

DevOps: the perfect ally for Science Operation

Rocio Guerra - ESA

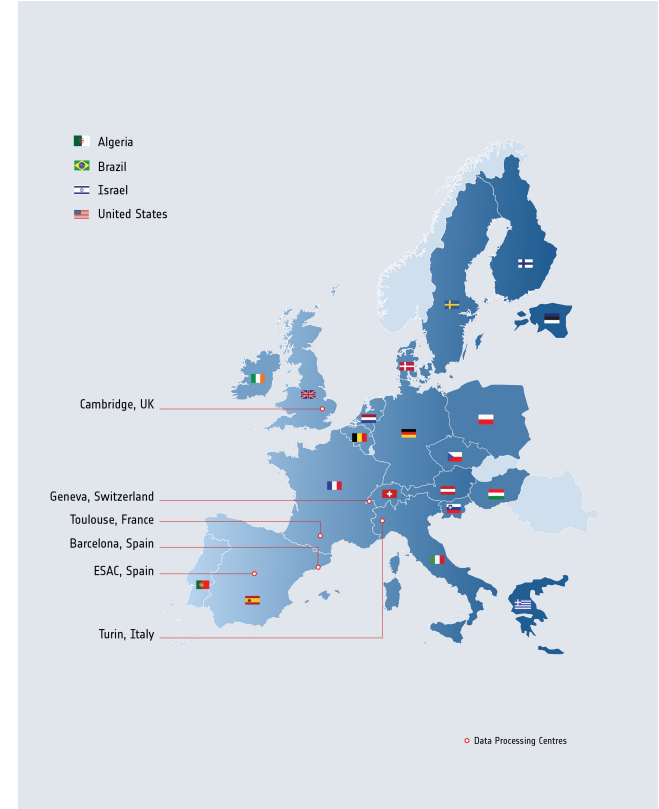
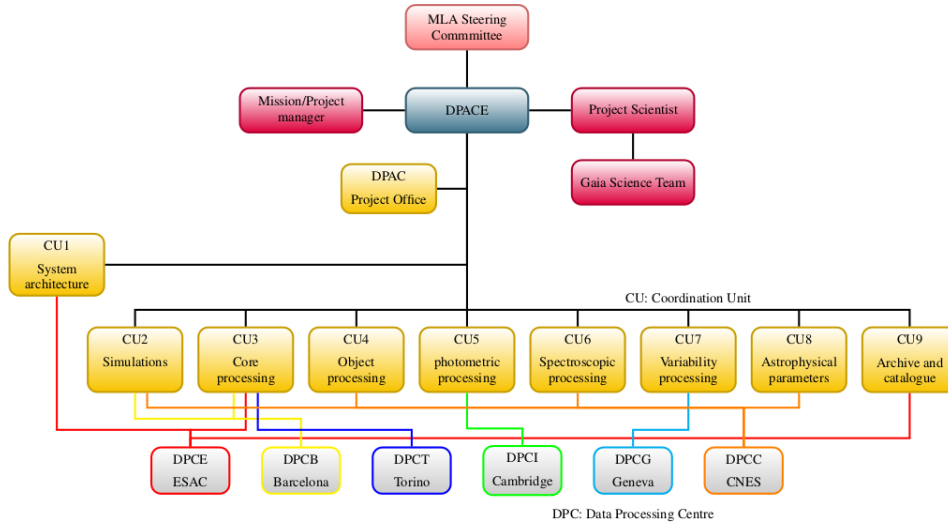
ADASS XXVIII, College Park MD, 15th November 2018

Gaia Data Processing - DPAC



• Large pan-European team

- Responsible for the processing of the Gaia data
- 450 expert scientists, developers and engineers
- 160 institutes, 24 countries + ESA



Gaia SOC from ESA viewpoint



- **Gaia SOC is an integral part of the Data Processing and Analysis Consortium (DPAC)**

- **A bit unusual for an ESA "classical" roles:**

- MOC's primary contact for all the payload/science-related mission aspects
- Mission planning product generation (Scan Law, Science Schedule, etc.)
- Payload health monitoring
- First-level product generation and downstream data dissemination
- Archive development + operations

- **"Special" roles:**

- CU1: System architecture + central repository (Main Database)
- CU3: Development + Operations of AGIS
- CU9: Support for catalogue creation
- DPCE: Data Processing Centre at ESAC

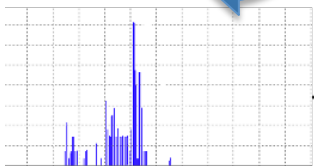
- **Cyclic and Daily processing**



Daily Pipeline



TM + HK ~ 35GB daily



Input
TM files

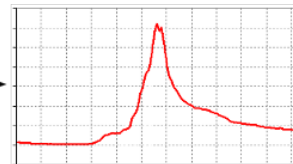
70M transits/day



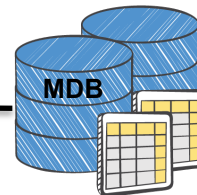
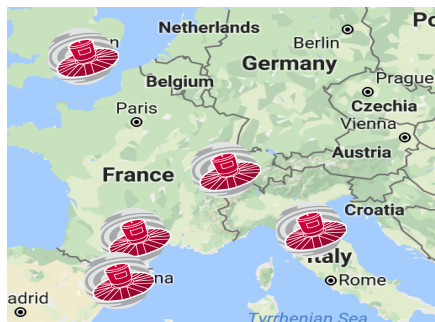
Output



Gbin files



~ 300-500GB



Our challenge: dealing with complexity!

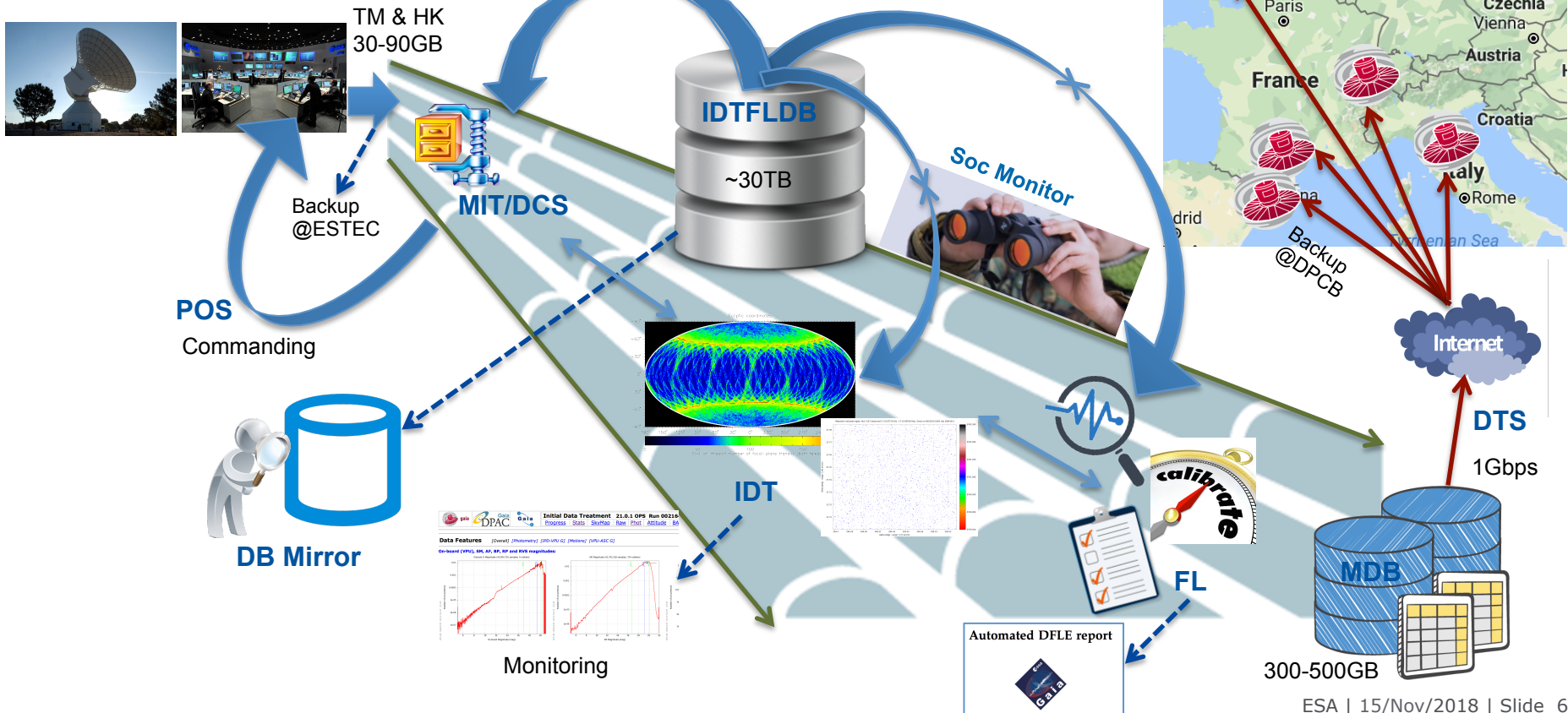


- **Operations are very complex**
 - Very ambitious top level requirements
 - Near real-time processing
 - Systems tightly coupled: many dependencies
 - Systems developed by external parties -> hard to integrate
- **Very strict performance requirements**
 - Science alerts can't wait! + Payload monitoring
- **Data volumes**
 - 35-90GB of TM inputs become 300-500GB to be served to the DPCs daily

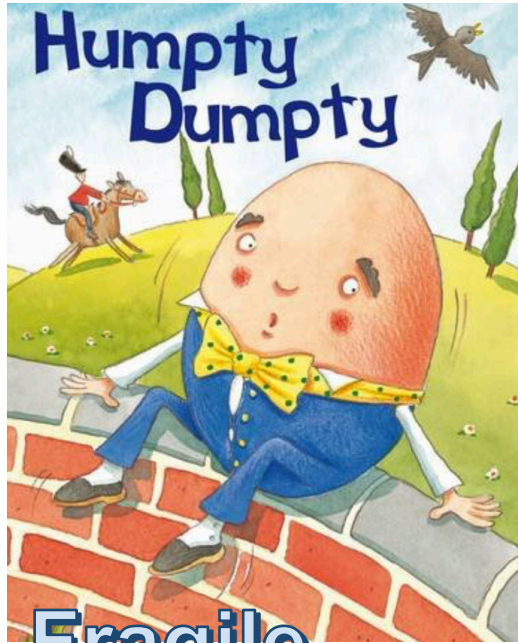
CURRENT DATE AND TIME	2018-11-08T18:55:59 (TCB)
MISSION STATUS	
Satellite distance from Earth (in km)	1,569,642
Number of days having passed since 25 July 2014	1567
OPERATIONS DATA (collected since 2014/07/25)	
Volume of science data collected (in GB)	58,834
Number of object transits through the focal plane	112,746,221,016
Number of astrometric CCD measurements	1,111,355,607,154
Number of photometric CCD measurements	224,751,711,050
Number of spectroscopic CCD measurements	21,550,324,020
Number of object transits through the RVS instrument	7,193,871,670



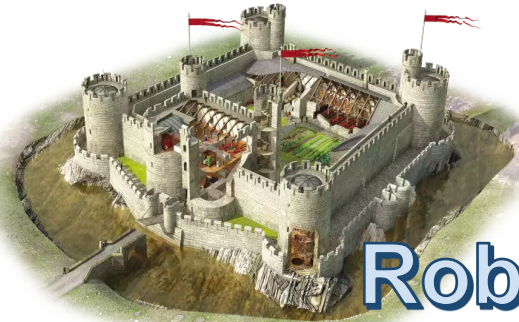
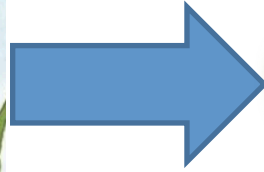
REAL Daily Pipeline



Our mission: creating a robust system



Fragile



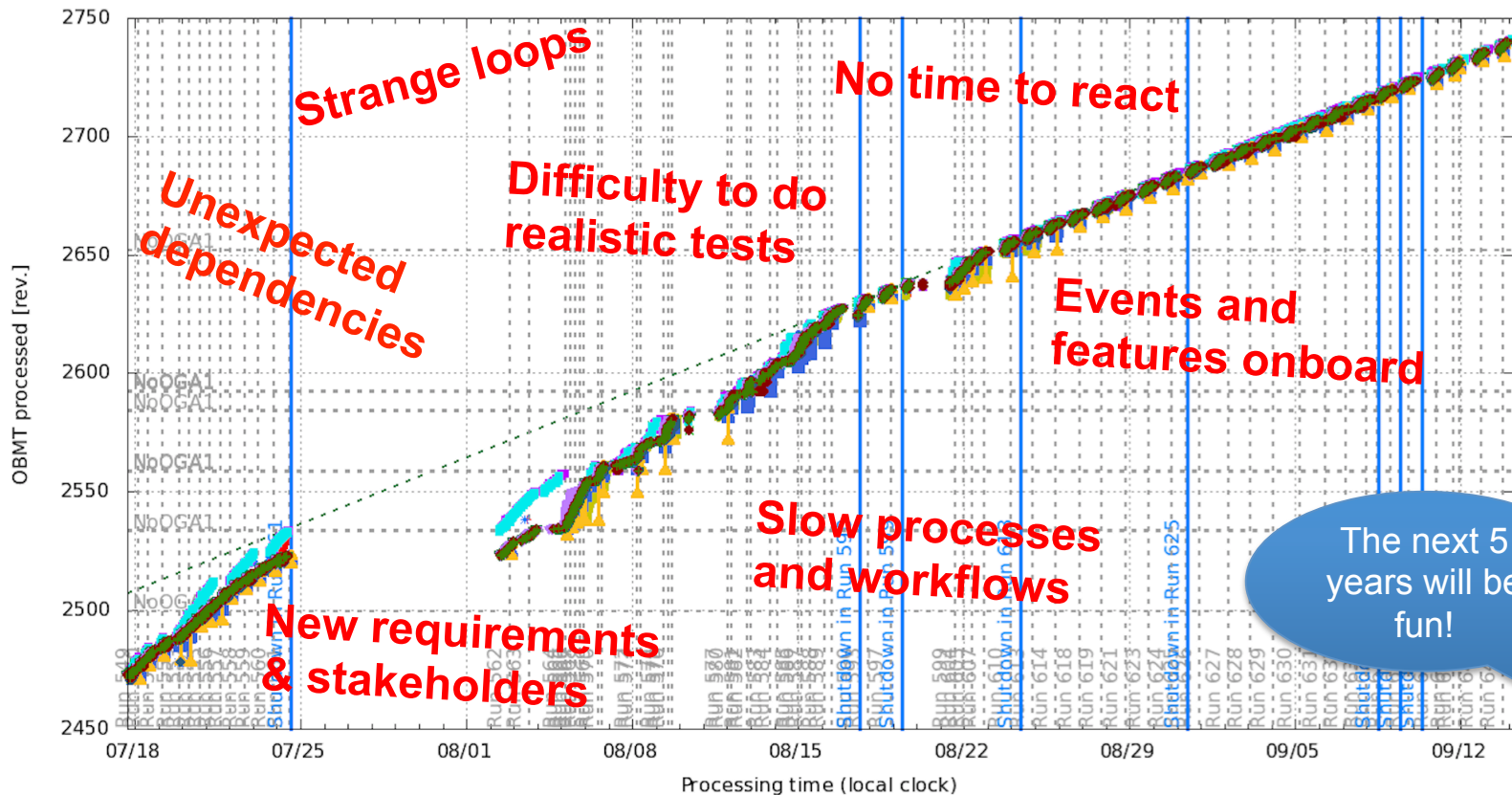
Robust

Highly Reliable
Scalable
Available

- **Robustness achieved by:**

- Large simulated datasets provided at different levels
- Tests campaigns (system, integration) including E2E and Operational Rehearsals – Test Readiness Review and Review Boards
- Formal CCBs, bug tracking and definition of strict issues workflows and approval cycles
- Identification of interfaces with the stakeholders
- Definition of detailed Operational Procedures

It seems robustness was not enough





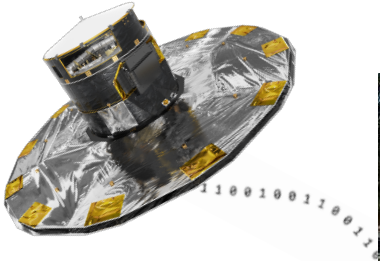
- **Military field: Team of Teams, 2015. General McChrystal**

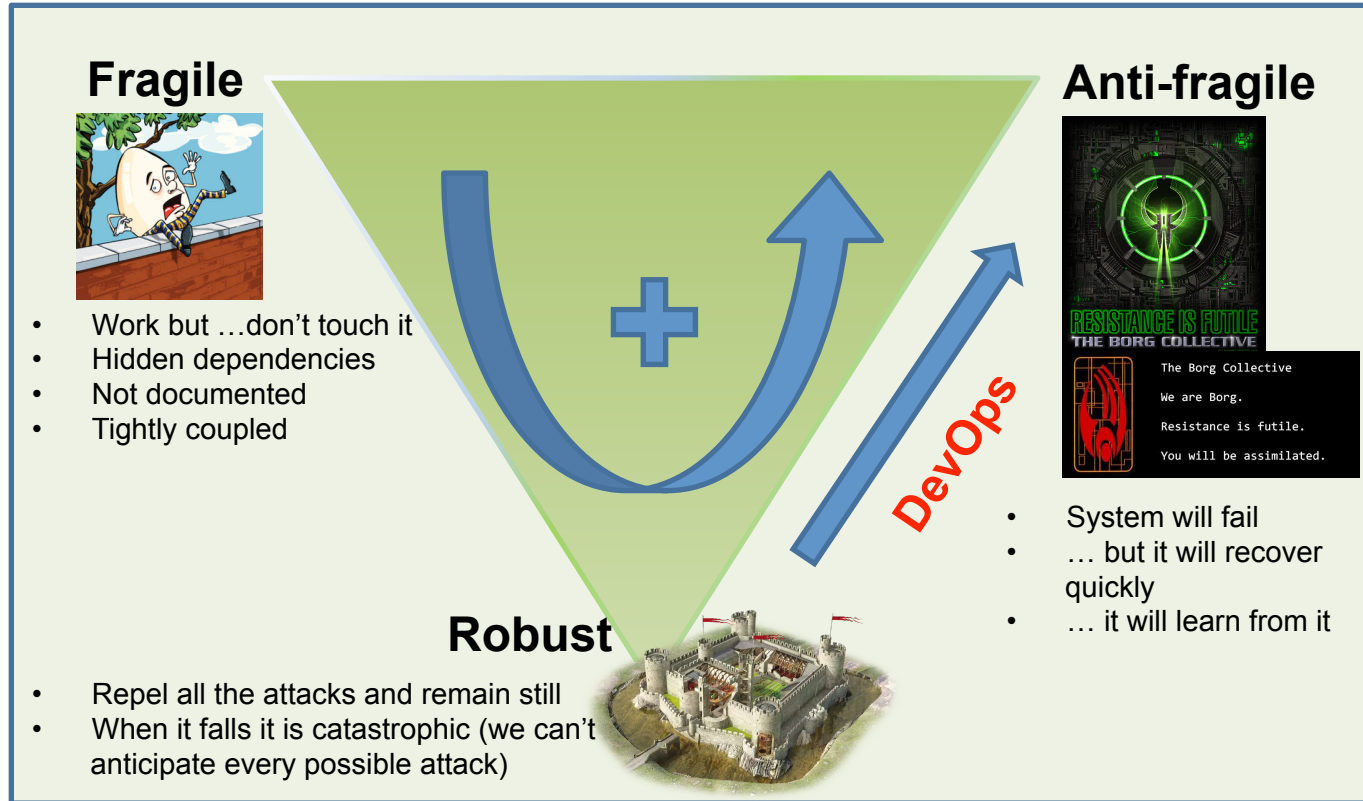
When General McChrystal took command of the Joint Special Task Force he quickly realised that conventional military tactics were failing. Al Qaeda in Iraq was a decentralised network that could move quickly, strike ruthlessly, then seemingly vanished in the local population. The allied forces had a huge advantage in numbers, equipment and training – but none of that seemed to matter. To defeat them, his Task Force would have to **acquire the enemy's speed and flexibility**.

- **And not for the intensive processing imposed by new astronomical missions either!**

- **We had applied strictly best practices, but:**

- They were insufficient to solve problems while keeping up with the performance requirements
- Too many unknown variables, scenarios
- Too little time to follow the processes as defined (but still needed to conform with quality standards)



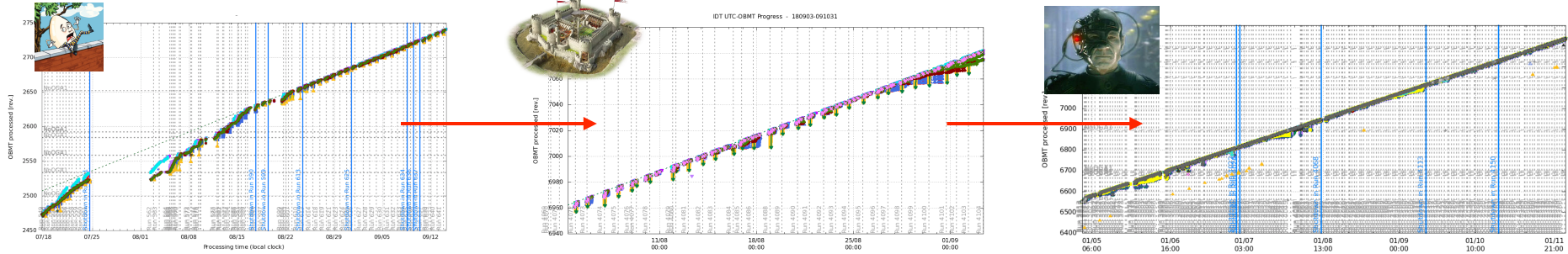


Transformation through DevOps

- **Embracing DevOps practices allows us to:**

- Velocity: fast time to production
- Continuous feedback and fast incorporation
- Stability and quality of operations: optimized and streamlined processes and procedures
- Reduced risks
- Empowered teams
- Saving time and money

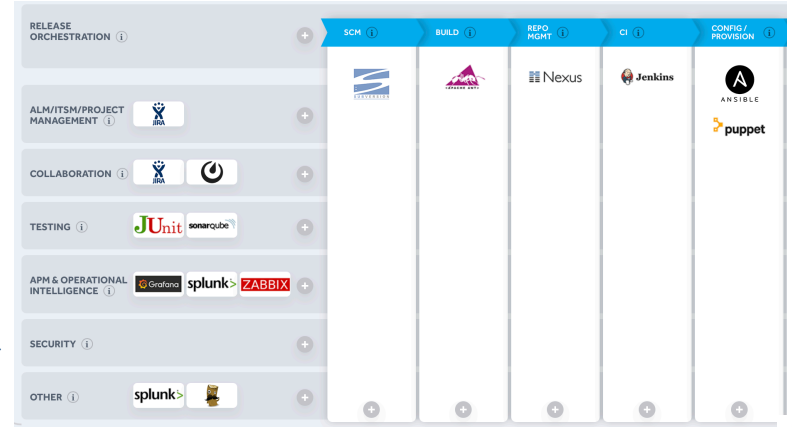
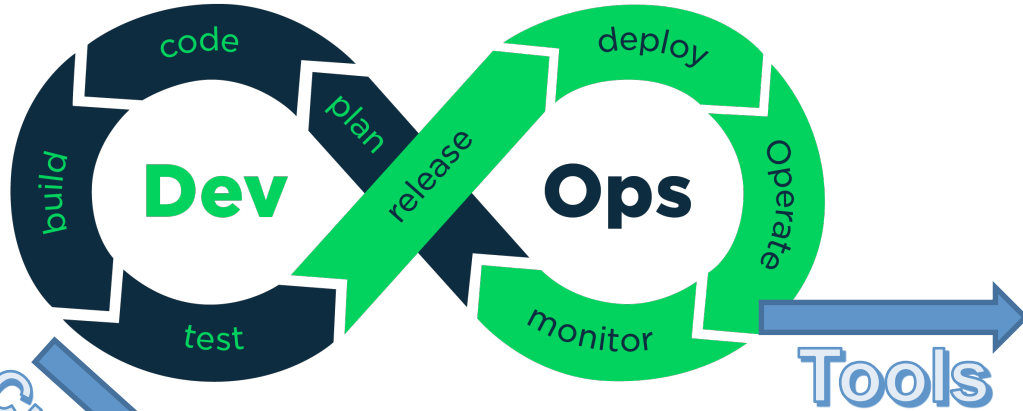
➔ **Antifragile system**



- More stable systems over longer periods
- Downtimes reduced
- Ability to detect and recover from failures faster

- Same with less
- Quality of the data produced
- Team motivated

DevOps is not just about tools



ermmm ... we don't use Dockers or Kubernetes (yet) 🤔

Cultural Change

- Automation
- Metrics
- Lean
- Collaboration

• Jenkins as the main Continuous Integration tool

- Recently introduced the usage of Jenkins Pipelines for implementation and integration of complex jobs: building, testing, QA and deployment: a big step towards automation
- SonarQube is the source analysis tool for continuous inspection of code quality. Fully integrated with Jenkins

Jenkins Pipeline **GT_20.4_21.3**

Full project name: CU1_GT/GT_20.4_21.3
GalaTools pipeline for automatically upload two different artifacts to Nexus, one for the 20.3 and the other for 21.2

Stage View

Checkout GT	Apply patch for 20.3 and build ...	Revert and build for 21.2 and build ...
3min 31s	1h 0min	1h 24min
2min 16s	1h 22min	1h 21min
4min 5s	1h 28min	1h 28min
3min 37s	1h 19min	1h 20min failed

Build History

#	Date	Status
#115	05-Nov-2018 13:38	Success
#116	04-Nov-2018 17:08	Success
#113	31-Oct-2018 16:38	Success
#112	31-Oct-2018 15:38	Success
#109	31-Oct-2018 11:43	Success
#108	31-Oct-2018 11:37	Success
#107	31-Oct-2018 11:26	Success
#106	31-Oct-2018 11:08	Success

Test Result Trend

Stage Time

Average stage times:
(Average full run time: ~3h 35min)

- #115: Nov 05 13:38, No Changes
- #113: Nov 04 17:08, No Changes
- #107: Oct 31 16:38, No Changes



Test Result Trend

PROJECTS Coverage

Project	Coverage
MDBDataConverter	90.4%
IDU	82.1%
MDBIntegrator	79.3%
MDBDataCustodian	78.7%
CalibrationTools	76.5%
FL	74.9%
MDBEventManager	73.9%
GalaTools	64.0%
MdbDict	64.0%
StatisticalTools	61.2%
IDTools	60.7%
CUB-validation-WP948	58.4%
DCS	56.4%
IDT	54.8%
MOBUtilities	51.1%
CalTeamGen	50.7%
MdbDataMgr	50.4%
CUB-validation-WP945	47.6%
ValidationTools	43.7%
CUB-validation-WP944	41.9%
CalTeamConverter	40.8%
MIT	40.6%
CUB-validation-WP942	37.1%
POS	35.6%
AGISTools	32.1%
CalTeamTools	27.5%
AGISLab	27.0%
CUB-validation-WP941	25.9%
AGIS	22.2%
Remat	21.3%
GOST	19.4%
DKP	18.0%
DTSTool	11.0%
CUB-validation-WP943	6.4%
CUB-validation-WP946	5.6%
CUB-validation-WP947	4.9%
CUBDataMapping	4.4%

PROJECTS

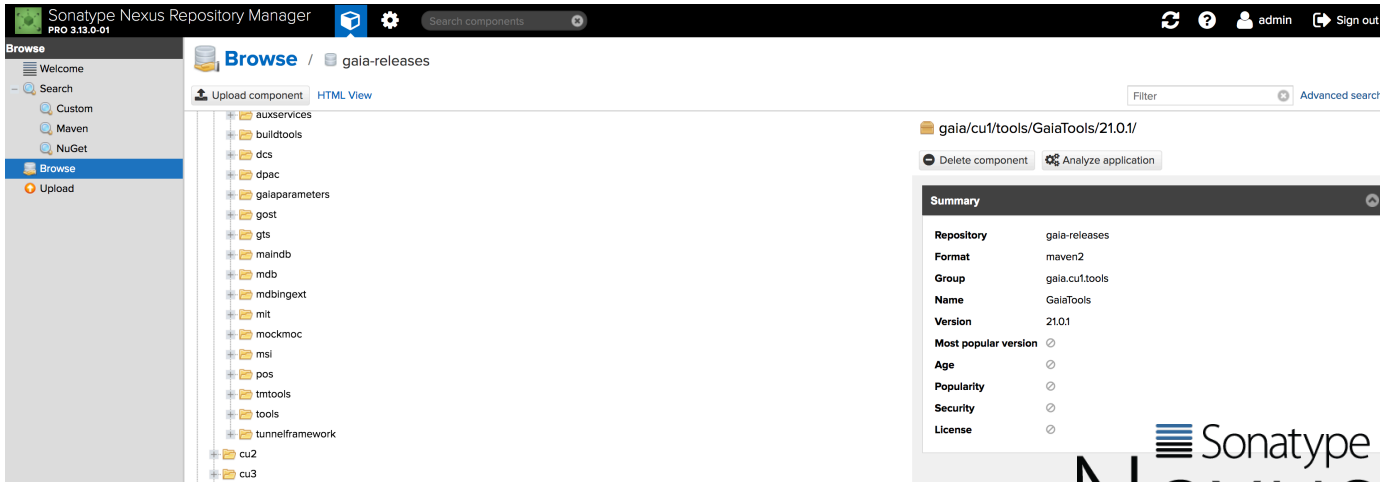
Size: Lines of Code Color: Coverage

Only the first 20 components are displayed

Artifact Repository and Infrastructure



- For storage of SW artifacts Nexus Professional Repository Manager
- IT Infrastructure automates the deployments, OS updates, etc. with Puppet and Ansible



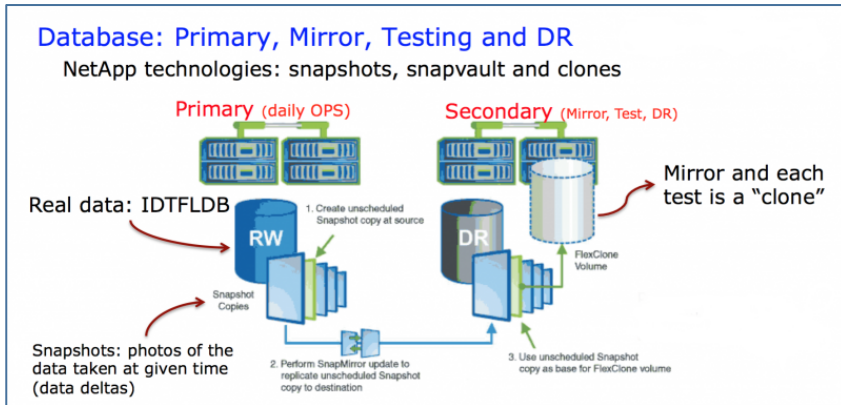
ANSIBLE



Sonatype
Nexus



- **Constant testing and re-evaluation**
- Main challenge: **reproduce DB** state, volume and load
- **Representative tests in "real time"**
 - Past: very costly – time and resources – to conduct one, let alone repeated test
 - Now we can do small, simple, fast experiments, trials
 - Clone of the working database (~30TB), same environment and receiving real time TM
 - -> Direct comparison with OPS
 - + backup!



- **Future plans:**
 - Testing management in Jira
 - Cucumber for automating system tests

Agility - Kanban

- **Kanban to visualize the work**
- Based on:
 - Making the policies and procedures explicit
 - Work in progress (WIP) limited
 - Manage the workflows
- Whiteboard with sticky notes (moving to **digital board in Jira** – useful for collecting metrics)
- **“Person of Interest”** role trains the team members and enhance the procedures



PERSON OF INTEREST

Procedures – Jupyter Notebooks



- **Jupyter Notebooks** for procedures and runbooks
 - Markdown text, executable code and output all inside a single document -> they all fit the procedures and runbooks need very well
- **Connection to operational logbooks**
- **Benefits:**
 - Up-to-date
 - Often reviewed, collaboration
 - Automation, reduce risks for manual intervention

The screenshot shows a Jupyter Notebook with two main sections. The first section, titled "Daily pipeline checks", contains code for generating a report. The second section, titled "Run POS Scheduler for POR generation", contains code for running a scheduler. The code includes comments in Danish and Python code using libraries like os, sys, and subprocess.

Message ID: 250	Entry time: Tue Nov 6 08:02:23 2018
System:	IDT
	Accountability
	JN
Subject:	Daily OPS Checks (JN)
Status:	Pass
	test.txt
Attachment 1:	dailyOPSChecks.pdf 455 kB Hide Hide all

dailyOPSChecksToday

November 6, 2018

1 Daily pipeline checks

```
In [1]: # #
# Procedure: dailyOPSChecks
# System: OPS - DAILY
# Description: Daily checks via OPERATEDDOP
# Location: OPER_PU/ops/dailyOPSChecks.pysh
# Dependencies: E:\directory\checkThreads.sh, checkDtpLegend.py, launchScript.sh, threads.sh
# Status: In progress
# Author: JN
# -
```

```
In [0]: $lsblk -e "Bogdir"
opsdir=1
# Gets the number of threads per Java process
Bogdir/launchScript.sh 2>/dev/null

node pid utb process
-----
gslal12 14262 20 HIT_ADI
gslal12 14308 12 HIT_OP
gslal12 14424 20 HIT_PTI
gslal12 14510 14 HIT_TMAOR_BTI
gslal12 14596 14 HIT_TMAOR_ADI
gslal12 14692 16 HIT_DBO
gslal12 14772 14 HIT_TMAOR_ORBIT
gslal12 14872 14 HIT_TMAOR_LFMS
gslal12 15160 19 DCS_SDR_FDR
gslal12 15336 26 DCS_OOBD
gslal12 15421 18 DCS_WORBL_1
gslal12 15504 21 DCS_WORBL_2
gslal12 15587 15 DCS_WORBL_3
gslal12 15670 19 DCS_WORBL_4
gslal12 15753 16 DCS_WORBL_5
-----
```



Takeaways



- **By embracing some DevOps practices the Gaia SOC enhanced the data processing and the effectiveness of the team**
- **Fundamental drivers: automation, lean, metrics and collaboration**
- **Continuous learning (it's a culture)**
- **Goal: collecting feedback from the systems and the team so that they can be used for improving them and achieving antifragile systems**





gaia



esa

Thanks QA

