Normalisation of Experiment Critical Services Data

Maria Girone, IT-ES

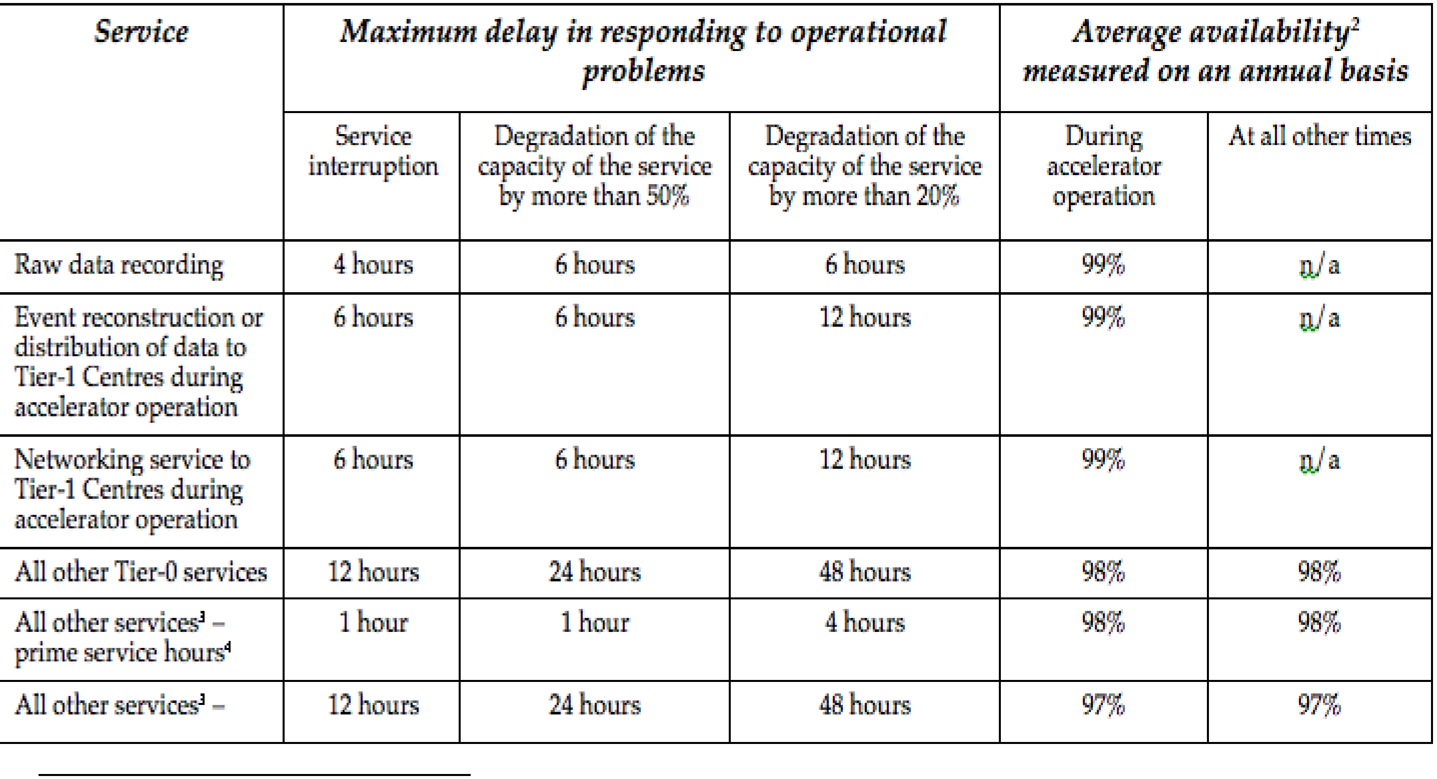
WLCG MB, 20 March 2012

Many thanks to Maite Barroso, Maria Dimou, Alessandro Di Girolamo, Ian Fisk, Stephane Jezequel, Maarten Litmaath, Stefan Roiser, Andrea Sciabà

Introduction

- Currently there are four **different definitions** of criticality, downtime and response time from the LHC experiments (see MB 10 Jan 2012)
  - CMS and LHCb used https://twiki.cern.ch/twiki/bin/view/FIOgroup/SDBUserDoc#Criticality
- Hard for service providers; can lead to sub-optimal response to incidents
- The purpose of this exercise is to propose common definitions for service incidents
  - following ITIL
  - WLCG MoU, Annex 3

MoU Annex 3



Service	*Maximum delay in responding to operational problems*			*Average availability[2] measured on an annual basis*	
	Service interruption	Degradation of the capacity of the service by more than 50%	Degradation of the capacity of the service by more than 20%	During accelerator operation	At all other times
Raw data recording	4 hours	6 hours	6 hours	99%	n/a
Event reconstruction or distribution of data to Tier-1 Centres during accelerator operation	6 hours	6 hours	12 hours	99%	n/a
Networking service to Tier-1 Centres during accelerator operation	6 hours	6 hours	12 hours	99%	n/a
All other Tier-0 services	12 hours	24 hours	48 hours	98%	98%
All other services[3] – prime service hours[4]	1 hour	1 hour	4 hours	98%	98%
All other services[3] –	12 hours	24 hours	48 hours	97%	97%

Assuming 200d LHC operations

Annual Downtime

14h

14h

14h

28h

28h

42h

- Written before operations began
- Response time referred to the maximum delay before action is taken
- Mean time to repair covered indirectly through the availability targets

Impact and Urgency

- For each WLCG service, each experiment defines:
  - The **Impact** on **operations** and **people** of a complete service failure
    - ⇒ the amount of "damage" done if no action is taken
  - The **time** before the full impact is reached
    - ⇒ how "urgent" it is to fix the service to prevent such damage from happening
  - We will call it **"Urgency"**

Example: Px → Computer Centre network cut has a very high impact but low urgency as the experiments have buffers

ITIL terminology

- Matches with ITIL Terminology
  - Impact - The effect on business that an incident has
  - Urgency - The extent to which the incident's resolution can bear delay
  - Priority - How quickly the service desk should address the incident (this is a combination of the other 2)

Services

- **“Functional” service**
  - A **high level** service corresponding to a particular **function** of the computing system
    - Example: data export from Tier-0 to Tier-1’s
    - **Defined in the WLCG MoU, Annex 3**
  - directly part of LHC computing operations
  - also included tools, desktop services and services for application development
- **“Specific” service**
  - A service contributing to one or more functional services
    - Example: FTS

CERN Functional Services

Operations related services
High bandwidth connectivity from detector area to computer centre
Recording and permanent storage in a MSS of raw and reconstructed data
Disk storage of reconstructed data
Distribution of raw and reconstructed data to Tier-1 sites in time with data acquisition
Prompt reconstruction, calibration and alignment
Storage and distribution of conditions data
Data analysis facility
Databases
VO management services

Tools and support services
Tools and services for application development (CVS, SVN, etc.)
Desktop services (email, web, Twiki, Indico, Vidyo, etc.)

Impact on operations and people

Level	Definition
10	Most ops services stop
9	Some ops services stop
8	One ops service stops
7	Most ops services disrupted
6	Some ops services disrupted
5	One ops service disrupted
4	Some “support” services stop
3	One “support” service stops
2	Some “support” services disrupted
1	One “support” service disrupted

Level	Definition
10	Whole VO affected
8	users affected > 50%
5	10% < users affected ≤ 50%
3	users affected ≤ 10%
1	A single user affected

Scale used for Impact

Urgency

- **Time** after the incident when the “full” impact is reached
  - Typically correlated to the experiment buffers, i.e. short service interruptions are normally not a problem

- Not to be confused with “response time”

Scale used for Urgency

Level	Time (hours)
10	0
9	0.5
8	1
7	2
6	4
5	6
4	12
3	24
2	48
1	72

Two Terms

- Introducing two metrics of Impact and Urgency helps evaluate how to treat services
  - Well designed systems have buffers and redundancy
    - A service may have a high impact if it fails, but that impact may be postponed for long periods (cf ex1)
    - Urgency helps with planning operational response
    - Impact helps with system design
  - Separating the concepts also helps in the experiment evaluation and doesn’t mix up how soon we have to fix something with the importance of the service

CERN Specific Services

Px → Computer Centre network
WLCG network (LHCOPN, GPN)
CERN Oracle online
CERN Oracle Tier-0 (including streaming)
Frontier front-end and Squid
CASTOR tape
CASTOR disk
EOS
Batch service
CE
LFC
FTS
VOM(R)S
BDII

Myproxy
gLite WMS
CVMFS Stratum0
CVMFS Stratum1
Dashboard
SAM
VOBOXes
AFS
CAF
CVS/SVN
Twiki
Mail and Web services
Hypernews
Indico
Savannah/JIRA/TRAC

Experiment input for CERN specific services

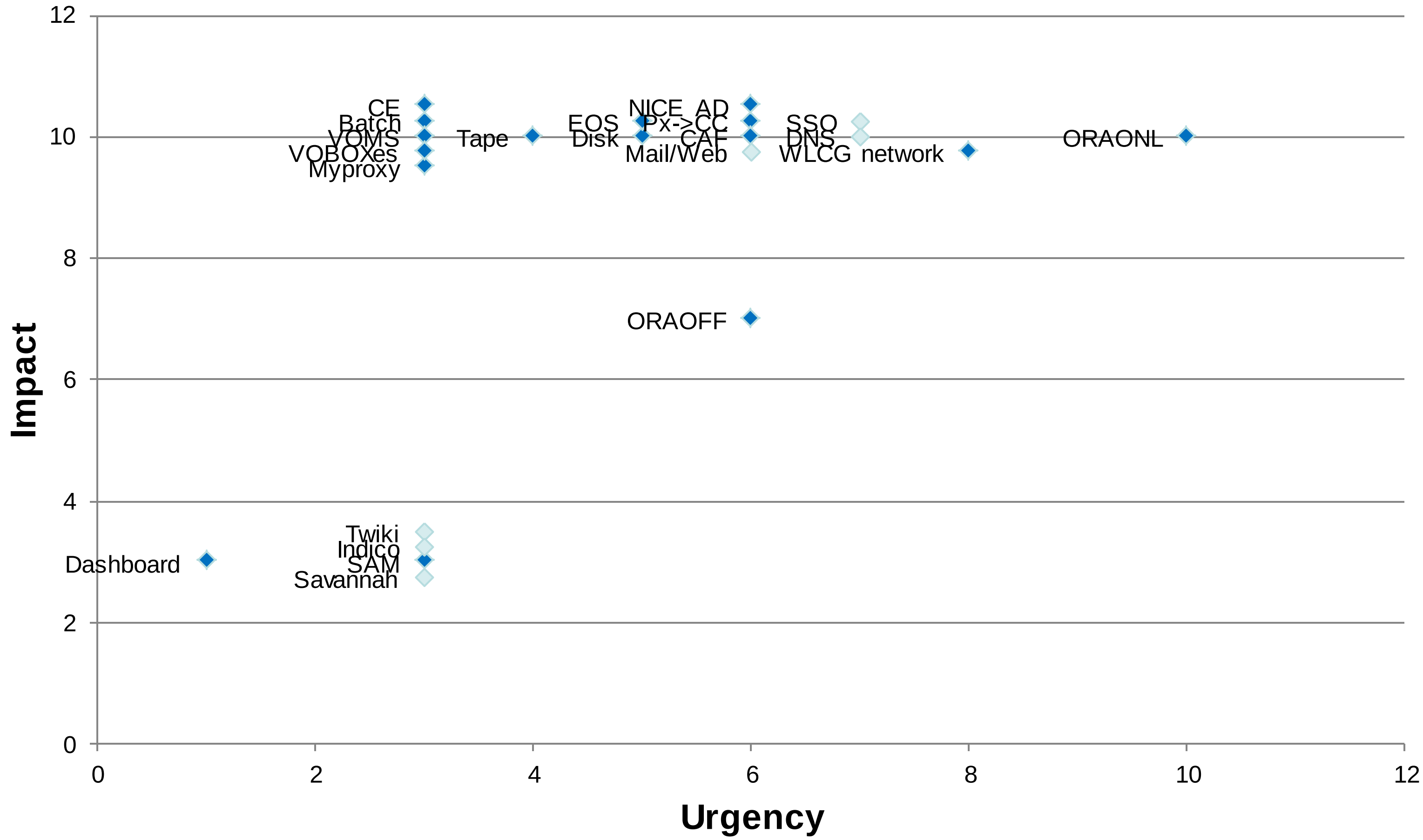
ALICE

Service	Urgency	Impact
Px → Computer Centre network	6	10
WLCG network (LHCOPN, GPN)	8	10
CERN Oracle online	10	10
CERN Oracle Tier-0 (including streaming)	6	7
Frontier front-end and Squid	-	-
CASTOR tape	4	10
CASTOR disk	5	10
EOS	5	10
Batch service	3	10
CE	3	10
LFC	-	-
FTS	-	-
VOM(R)S	3	10
BDII	-	-

Service	Urgency	Impact
Myproxy	3	10
gLite WMS	-	-
CVMFS Stratum0	-	-
CVMFS Stratum1	-	-
Dashboard	1	3
SAM	3	3
VOBOXes	3	10
AFS	-	-
CAF	6	10
CVS/SVN	-	-
Twiki	3	3
Mail and Web services	6	10
Hypernews	-	-
Indico	3	3
Savannah/JIRA/TRAC	3	3

Service	Urgency	Impact
SSO	7	10
DNS	7	10
NICE AD servers	6	10

ALICE distribution

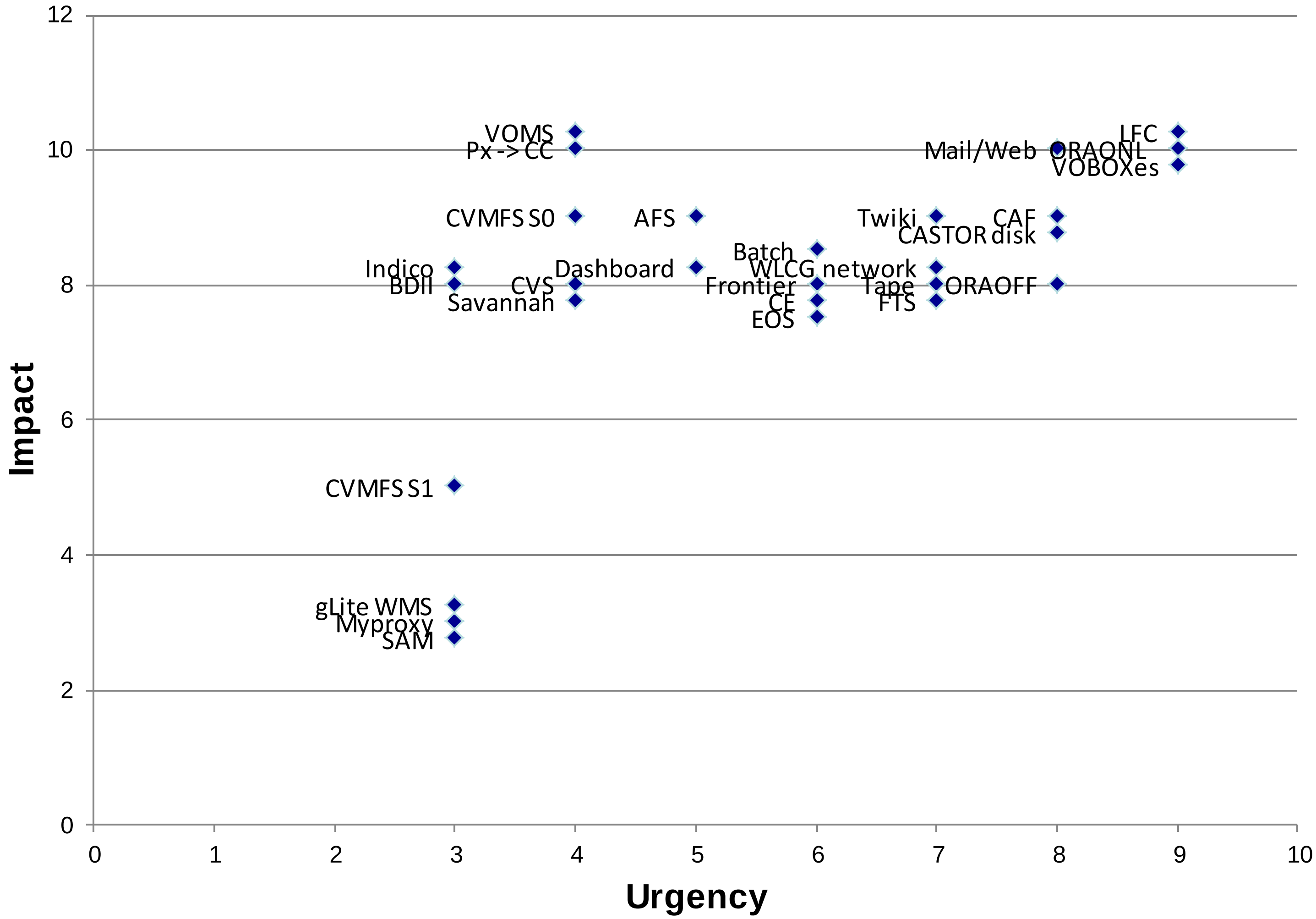


ATLAS (Draft)

Service	Urgency	Impact
Px → Computer Centre network	4	10
WLCG network (LHCOPN, GPN)	7	8
CERN Oracle online	9	10
CERN Oracle Tier-0 (including streaming)	8	8
Frontier front-end and Squid	6	8
CASTOR tape	7	8
CASTOR disk	8	9
EOS	6	8
Batch service	6	8
CE	6	8
LFC	9	10
FTS	7	8
VOM(R)S	4	10
BDII	3	8

Service	Urgency	Impact
Myproxy	3	3
gLite WMS	3	3
CVMFS Stratum0	4	9
CVMFS Stratum1	3	5
Dashboard	5	8
SAM	3	3
VOBOXes	9	10
AFS	5	9
CAF	8	9
CVS/SVN	4	8
Twiki	7	9
Mail and Web services	8	10
Hypernews	na	na
Indico	3	8
Savannah/JIRA/TRAC	4	8

ATLAS distribution (draft)

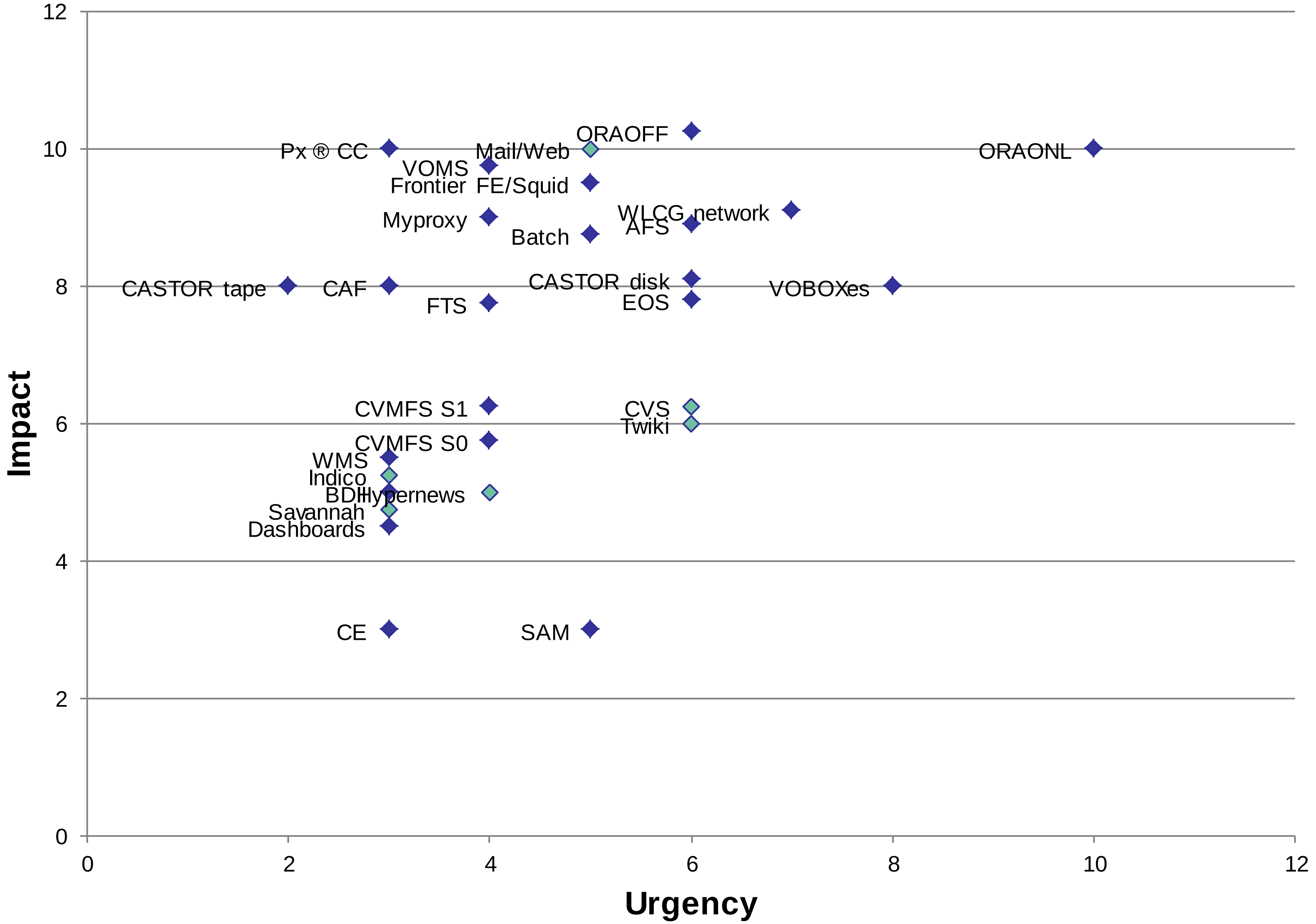


CMS

Service	Criticality	Impact
Px → Computer Centre network	3	10
WLCG network (LHCOPN, GPN)	7	9
CERN Oracle online	10	10
CERN Oracle Tier-0 (including streaming)	6	10
Frontier front-end and Squid	6	10
CASTOR tape	2	8
CASTOR disk	6	8
EOS	6	8
Batch service	5	9
CE	3	3
LFC	NA	NA
FTS	4	8
VOM(R)S	4	10
BDII	3	5

Service	Urgency	Impact
Myproxy	4	9
gLite WMS	3	5
CVMFS Stratum0	4	6
CVMFS Stratum1	4	6
Dashboards	3	5
SAM	5	3
VOBOXes	8	8
AFS	6	9
CAF	3	8
CVS/SVN	6	6
Twiki	6	6
Mail and Web services	5	10
Hypernews	4	5
Indico	3	5
Savannah/JIRA/TRAC/eLog	3	5

CMS distribution

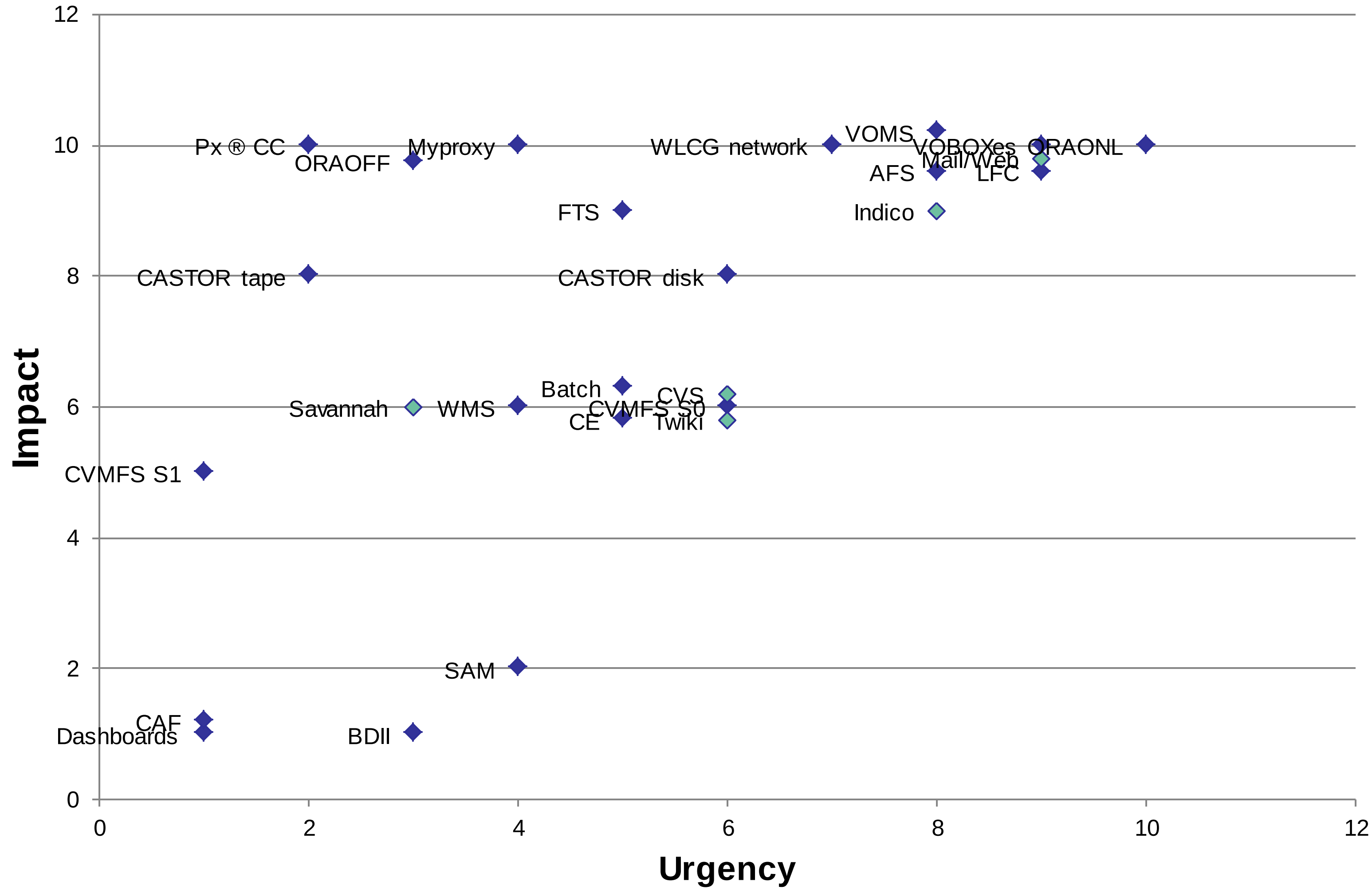


LHCb

Service	Criticality	Impact
Px → Computer Centre network	2	10
WLCG network (LHCOPN, GPN)	7	10
CERN Oracle online	10	10
CERN Oracle Tier-0 (including streaming)	3	10
Frontier front-end and Squid	NA	NA
CASTOR tape	2	8
CASTOR disk	6	8
EOS	NA	NA
Batch service	5	6
CE	5	6
LFC	9	10
FTS	5	9
VOM(R)S	8	10
BDII	3	1

Service	Criticality	Impact
Myproxy	4	10
gLite WMS	4	6
CVMFS Stratum0	6	6
CVMFS Stratum1	1	5
Dashboard	1	1
SAM	4	2
VOBOXes	9	10
AFS	8	10
CAF	1	1
CVS/SVN	6	6
Twiki	6	6
Mail and Web services	9	10
Hypernews	NA	NA
Indico	8	9
Savannah/JIRA/TRAC	3	6

LHCb distribution



CERN Alarms per VO and Area

September 2011 – March 2012

Type of Problem	ATLAS	CMS	ALICE	LHCb	Total
FileTransfer	0	3	0	0	3
FileAccess	4	1	0	0	5
Databases	2	2	0	0	4
Storage	0	1	0	0	1
Network	0	0	0	0	0
LocalBatch	3	2	0	0	5
Middleware	0	0	1	0	1
Other	1	0	0	1	2
Total/VO	10	9	1	1	21

Next Steps

- Analyze discrepancies among the experiments in the impact and urgency assignment of individual services

- Based on this two-dimensional assessment, service priorities can be set
  - Further input from operations
    - Frequency of incidents (alarms)
  - Provide guidance on use of **alarms** (as opposed to tickets)

Backup Slides

WLCG Operational Targets

Time Interval	Critical Tier0 Services (see MoU)	Target
30’	Operator response to alarm / call to x5011	99%
1 hour	Operator response to alarm / call to x5011	100%
4 hours	Expert intervention in response to above	95%
8 hours	Problem resolved	90%
24 hours	Problem resolved	99%

Targets approved by WLCG Overview Board

99% of problems resolved in 24h

Time Interval	Tier1 Services	Target
1 working day	All services – problem resolved	95%

Time Interval	Tier2 Services	Target
1 working day	All services – problem resolved	90%

Targets discussed at WLCG Grid Deployment Board

Tier-1 functional services (from MoU)

Operations related services
Raw and reconstructed data import from Tier-0
Simulated and processed data import from other WLCG centres
MSS archival storage of raw, reconstructed, processed and simulated data
Disk storage for data and temporary files
Provision of data access to other WLCG centres
Data analysis and reprocessing
Other experiment services
Network and data transfer services to Tier-0 and Tier-1 sites (high bandwidth) and to Tier-2 sites
Databases

Tier-2 functional services (from MoU)

Operations related services
Disk storage for data and temporary files
Provision of data access to other WLCG centres
Data analysis
Simulation and data processing
Other experiment services
Network and data transfer services

ES

Tier-1/2 and WLCG Services

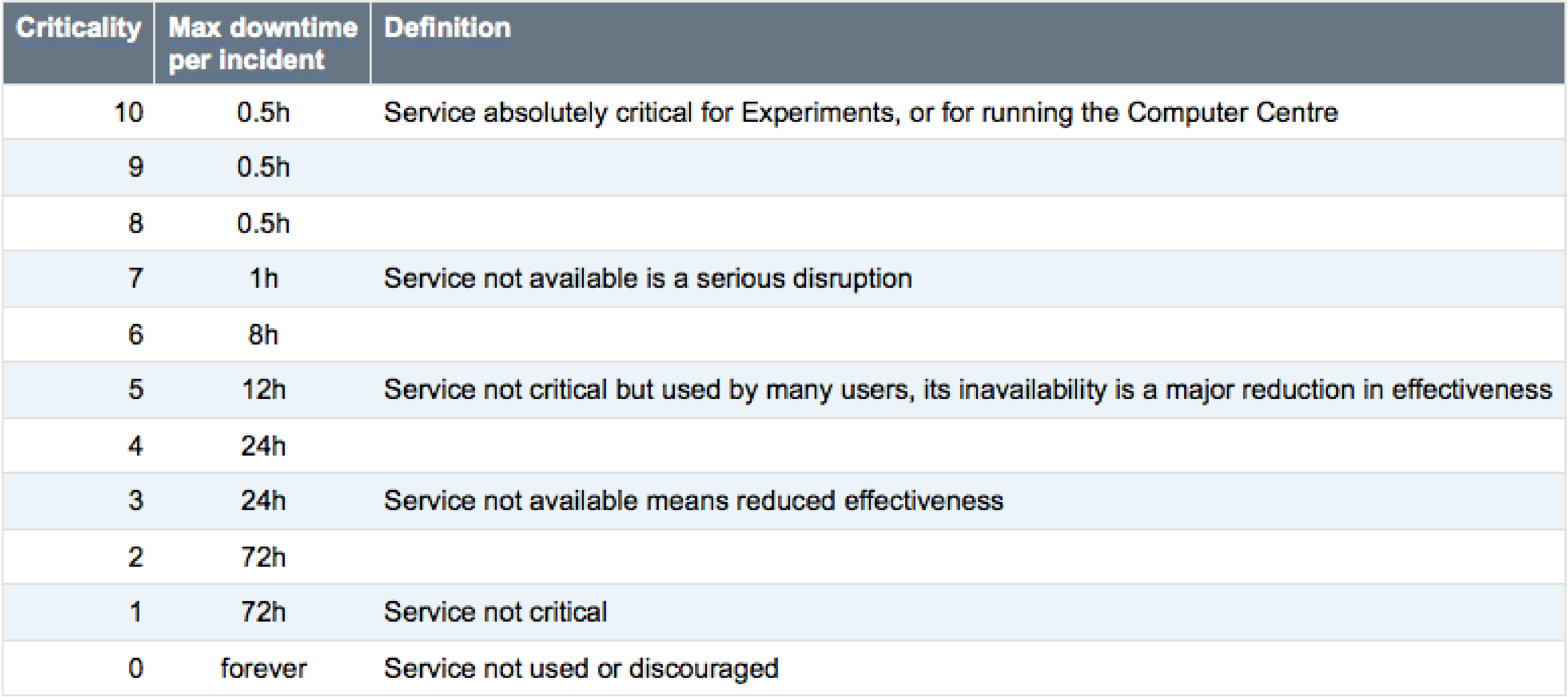
CERN IT Department

Tier-1
WLCG network (LHCOPN, GPN)
Frontier front-end and Squid
SE (includes SRM)
Batch service
CE
LFC
FTS
Oracle
VOBOXes

Tier-2
Squid
SE (includes SRM)
Batch service
CE
Oracle
VOBOXes

WLCG
GOCDB
GGUS
EGI Operations Portal

https://twiki.cern.ch/twiki/bin/view/FIOgroup/SDBUserDoc#Criticality



Criticality	Max downtime per incident	Definition
10	0.5h	Service absolutely critical for Experiments, or for running the Computer Centre
9	0.5h	
8	0.5h	
7	1h	Service not available is a serious disruption
6	8h	
5	12h	Service not critical but used by many users, its inavailability is a major reduction in effectiveness
4	24h	
3	24h	Service not available means reduced effectiveness
2	72h	
1	72h	Service not critical
0	forever	Service not used or discouraged