

Rucio, FTS and Network: Thoughts and High Level Requirements

J. Letts (UCSD), D. Piparo (CERN) - DOMA Review Mini-Workshop I/2 - June 1, 2021



The Role of Common Software Solutions

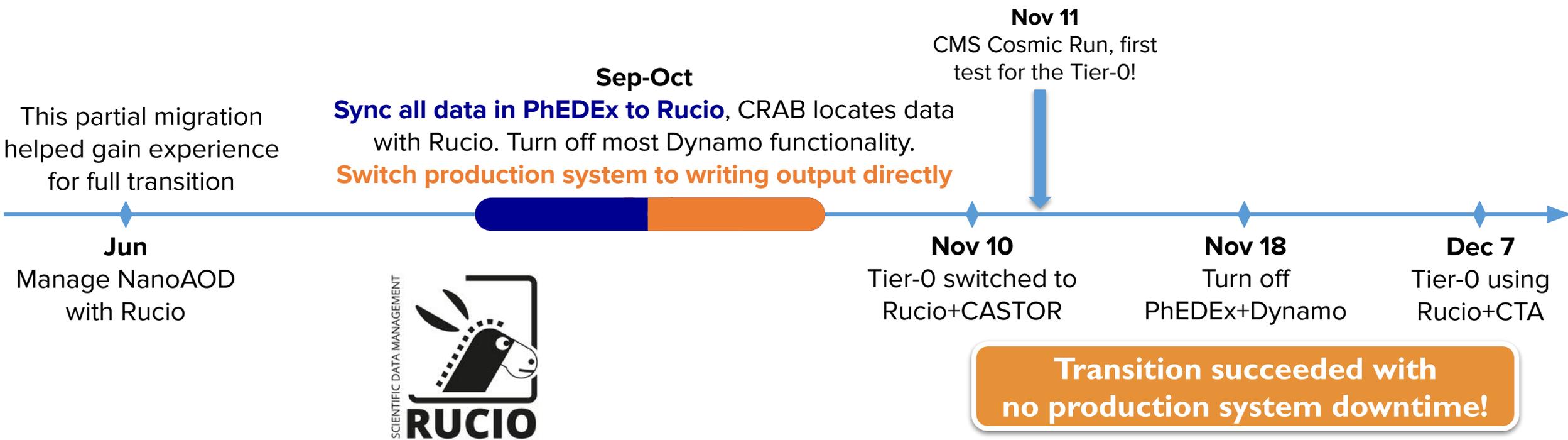
- Adopting common software and services for the sustainability of the CMS stack
 - An overarching direction of CMS Offline Software and Computing
- Examples of **community software/services in CMS**:
 - ROOT, Geant4, event generators, FastJet
 - DD4Hep for detector description
 - CRIC, FTS, CTA
 - Rucio
 - And much more: HTCondor, Jenkins, GCC, Clang, Git, TBB, ...





Rucio 1/2

- Relatively new in CMS
- Switched from PhEDEx+Dynamo in November 2020
- Current activities mainly concentrated on
 - Gaining expertise with the DDM
 - Improving the integration with the workload management
- **We are very satisfied with the new tool and the relationship with the development team.**
 - *Thanks Martin, Mario & all!*





- We appreciate the current OSS model behind Rucio: open and transparent
- **An asymmetry wrt other HEP common sw projects might be there**
- ROOT, G4, FTS, CRIC, DD4Hep, Generators: none has its core development team inside a particular experiment
- **Not a problem per se today - we are very happy with the collaboration with Rucio Devs**
- **Unforeseen issues affecting sustainability could arise in the next 5 years**
 - Changes of research direction/priorities of the team where Rucio is mainly developed
 - Person power cuts forcing the current team to prioritize only one experiment
 - Senior developers moving to other challenges and not being replaced
 - Hard to identify and fund developers outside the core development team, e.g. in CMS
- Can the the aforementioned risks be mitigated by **complementing the current development effort with a centralized, experiment independent supporting contribution?**
 - In a nutshell: what could be an even more sustainable funding model for the product



FTS and Network

- **FTS: a single entry point offering opportunities in the area of network management**
 - A place where we see improvements in packet marking, traffic shaping and SDNs being plugged
 - Tightly coupled with RUCIO
- **CMS heavily depends on a reliable bulk transfer tool even today**
 - Even more in Run 4 and beyond
 - Today: ≥ 20 GB files become troublesome
- Discussing now the future operation mode of Tier-0: file sizes are a crucial topic
 - What will be the new average and max file size FTS will be able to handle?
 - **Need solid transfer for 100 GB files at least** (e.g. good handling of tails of file size distribution)
- Larger files means also **fewer scaling challenges for all kind of bookkeeping tools**
 - An opportunity to simplify (and thereby make more sustainable) our future computing software tools



Network

- **CMS treats network more and more like any other computing resource**
 - Towards network management
- HL-LHC network challenge: **not just giving HEP enough capacity**
 - **But best use of the shared network infrastructure within feasible limits**
 - Respecting and enforcing room for other scientific and academic network traffic
- **CMS cannot drive network research**
 - Lack of network engineering expertise within the Collaboration and unlikely to build up a substantial one
- **CMS can engage (within person power constraints and with favourable cost/benefit ratio) with network engineers to improve the current system**
 - Testing, providing use cases, describing operations and approaches adopted
 - **A single point of contact with CMS (and ATLAS) transfers and network capabilities: FTS**
- **Priority: a good, reliable, complete network usage monitoring for transfers and xrootd traffic**
 - *It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.*