

DPG, DPP & Other News

Predrag Buncic

CERN



ALICE

Step back

LS1 Objectives: Software Framework

Support data taking and processing

Production

- Use HLT farm for processing
- ~~CAF(s)~~ → CAF on demand

Calibration

- Online calibration
- Improve space-point calibration

OCDB

- Eliminate TPC specific storage

Software

Infrastructure

- Migrate SVN to IT
- SVN → Git
- Savannah → JIRA
- Cleanup obsolete code
- CVMFS for s/w distribution

Improve software quality

Continuous integration

- CTest/Cdash

QA site for RCs

- Monitoring
- Reporting

QA driven release process

Separate Analysis package

- Faster release cycle

Improve software performance

Simulation

- G3 → G3, FLUKA-VMC
- De-calibration for embedding
- Geometry for upgrade

Reconstruction

- Improve digitization (algorithms)
- Clustering (HLT → Offline)
- Tracking (external seeding)
- Use TRD point in track fit

Analysis

- ESD, EOD size
- Flat Data structures
- Format for data preservation
- microAOD

R&D

Parallelization studies

GPU algorithms

Low level vectorization

Optimize memory usage

Study alternative algorithms

- Create task force to continuously monitor performance and suggests improvements

Simulate scenarios for HLT/Offline integration

LS1 Objectives: Distributed Computing

Performance

Improve merging

- Reduce time to finish for jobs

Improve analysis efficiency

- Address Lego trains performance
- Better monitoring
- Identify slow jobs, hot files, problematic files and sites

Monitoring

Aggregation of IP Traffic

Correlate Job efficiency w/network traffic

Aggregate ML, QA and Logbook

MonaLisa API

QA and Testing

Continuous testing AliEn releases

Data Management

File access monitoring

Active Data Management

- Transparently move unused files to archive SE

Reduce size file catalog

- Remove GUIDs

Consolidation

jAliEn

- Minimize dependencies to other software components

Using CVMFS for software distribution

Prepare for Cloud(s)

Offline contribution to O2 TDR

Submission of the TDR & UCG to the LHCC
20/4/2015

Presentation of the TDR & UCG to the LHCC
2/6/2015

2015

February

March

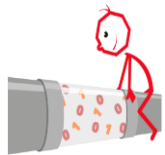
April

May

June



Architecture



Data flow



Data model



Computing platforms



Tools



Simulation



Calibration



Reconstruction



DQM



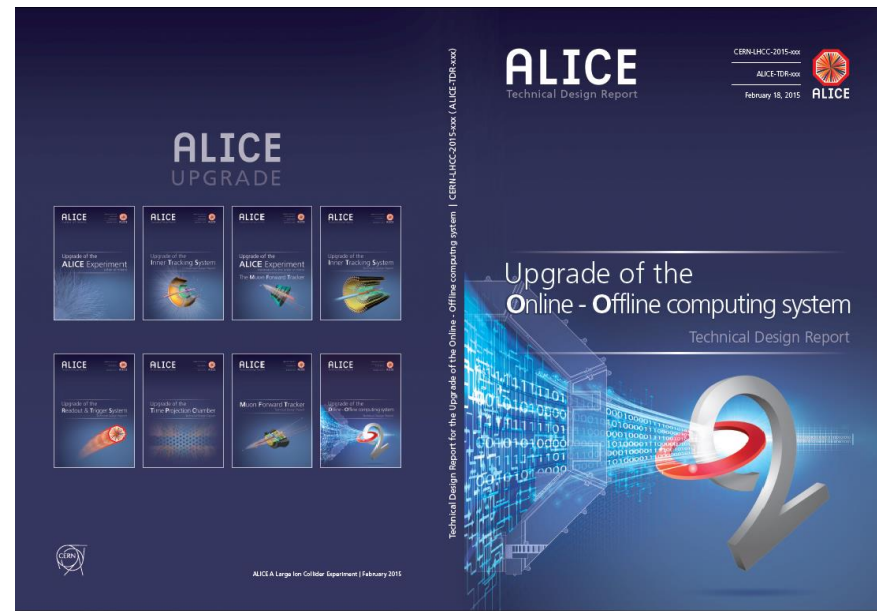
Control Configuration Monitoring



Software Lifecycle



Software Framework

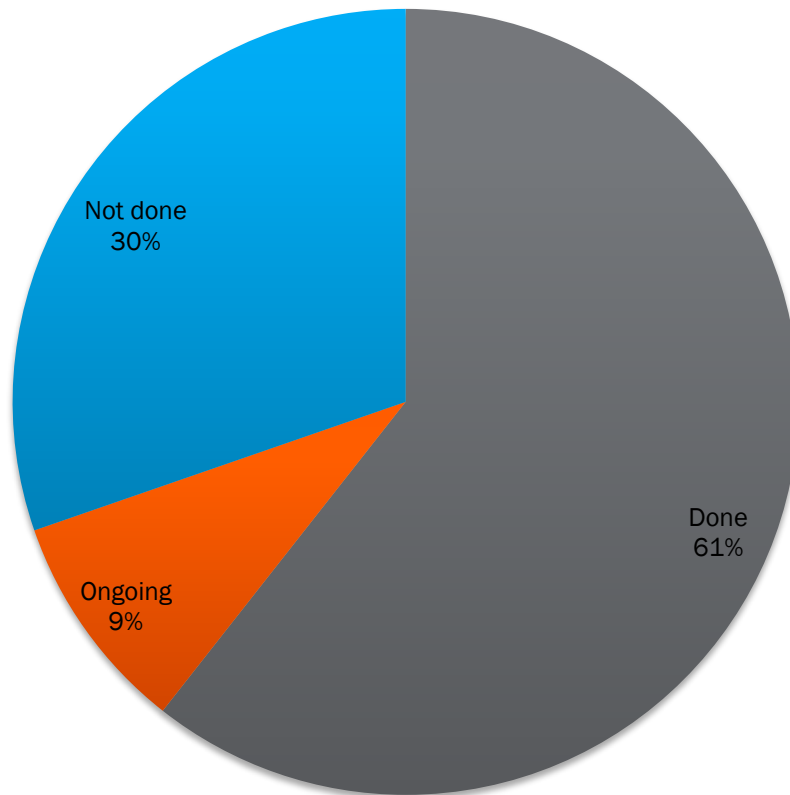


<https://cdsweb.cern.ch/record/2011297/>



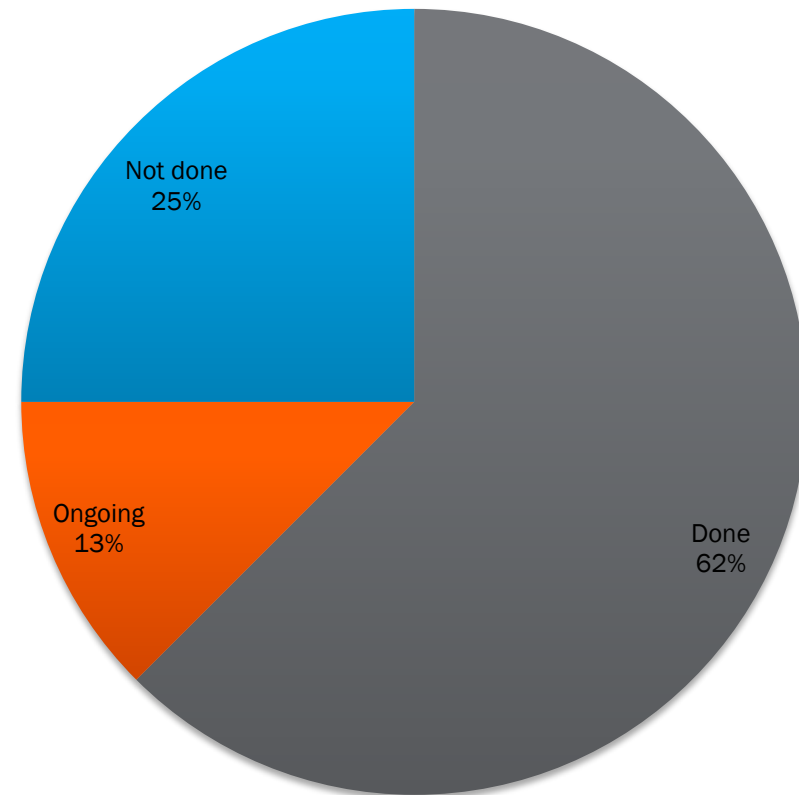
Task completion summary

Software Framework



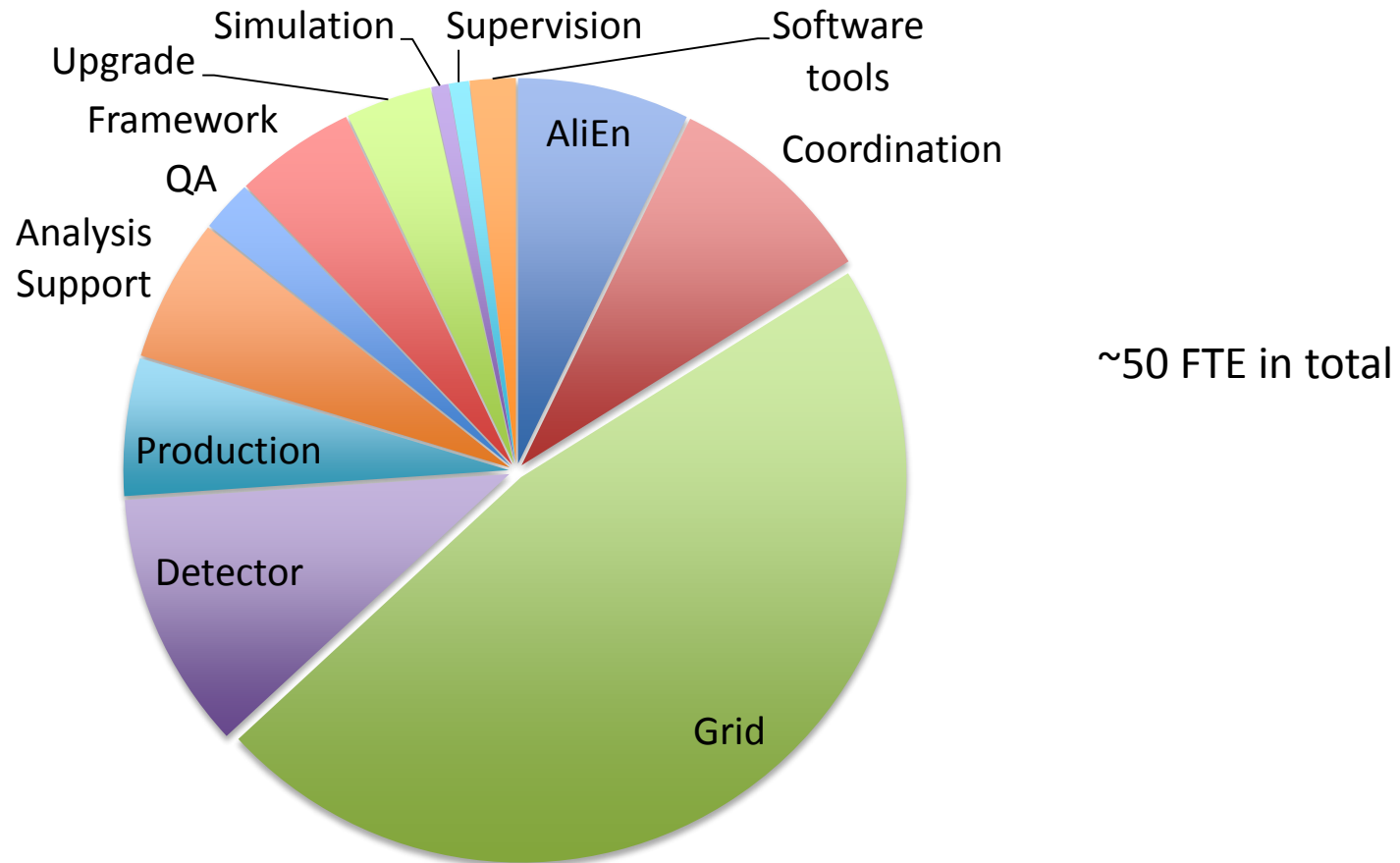
- Focus on functionality and software quality, not working on CPU performance

Distributed Computing



- Focus on operational stability and efficiency, avoiding possibly disruptive developments

Offline related effort in ALICE



- The biggest effort goes into Grid operations
- Almost no effort is invested in simulation
- Coordination is already consuming the effort comparable to Grid s/w development, framework development or detector software development

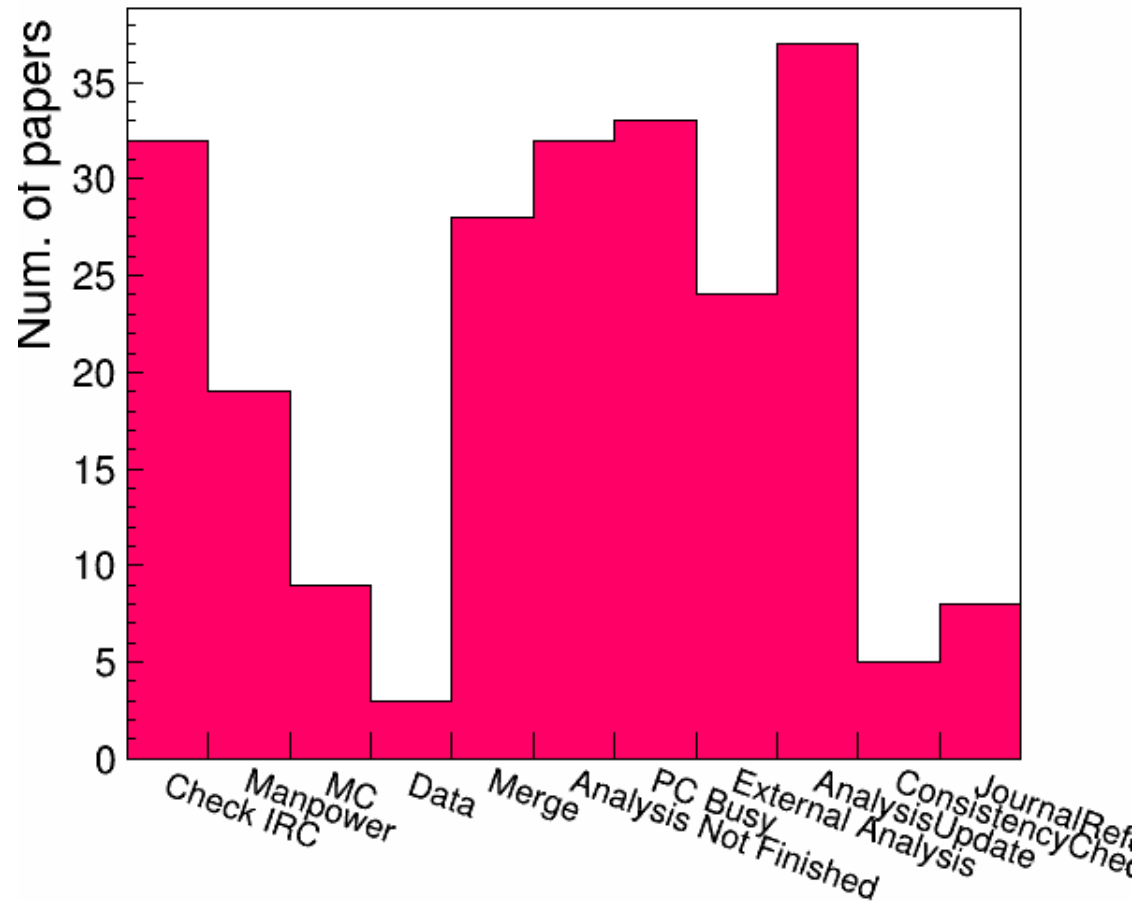


Manpower problems

- The perception is that data processing takes too long and is prone to many errors and delays
- Enough FTEs?
 - Enough people in the right place and able to do a job? No.
 - We need more people dedicated to offline activities in detector groups.
- More workers or more managers?
 - If the organization is OK, we need only more people
 - If not, we have to reorganize and/or add more people and more managers
- More experts or many more FTEs?
 - Expert retention and long-term stability remains a big problem
 - The alternative (overlapping short term staff) requires many more FTEs
- We need to do more with less

How to speed things up?

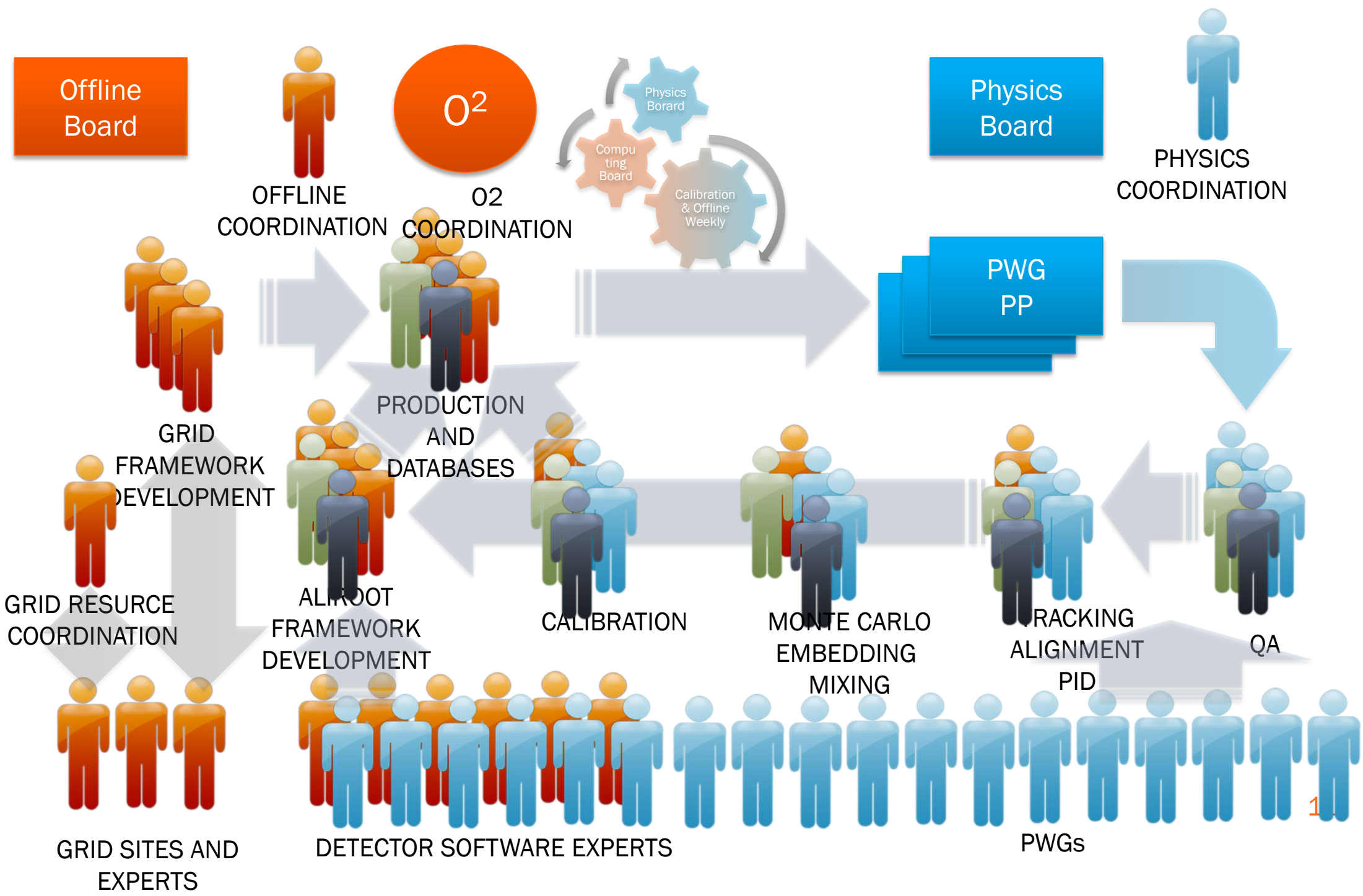
Reasons slowing down papers



➔ Main reasons slowing down:

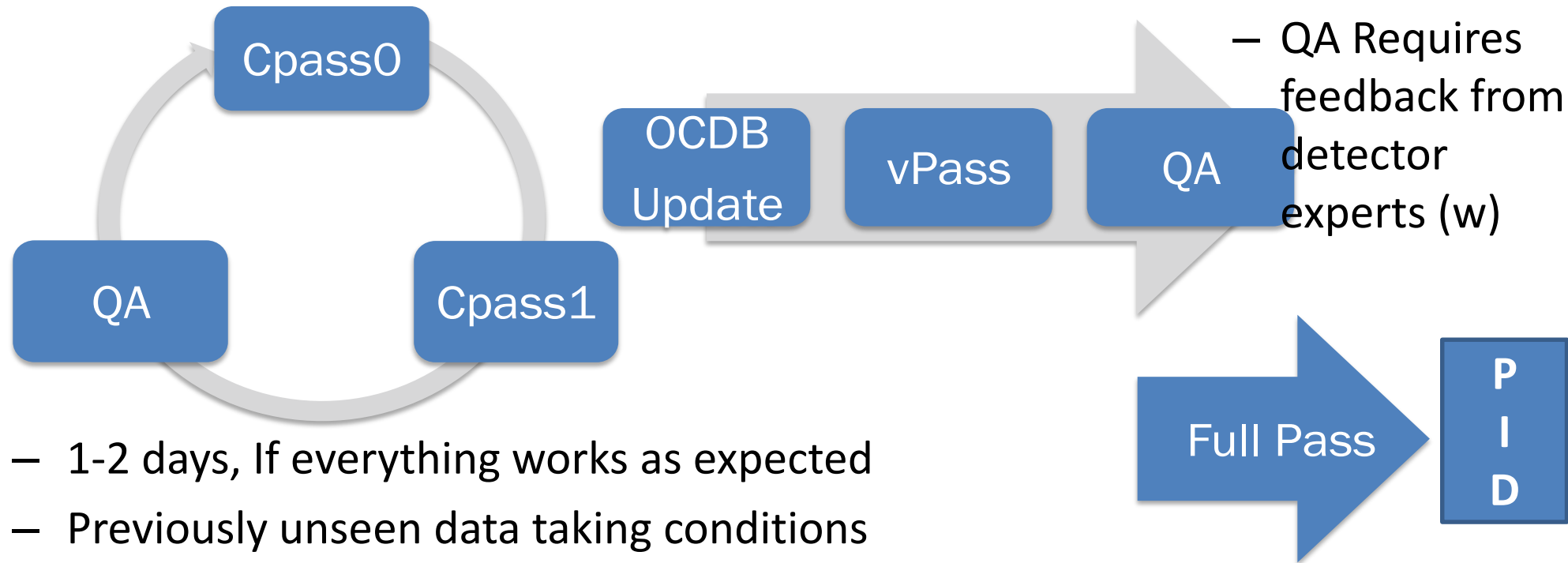
- Checks required by IRC
- Analysis not finished at PB paper approval
- Analysis update
- Merging of several analysis
- Missing info from external analysis
- PC busy

➔ Problems related to MC/data seem to affect only specific papers



Why data processing takes so long?

- Process designed to maximize good use of computing resources and minimize chances of bad data passed to people doing analysis

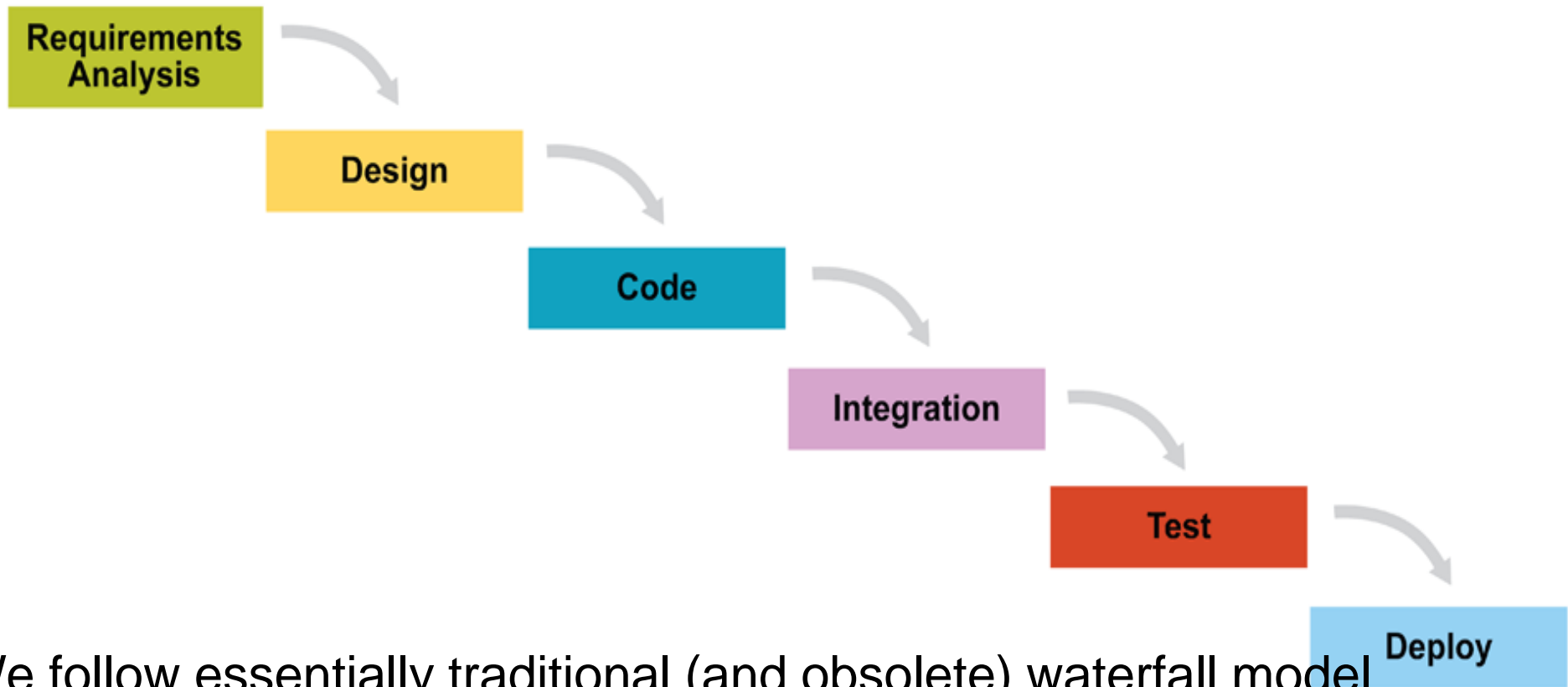


- 1-2 days, If everything works as expected
- Previously unseen data taking conditions
 - Require manual calibration
 - May require software update, validation, packaging, deployment, re-run, QA (w,m)

- Final PID is attributed at analysis stage (m)



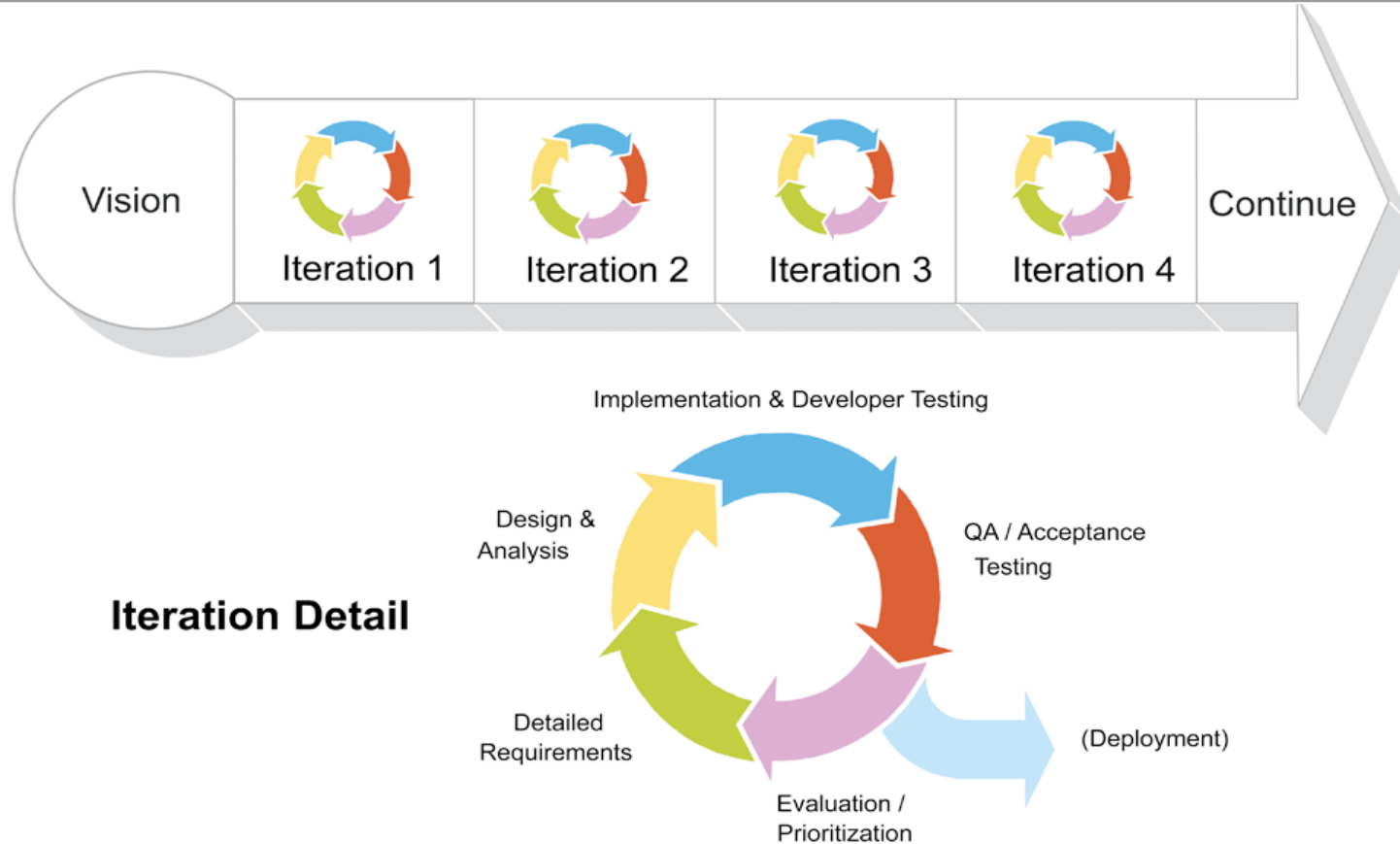
Waterfall approach



- We follow essentially traditional (and obsolete) waterfall model
 - Hoping to get everything right after one full reconstruction pass
- Software build, test, validation process is slow
 - Infrastructure put in place to support software QA not yet used on regular bases



Agile approach



- Trades the traditional phases of “waterfall” development for the ability to develop a subset of high-value features first, incorporating feedback sooner.
 - Incremental product development using one or more, self-organizing teams of about seven people each.
 - Structure of roles (project owner, scrum master, development teams) and meetings
 - Fixed-length iterations (sprints) which never more than 30 days.



Streamlining data processing

Cpass0

Cpass1

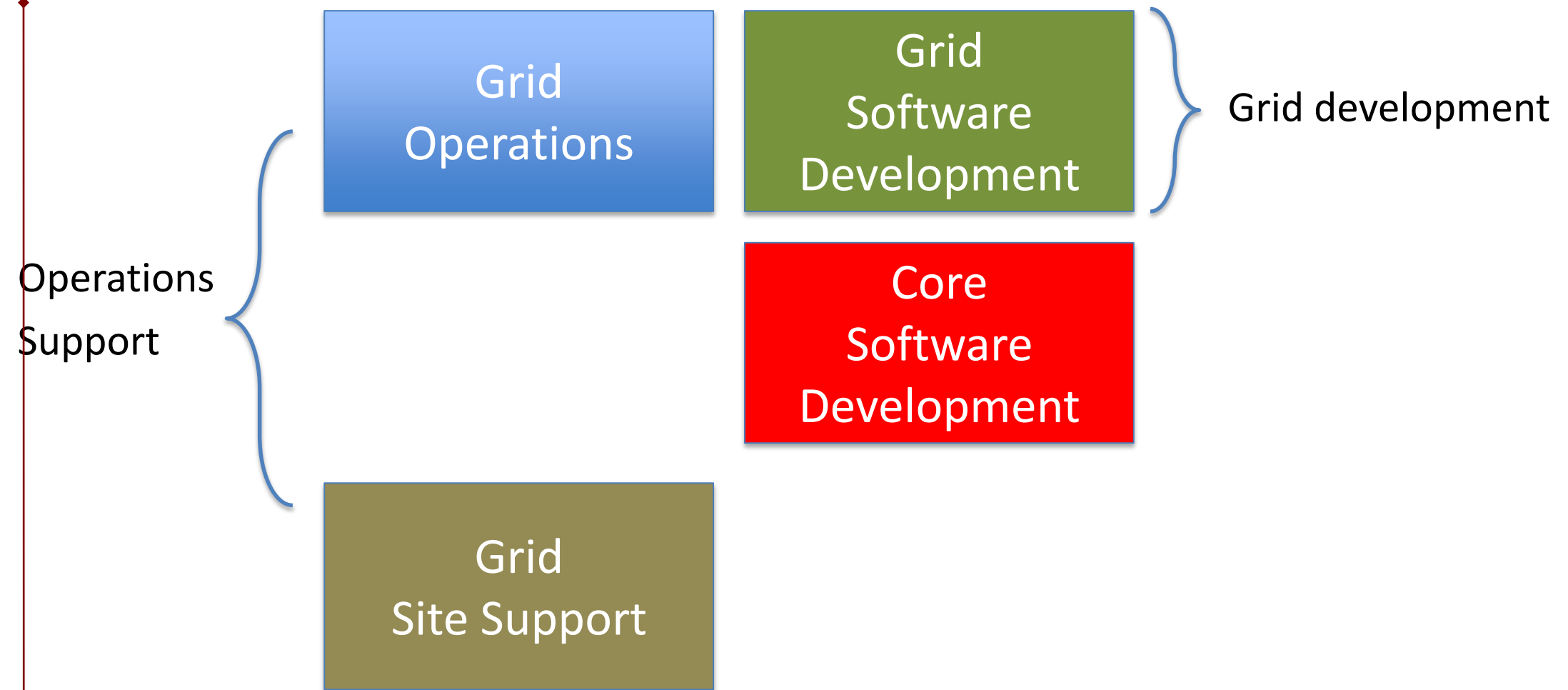
OCDB
Update

PPass

QA

- Instead of procedure with several QA checkpoints that currently often resets the process and involves developers, try to run the calibration/reconstruction until end and only then do comprehensive QA
 - Do not stop for 'broken' runs (whatever the reason), recuperate these in subsequent passes, if problems fixed.
 - Only hard stop if reconstruction fails for all
- Define upfront the quality of reconstructed data that need to be achieved after each reconstruction pass and stage/prioritize the analysis activities accordingly
 - In general, this will allow for a fast initial QA and some types of analysis

Streamlining the organization





Streamlining the organization

Operations Support

Grid Operations

Grid Software Development

Grid development

Core Software Development

PWG_PP
Tracking, Calibration, QA,
Simulation

Grid Site Support

- Calibration
- Global Tracking & Alignment
- Simulation
- Analysis framework
- PID



Streamlining the organization

Grid
Operations

Grid
Software
Development

Core
Software
Development

PWG_PP
Tracking, Calibration, QA,
Simulation

Grid
Site Support

Detector
Software
Teams

- Calibration
- Reconstruction
- Detector simulation
- QA

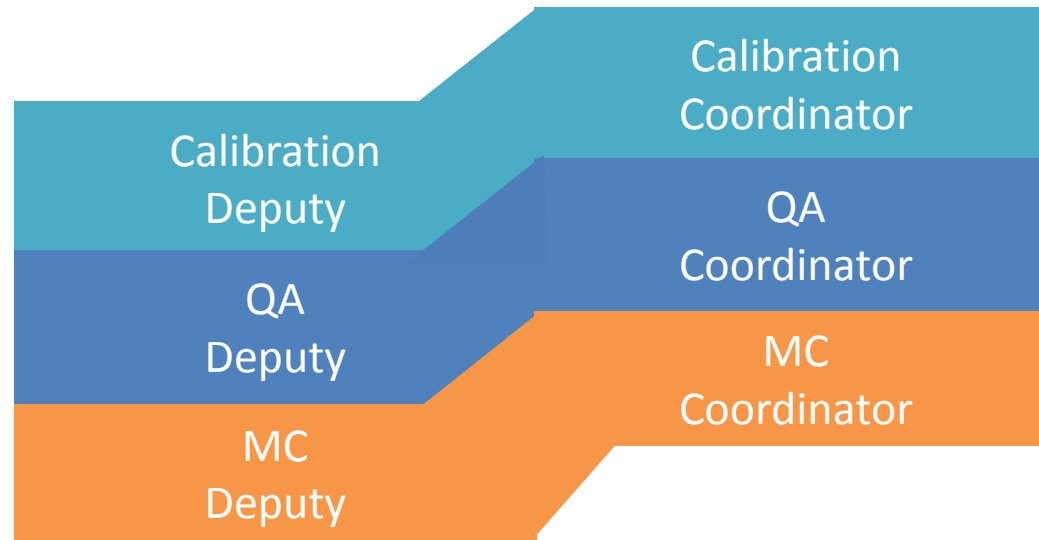
- Engage additional manpower from the collaboration and optimize use of the existing manpower by reorganizing the responsibilities between Offline and PWG-PP
- The proposal is to separate development and operations activities in order to
 - Force the experts to document all procedures and allow operations to be run by the non-experts
 - Allow experts to focus on longer term developments
 - Concentrate tracking expertise from relevant detector groups
 - Create a group in charge of data preparation according to agreed upon plan and aiming to achieve desired quality of data after each processing step



Central Barrel Tracking Group

- Evolution of Tracking Task Force
- Combined the effort and expertise from
 - ITS, TPC, TRD, TOF
- Delivers combined (global) tracks and PID
- Responds to calibration and QA requests by DPG
- Includes manpower from all participating detector projects + additional manpower from the collaboration
- Barrel Tracking Group has a coordinator and a deputy

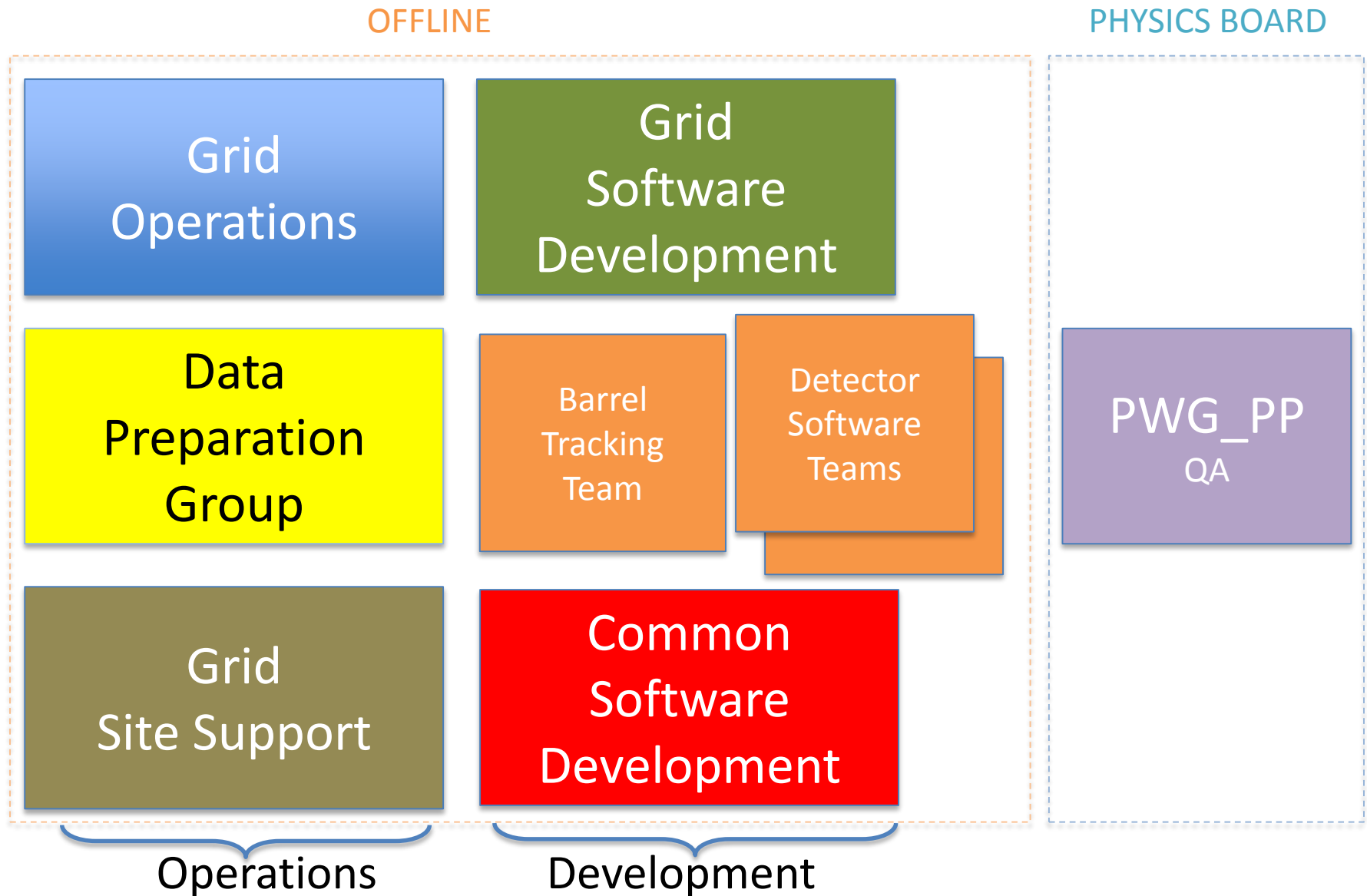
- With Physics Board define the quality that needs to be met by the reconstruction/simulation software in order to perform the required measurements
- With Run Coordination define the usable datasets that need to be reconstructed
- With PWGs define datasets that need to be simulated
- Coordinate and executes the calibration and QA activities that lead to reconstruction
- Coordinate, configure and test the corresponding MC productions
- Define and describe the datasets useful for analysis
- Issue tickets to Offline and detector experts for blocking software issues
- Provide the status reports to Physics Board get the input on priorities
- In cooperation with Offline and Physics Coordination Identifies areas and/or open tasks where the manpower is needed and defines service tasks to cover them



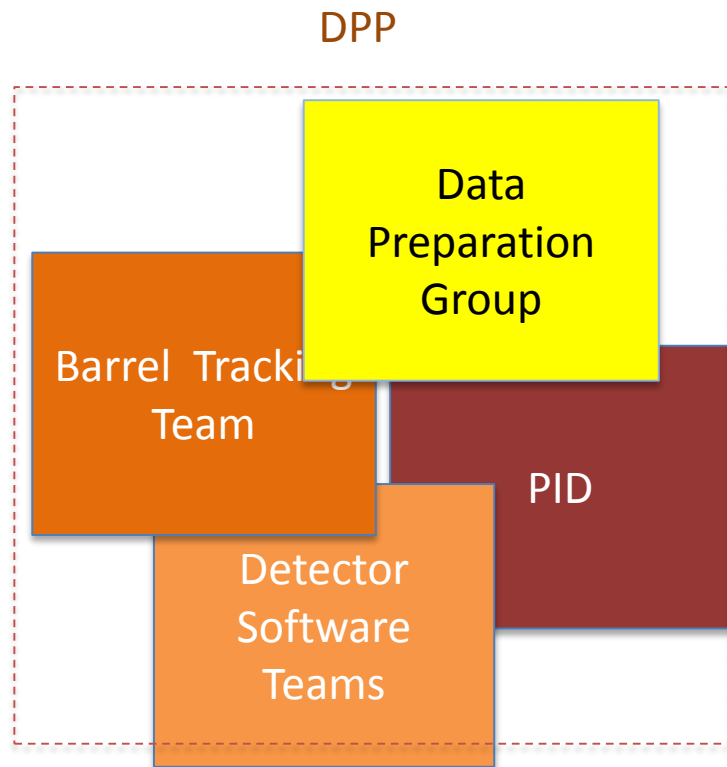
- Operations team (not necessarily the experts and developers themselves)
 - Facilitators, their job is to remove any impediments that obstruct a team's pursuit of its sprint goals
 - 1 year mandate (6 months as deputy + 6 months coordinator) @ 50%
- Reconstruction does not figure explicitly here
 - Preparation work shared between Calibration and QA,
 - Execution carried out by the Offline



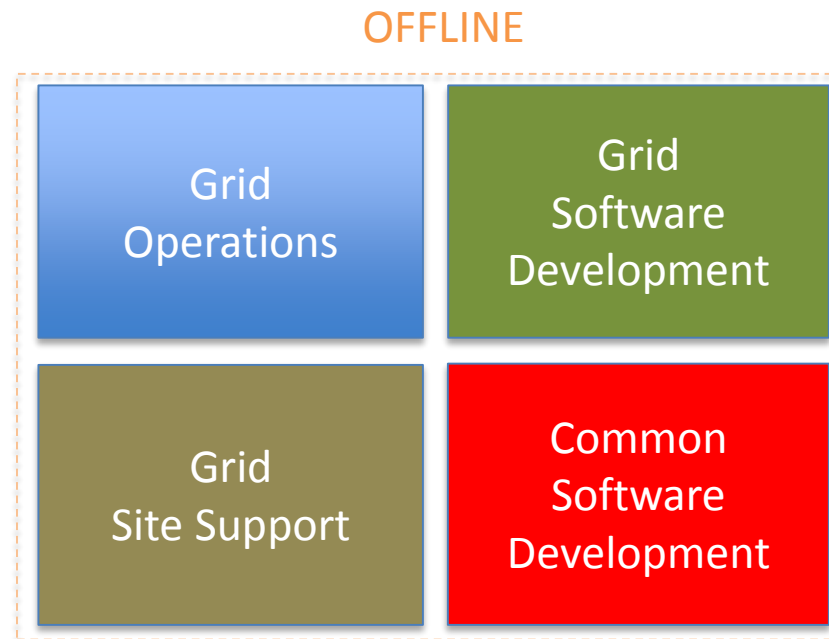
Incubation stage



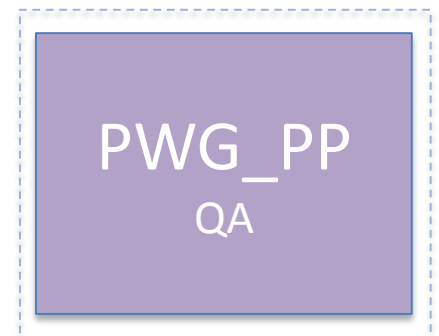
Final stage: Data Preparation Project



- Once DPP is established and demonstrates its viability we will proceed with creation of an independent project charged with data preparation and detector software developments



PHYSICS BOARD



- **Offline Role:**

- The Offline will focus on Computing aspects (Grid) and common software development (frameworks) as well as its contribution to O2 project

- For this plan to work, the detector groups must find sufficient and qualified resources to fulfill their obligation
 - Detector simulation
 - Calibration
 - Reconstruction
 - QA
- At least in the Grid operations/development there is a large amount of overlap between the operations and development
 - We need to find a way to strengthen that corner
- By dividing the responsibilities between development and operations we are risking that part of the collaboration will consider another part as “service providers”
 - This is seen very often when a bug is observed and no effort of debugging is done

- Most of the planned software improvements during LS1 were completed
 - ¼ of planned tasks cancelled or not successfully completed
- The accumulated delays can be attributed to a general lack of manpower, increased load due to work on O2
- Streamlined data processing proposed for Run 2 using modified QA process and agile development approach
 - Data Preparation Group and Central Barrel Tracking Team to be created within the Offline Project to incubate and prepare the ground for the future Data Preparation Project
 - Data Preparation Project should take the ownership of detector software stack (calibration, reconstruction, PID) and QA process
- The Management Board endorsed this proposal and now we have to start to implement it