

COD activity in EGI-InSPIRE

Marcin Radecki

CYFRONET, Poland & COD Team

- Overview of Grid oversight
- Highlights on what is important for keeping the infrastructure stable
- Operational model
 - procedures
 - ~~tools~~
- Operational model metrics

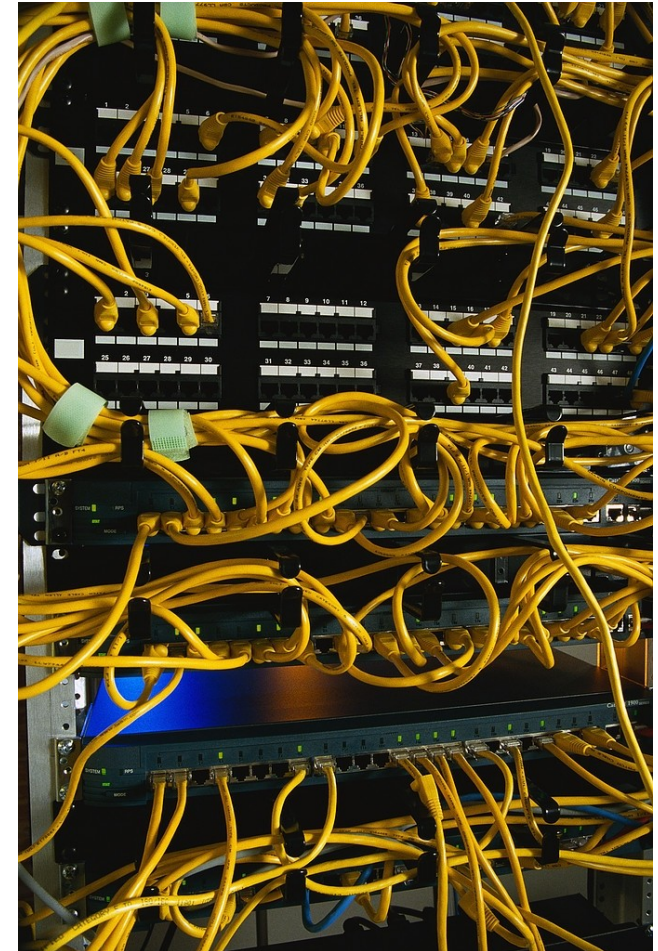
Grid Oversight – what's that?

“Ensure that sites, operational and middleware services are functional, reliable and responsive”

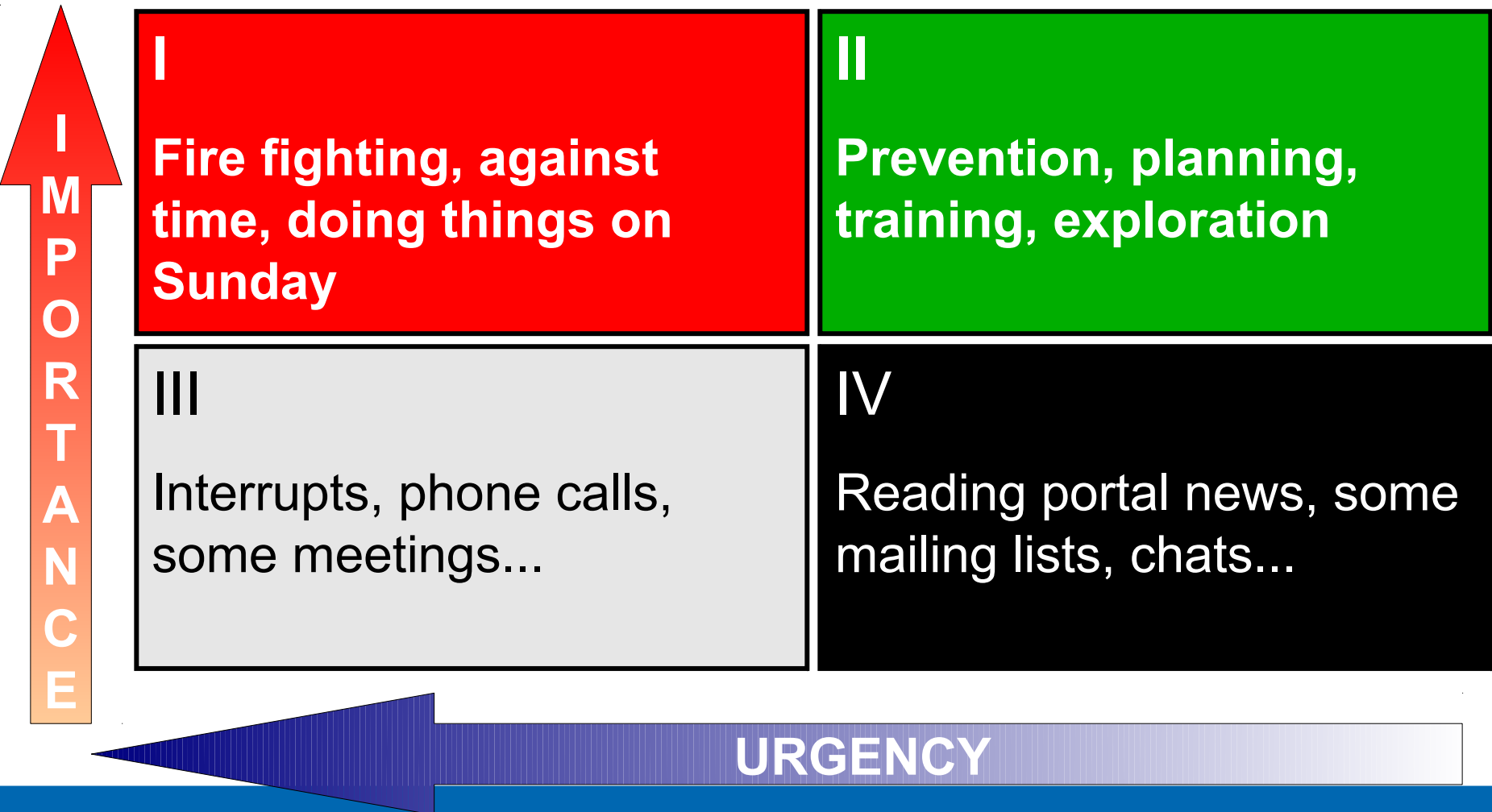
EGI-InSPIRE DOW

- Make sure operational problems are **detected**, **diagnosed** and **fixed** efficiently
- EGI resources are located in more than 300 sites, 59 countries
- NGI coordinates sites within country and:
 - provide infrastructure monitoring (nagios)
 - provide operations support (ROD)
 - provide helpdesk facility
 and **interface all the above** with EGI to provide a consistent all NGI view

**Grid Oversight requires
procedures, tools and staff
to make the infrastructure working well**



Principles of being effective



- notice a problem ASAP
- diagnose
- act precisely (without dead ends and U-turns)
- The above requires:
 - _ tools (monitoring, dashboard, GGUS)
 - _ well defined procedures
 - instruction on how to proceed in case of a failure
 - cover all aspects, details, nuances
 - _ collaboration
 - exchange experience, pass knowledge, get help on-line



- Service availability monitoring in Grid
 - Services are remote – impact of computer network
 - Complexity of Grid middleware
 - monitoring functionality for the user (replica management)
 - ...vs. testing atomic functionality
 - middleware error messages: <https://twiki.cern.ch/twiki/bin/view/LCG/BestErrorMessages>
 - Nagios – a monitoring system aware of the dependencies between functional components
 - do not tests services on a host if the host is not reachable
 - also a source of issues during transition from SAM to nagios...

- What is reported to site admin?
 - command which returned an error
 - error message e.g. (top 4):
“CGSI-gSOAP: Error reading token data: Success”
- **Experience is indispensable**
 - ...or support
 - documentation
 - knowledge base etc.



- Ideas that will not work
 - _ Search the error message and explanation in middleware manual
 - _ Ask the middleware developer for help
- Time consuming ideas
 - _ understand the software by yourself “Use the source (code), Luke!”
- Practical, working (usually) solution
 - _ search the knowledge bases
 - <http://goc.grid.sinica.edu.tw/gocwiki/SiteProblemsFollowUpFaq>
 - <https://weblog.plgrid.pl/baza-wiedzy/>
 - be warned some entries may be out of date!
 - _ see if someone encountered it before
 - in GGUS tickets – there is nice search engine, worse than knowledge base as may contain no solution
 - _ ask expert
 - your NGI 1st line support
 - post an e-mail to lcgrollout mailing list

- Define who and what does in case of a service failure
- Actors
 - Site Admin
 - Regional Operator on Duty Team (ROD)
 - 1st line support – technical support for Site Admin
 - Central Operator on Duty Team (COD)
- Items to operate on
 - **alarm** – problem reported by monitoring system. Contains info about time, localization of the failure. Appears in dashboard of ROD and COD.
 - **ticket** – record of a problem handling. Created when an alarm cannot be quickly turn off. Exists in GGUS (or regional helpdesk)

Operations Support Model Time Principles

- Model depends on timely actions
 - first 24h – time for site & technical support team
 - [24,72) - time for ROD to clear the problem OR record it in GGUS
 - [72,∞) - model malfunction, COD comes into the game
 - ticket not handled on time (expiration date passed) → COD
 - ticket not solved in 30 days → COD
- Metrics aim: indicate problems with operating model
 - items not handled on time
 - items not handled according to procedures
 - assess workload on ROD & COD teams

Model implemented (example)

Monday, 7 P.M.
Tuesday, 8 A.M.
Tuesday, 9 A.M.
Tuesday, 7 P.M. 24h passed
Wednesday, 8 A.M.



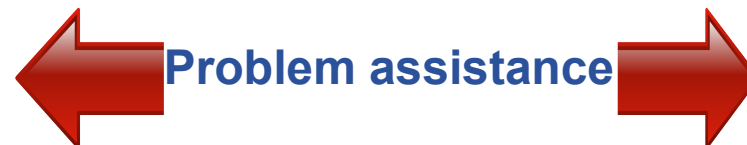
Regional Dashboard



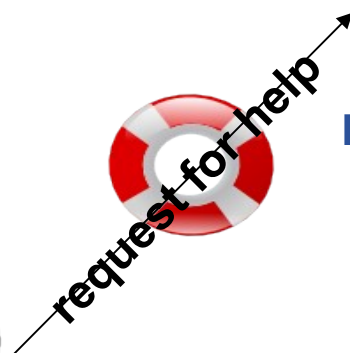
Regional Operator



1st line support



Problem assistance



Site

An „item” in the dashboard is either alarm or ticket that the relevant party (COD, ROD, 1st line) should take action upon.

Description	Number of items appearing in COD dashboard indicates the amount of work that the operator has to deal with. It could also be used to assess the quality of support process . There should be no items in COD dashboard if the support process is working in a timely manner.
What is measured	Number of items in COD dashboard that need immediate action, appearing on a given day. Items not done on a given day will be counted again the next day.
Purpose	To estimate the amount of daily work of COD operator and quality of support process.
Source of data	Operations Portal

An „item” in the dashboard is either alarm or ticket that the relevant party (COD, ROD, 1st line) should take action upon.

Description

Number of **items appearing in ROD dashboard** indicates the amount of work that the operator has to deal with. In general it **cannot be used** to assess the quality of support process.

What is measured

Number of items in ROD dashboard that need immediate action, appearing on a given day.

Purpose

To estimate the amount of daily work of ROD operator.

Source of data

Operations Portal

Metric = (alarms_closed_with_OK/alarms_closed_in_total)

Description

Regional ops. support staff can close an alarm if the actual state of the service is OK or some ERROR state. In general **they should fix problem** and close alarm only if the actual service state is OK.

What is measured

Fraction of alarms closed with OK status over some time period e.g. 1 month.

Purpose

Assess regional support quality, make sure model time rules are followed.

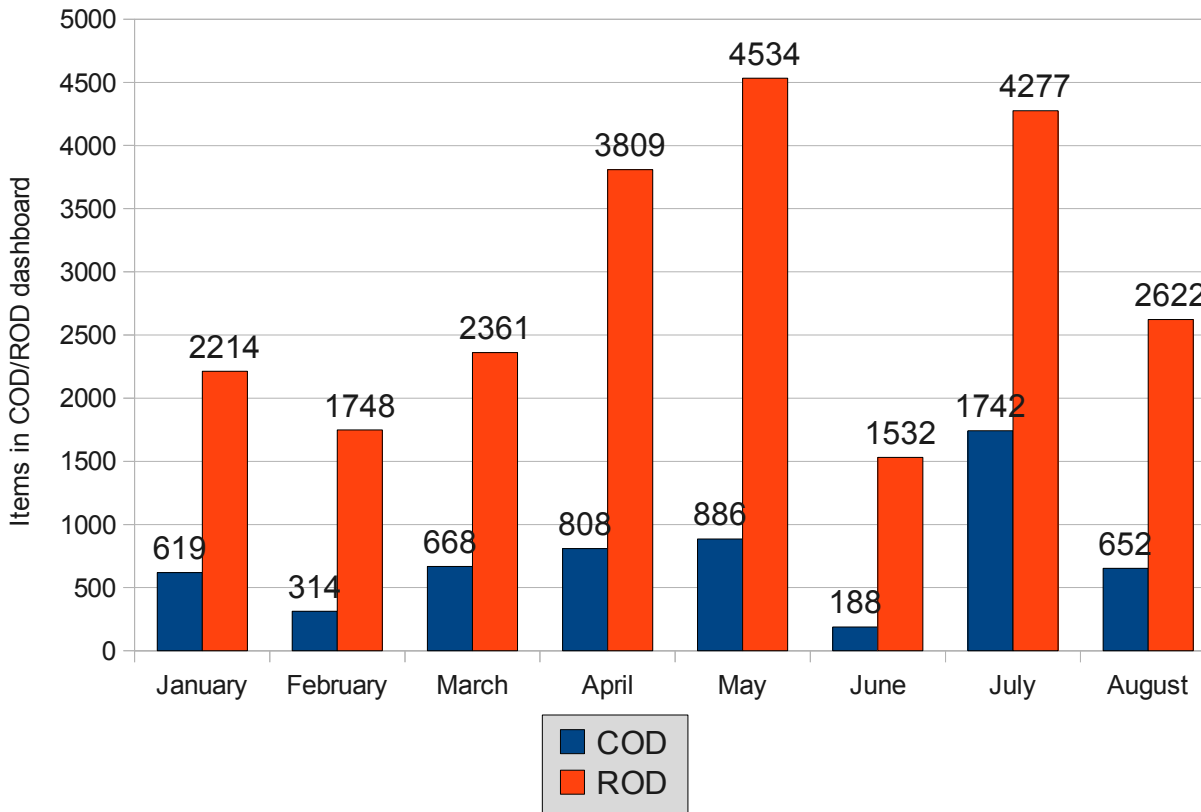
Source of data

Operations Portal

Workload: values

COD/ROD workload

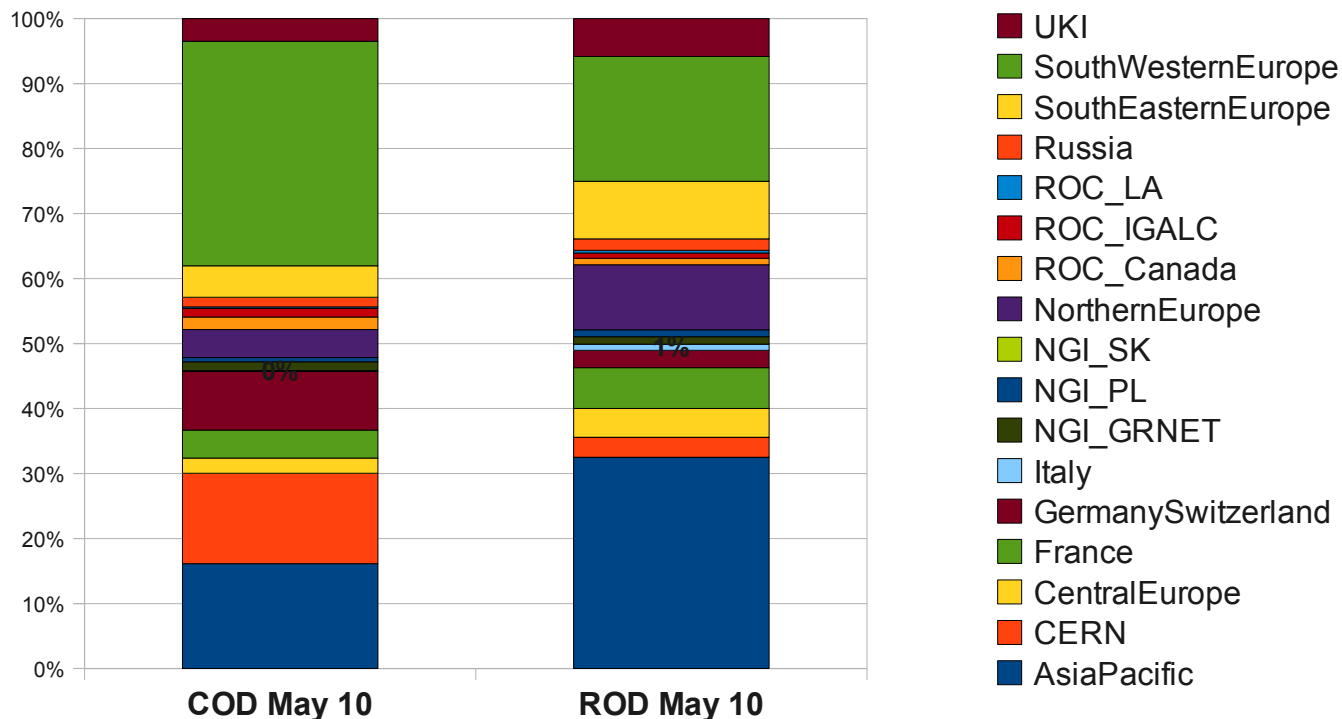
January - August 2010



- March - May 2010
 - _ New monitoring system introduced
 - _ End of EGEE-III, staff change
- May 2010
 - _ end of SWE ROC – generated 35% of COD workload
- July 2010
 - _ CERN COD workload share: 71%
- Conclusions
 - _ RODs do a good job
 - _ Alarms should not age on bank holidays

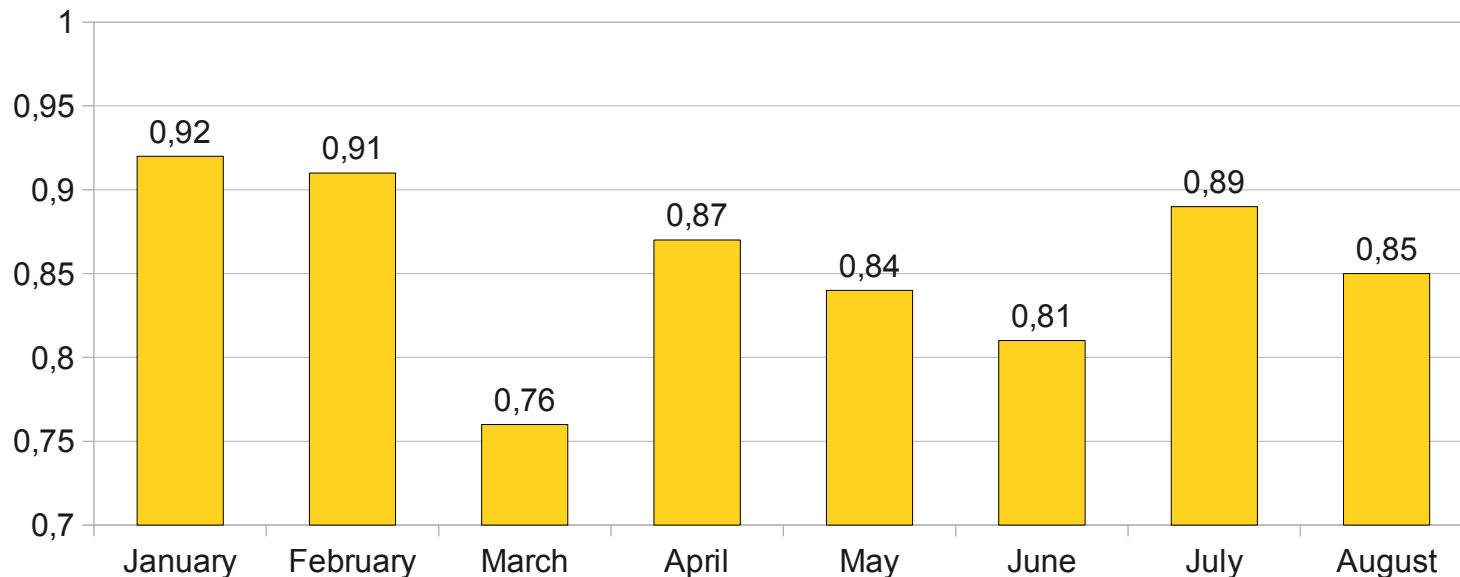
Workload Origin Analysis

C-COD and ROD workloads in May 10



- Metrics data sheets: <https://documents.egi.eu/secure/ShowDocument?docid=155>
- Daily metric data can be found also at: <https://operations-portal.in2p3.fr/dashboard/metrics>

Quality
Jan - Aug 2010



- **March** – new monitoring system introduction
- Known issues: <https://wiki.egi.eu/wiki/Operations:OperationsSupportMetrics>
- Operations Support Metrics were proposed as QR contents

- COD work is consistent with demand on higher site availability, however...
- Ratio is not so straightforward
 - example: if you have more CEs then you can have a lot of failures on one of them and still 100% availability

...but if you have “0” ROD workload it is likely you're 100% available ;-)

Questions

