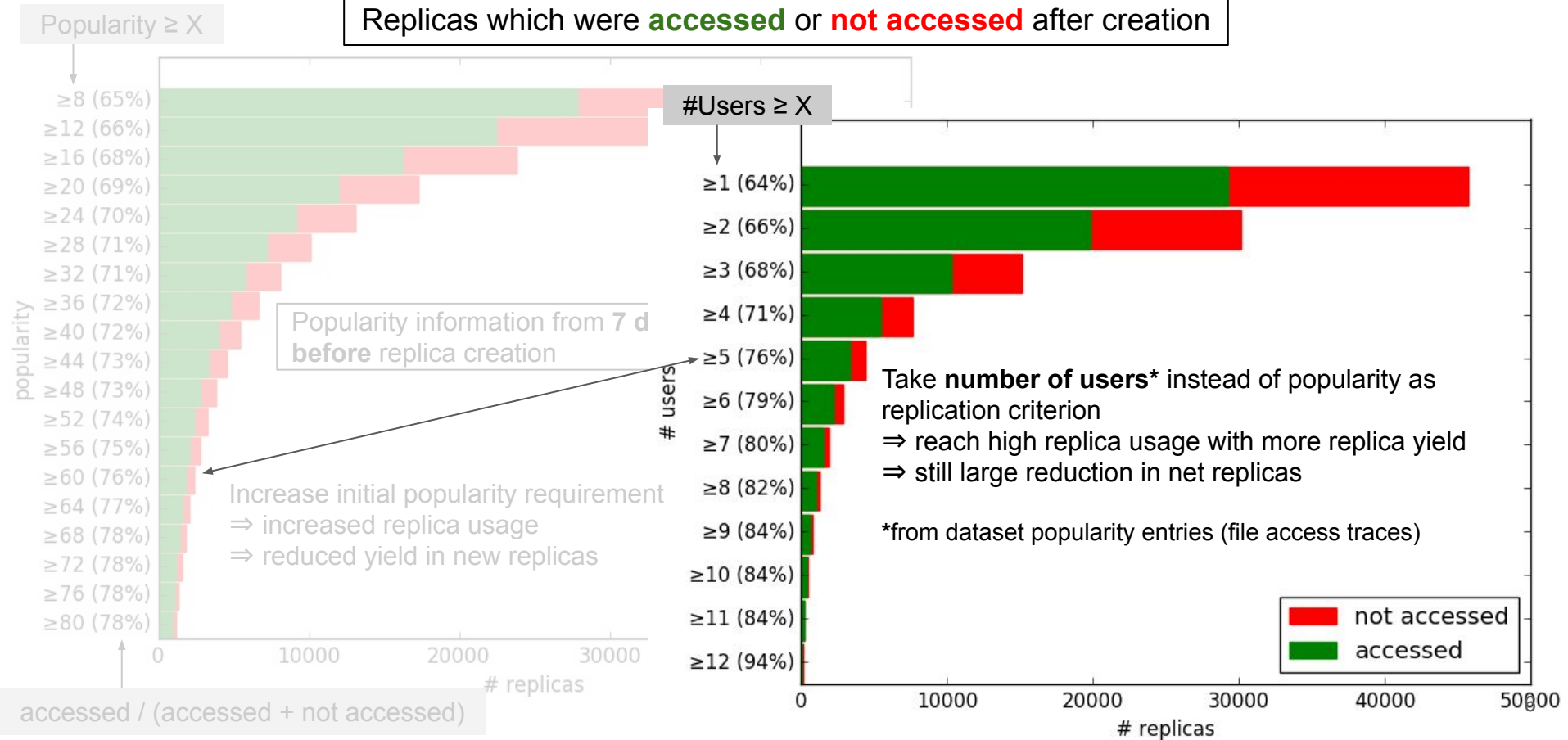
Replicas which were **accessed** or **not accessed** after creation



accessed / (accessed + not accessed)

Replica usage after creation - Popularity

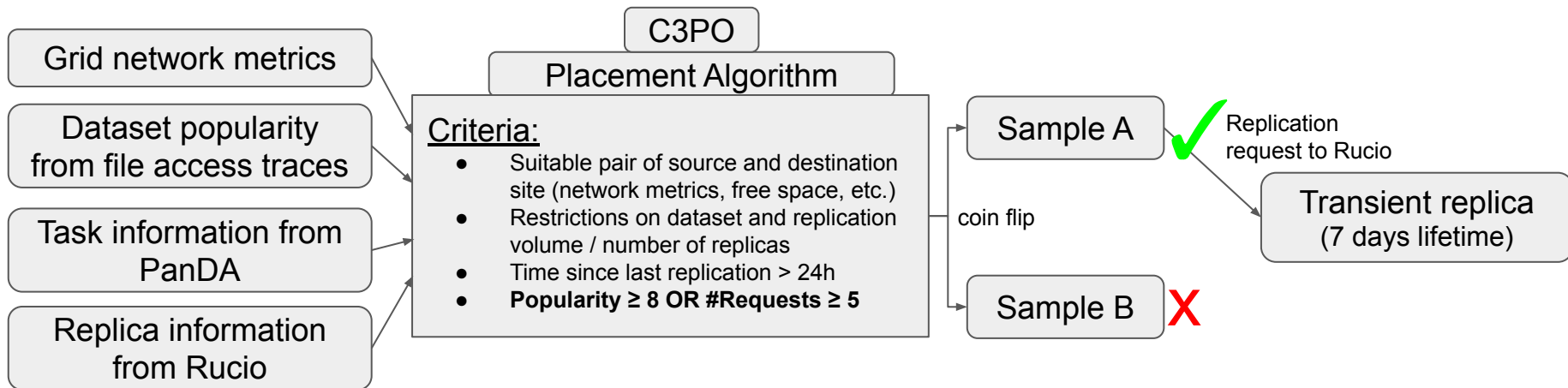
Replicas which were **accessed** or **not accessed** after creation



C3PO impact analysis

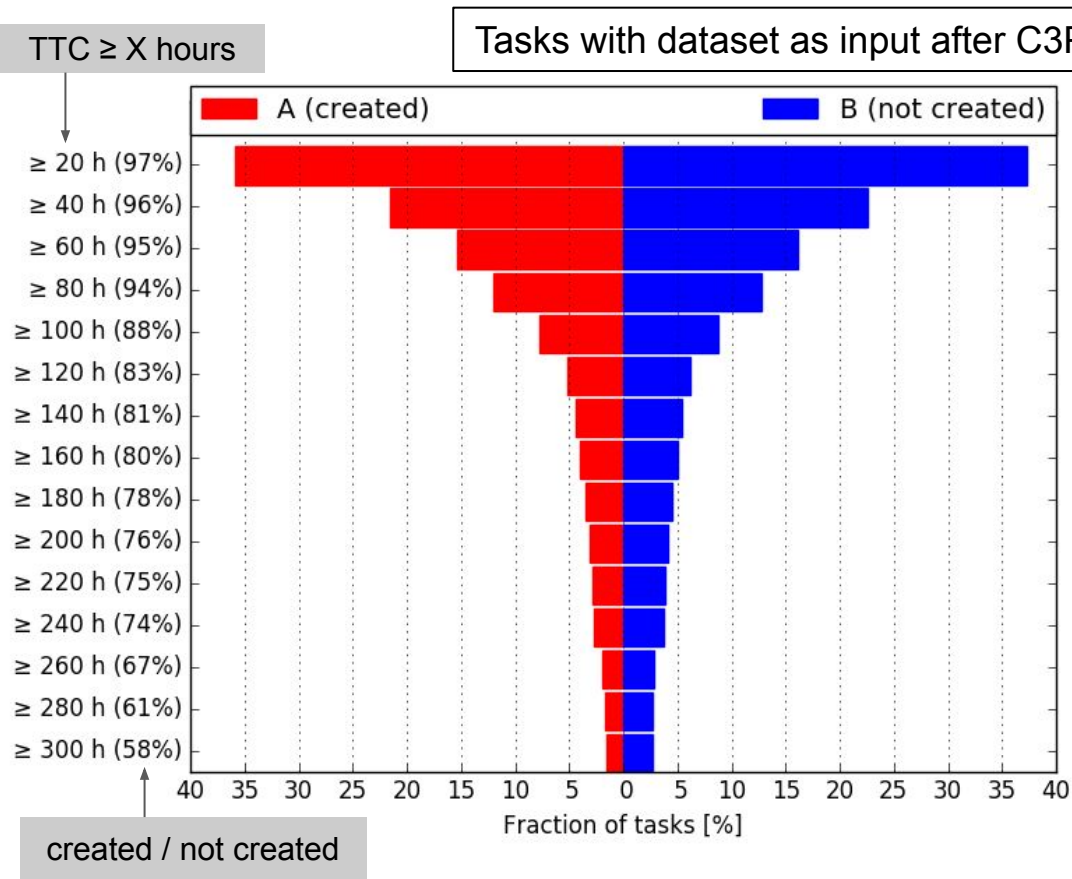
- Attempt to measure effect of C3PO operations on Grid throughput in terms of processed tasks and jobs
 - Metrics like replica access after creation indicate how well C3PO selected popular datasets ↔ no gauge for impact on efficient Grid processing
 - Main problem: measurement of metrics related to usage of C3PO selected datasets vs. other datasets doesn't really allow for a one to one comparison
- ⇒ Decided to run C3PO in an **A/B testing mode** for a period of time

C3PO impact analysis - A/B testing



- Direct comparison of C3PO decisions being applied vs. not being applied
 - After positive C3PO decision: coin flip (based on dataset name)
 - Decisions split into Sample A (replica is created) and Sample B (replica is not created)
 - Test period ~1.5 months
- For datasets that fall into Sample A or Sample B, compare metrics that are affected by (temporary) inaccessibility of input data or high workload on sites

C3PO impact analysis - Task TTC



- TTC (time to completion): time from point of creation to completion of the task
- Difference between **created** and **not created** replicas starts to occur in the tails of the TTC distribution
 - Statistically limited
 - Pronounced difference only in long tails
- Noticeable effect, but concerns only a small fraction of tasks

Summary

- Dynamic data placement agent C3PO
 - Developed and operational during Run-2 phase of ATLAS
 - Usage efficiency of resulting newly created replicas >60% (depending on parameters like target replication site, data type/format, etc.)
 - Altering initial C3PO decision criteria effects resulting usage probability
 - C3PO replicas tend to survive longer than nominal 7 days lifetime ↔ continuously accessed → efficient use of available disk space
- C3PO impact analysis with A/B testing
 - Metrics like task TTC indicate that replicas created by C3PO have some impact on Grid processing of their associated datasets
 - Only small effects (on very limited number of Grid tasks)
 - In general difficult to unambiguously attribute observed differences to C3PO replicas
- For a possibly better performance a strong integration with WFMS is needed.
- Generally, C3PO was able to spot popular datasets but for a better placement of additional replicas scheduling information should be taken into account.