



# What's coming for EOS Data Durability

**Manuel Reis**  
**Maria Arsuaga**  
**(IT-ST-PDS)**

# Overview

## EOS Ops Data Durability

- What it is and why it is.
- Achievements

## Method Overview

## Current Status

## Goals and Visions

- Promote proactiveness investigations
- Target portability – non CERN specific
- Software consolidation - Python3 & CI/CD

## Conclusions

# EOS Ops Data Durability

Can we quantify the amount of problematic files?

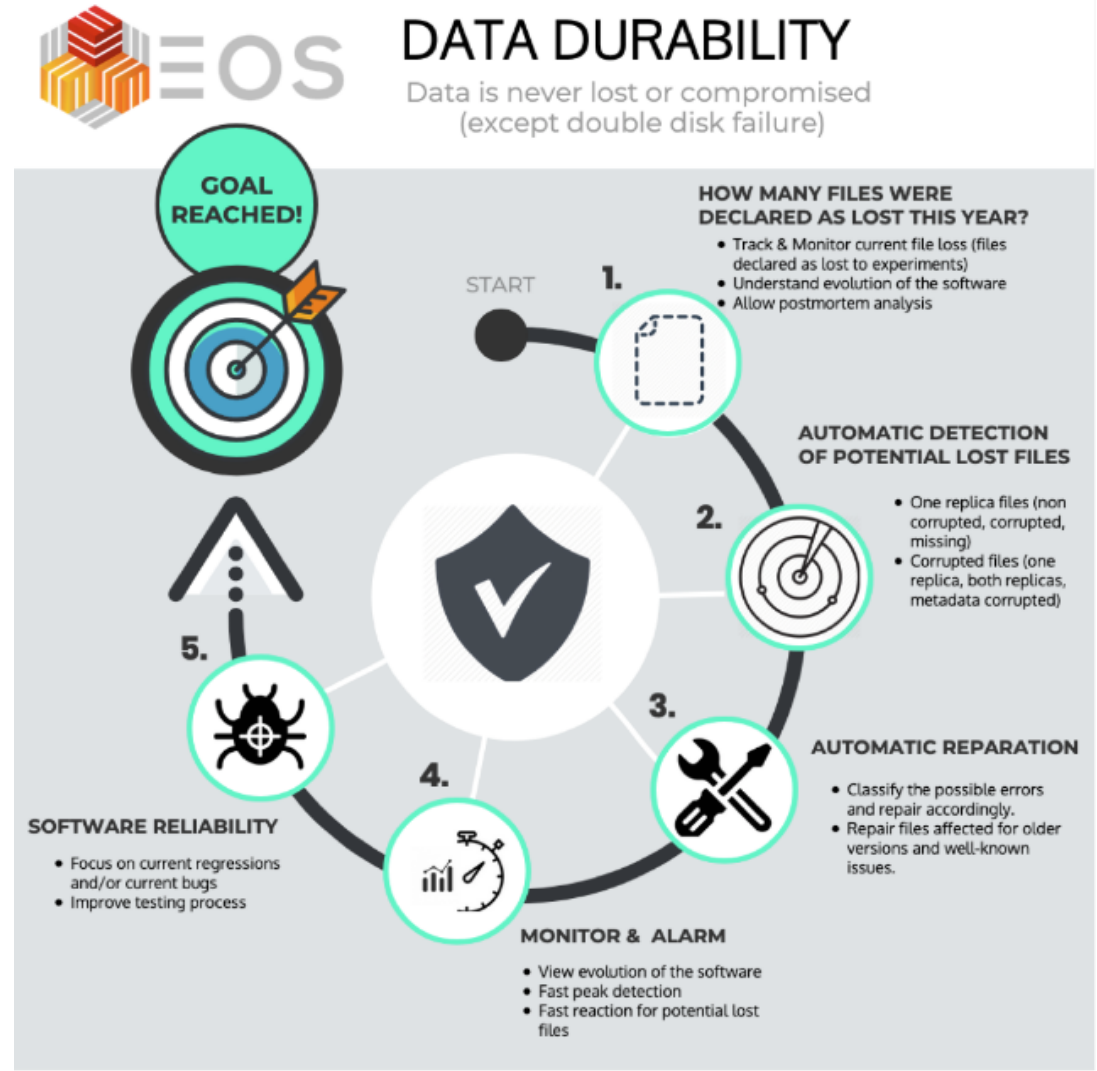
Can we grouped them by their “symptoms”?

It is extremely important do have an updated knowledge about the state of the data

As Software evolves:

- Old bugs are resolved (ideally)
- New ones may be introduced (unfortunately)

Identify, resolve and monitor files in problematic states.



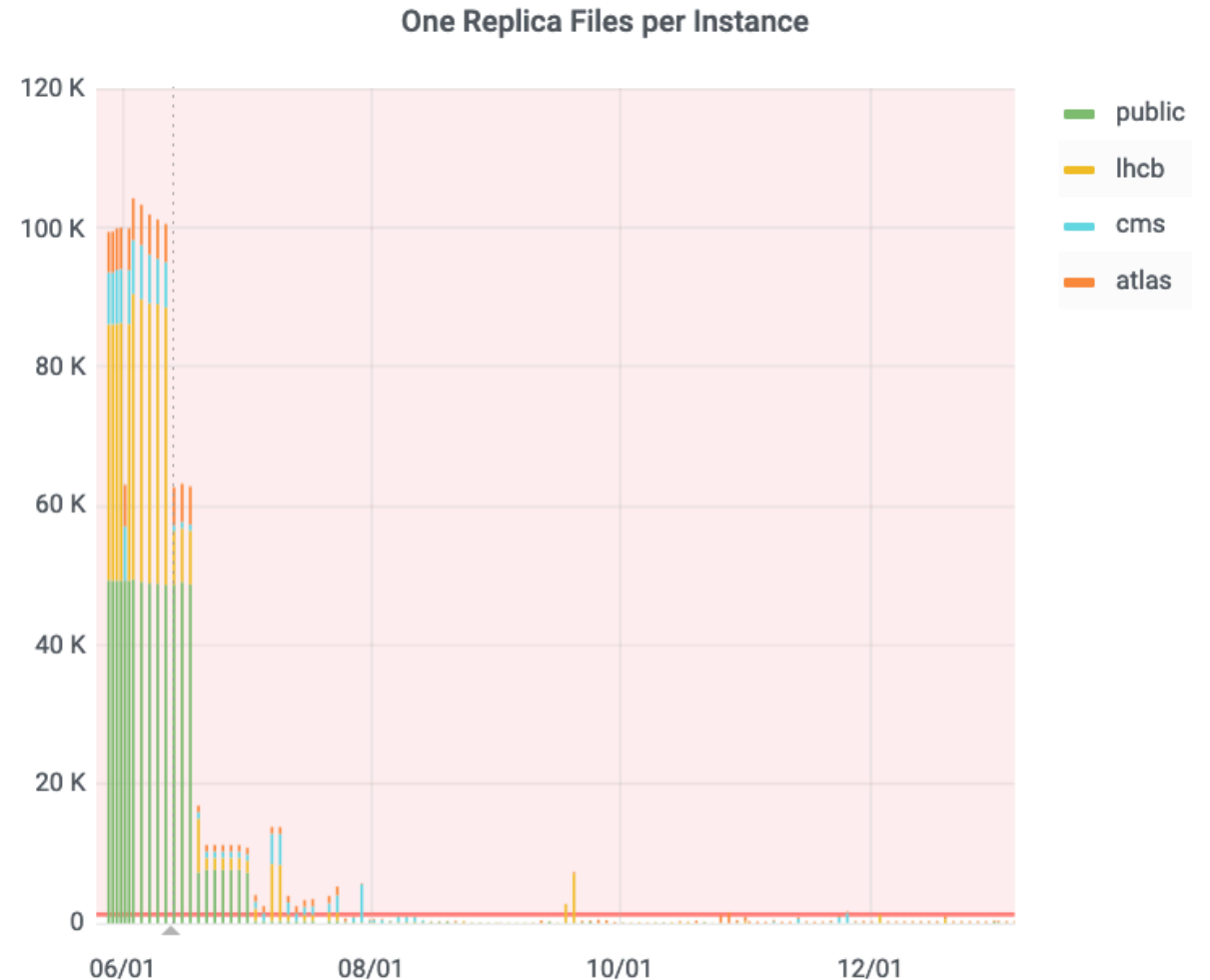
# Past achievements

**This allowed to restore a significant amount of files to a more stable state**

- Durability
- Integrity

**Wigner draining campaign succeed in time with automatic reparation during the draining process.**

**The focus is on the currently occurring problems.**



# How are incoherencies spotted

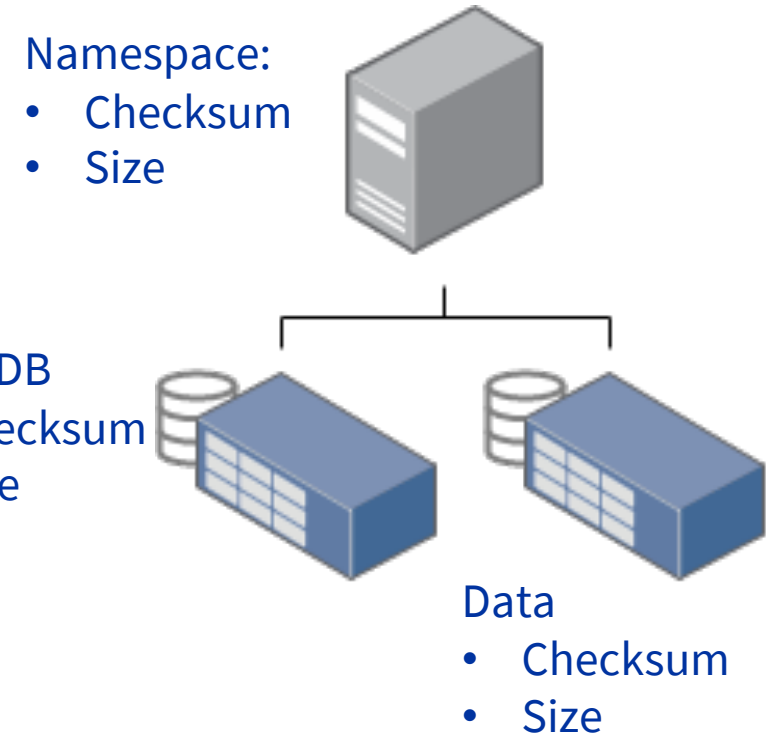
## The metadata is the first to be analysed

- Aborted, removed, recycle, detached, etc.

## The validation procedures check if the data is consistent and if correctly replicated

- System has multiple layers where information about the file or its replicas is stored.

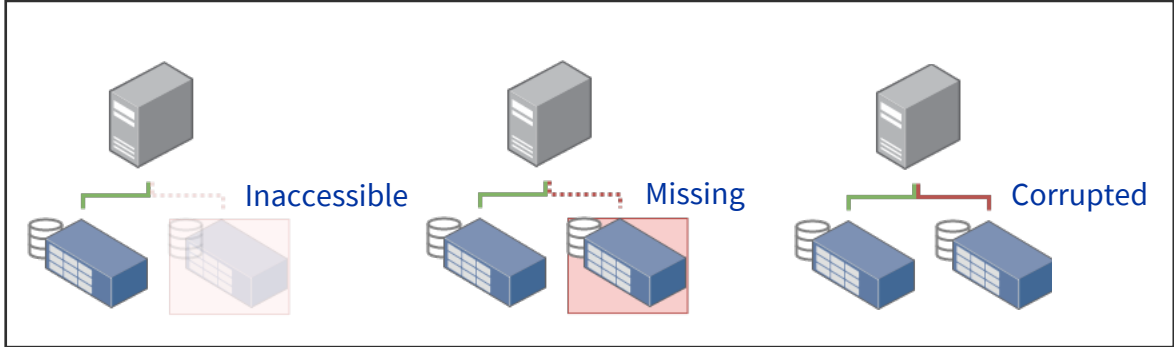
## For a complete analysis on the state a file, the checksums and sizes must be fetched from the different sources and cross-checked



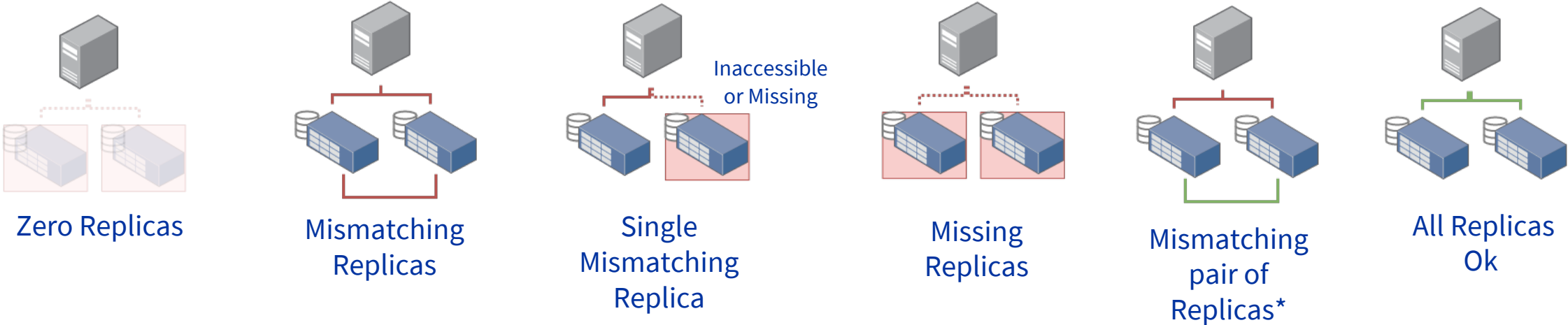
# Classification overview – branching

## The concepts to have in mind:

- Over or under replication (missing or temp. unavailable data)
- Mismatching data



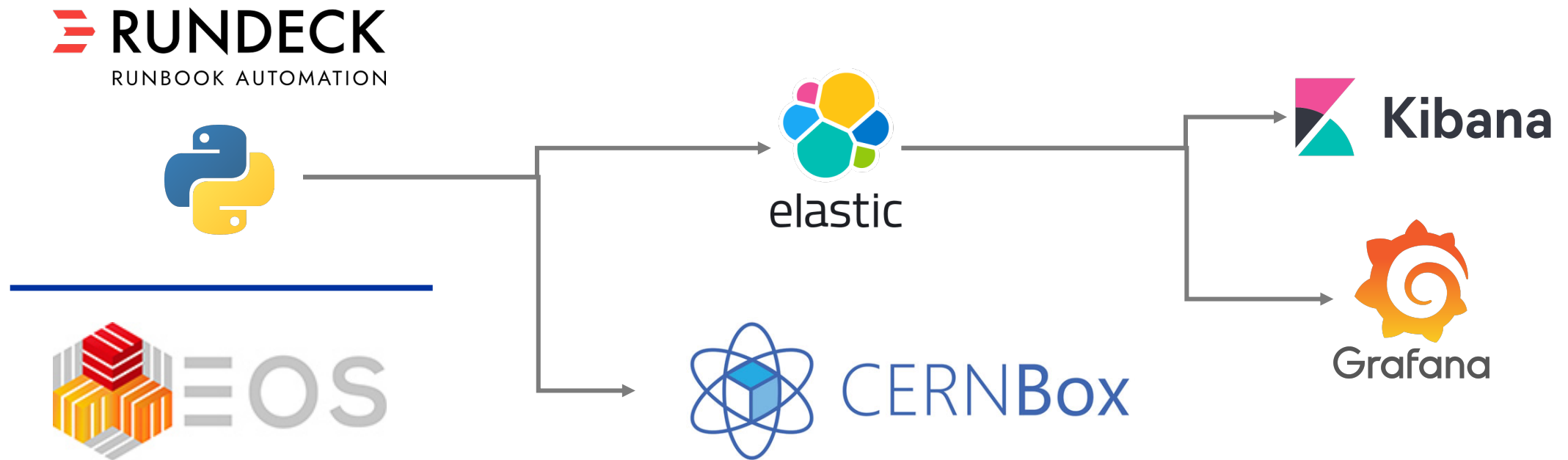
One Replica OK



# Tool-set overview

Detect, Classify, Repair & Monitors

Make the data available for further analysis



# What's being done

**Weekly summary indicating the weekly growth/decrease in the number of problematic files by classification**

**Centralized way to recover files from backups in homes and projects**

## **Tool-set consolidation:**

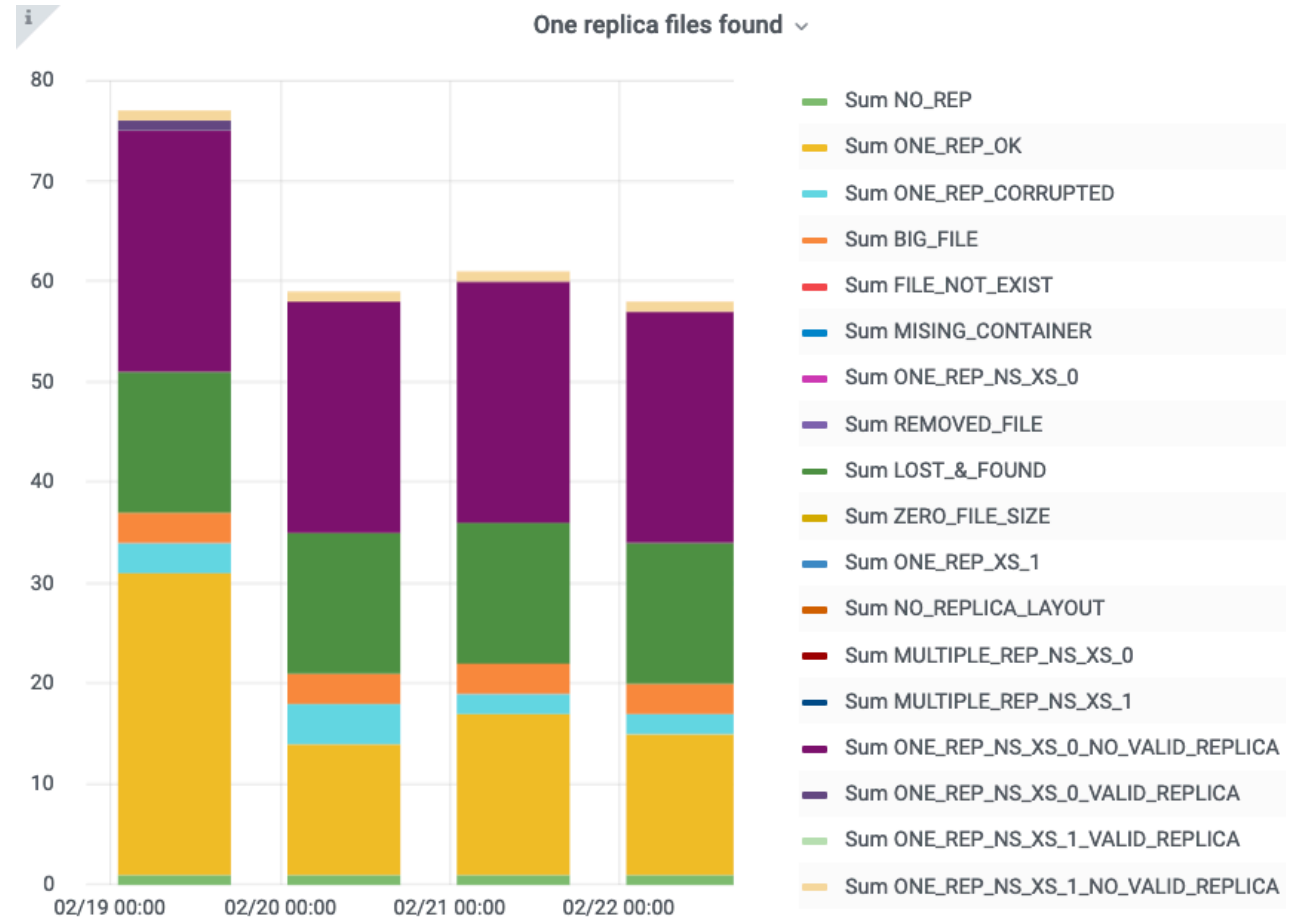
- Python3 migration: Python 2 no longer supported
- Redesigning and restructure tool-set (Refactor/Porte/Rewrite)
  - Including new and instance specific features
- Enhancing the repository CI pipeline
- Improving the management of dependencies and requirements.



# Goals

## Promote fast reaction and proactiveness

- Improve monitoring, alerting and reporting
  - Gather finer details.
    - FUSEX access; creation/modification timestamp, SW version
    - Increased correlation between SW version and file problem
  - Tag based classification
- Investigation - automated log search
  - Prioritizing recent problems and secondly resolve backlog



# Goals

## **Portability - extend the tools to a non CERN specific configuration**

- Flexible and configurable flows
  - Allow users to specify which actions to take for each scenario

## **Enhance integration with `eos-ns-inspect` and `eos fsck`**

- Wider analysis
- Completeness
- Sanity checks

## **Extend the tool to handle `rain` layouts**



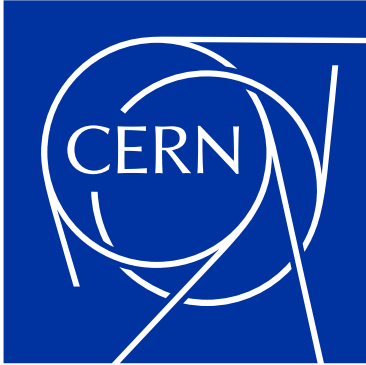
# Conclusions

**Important role on the self-healing of EOS and monitoring.**

**The first generation with its automatic detection and repair mechanism showed its value specially during the Wigner draining campaign.**

**The experiments instances are in a good shape and now we need to focus on new goals and problems. Targeting proactive investigation.**

**Provide more detailed information to the operators and developers regarding the source of the problems.**



[home.cern](http://home.cern)