ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

WP2 – Task 2.2 update

Paul Millar
Regular WP2 meeting

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

QoS, an introduction

Lots of options when provisioning storage hardware

- Speed: NVMe, SSD/flash, enterprise/consumer SAS/SATA HDD, … “tape”
- Specialist: low endurance SSD, SMR, “cloud” drives, …
- Aggregation: RAID, RAIN/object store (replication, erasure codes)

Leads to a combination explosion

Describe how storage behaves, not how it is stored.

Why?

- Save money: use the cheapest storage that’s “good enough”.
- We want to support small amounts of specialist hardware

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

QoS, an introduction

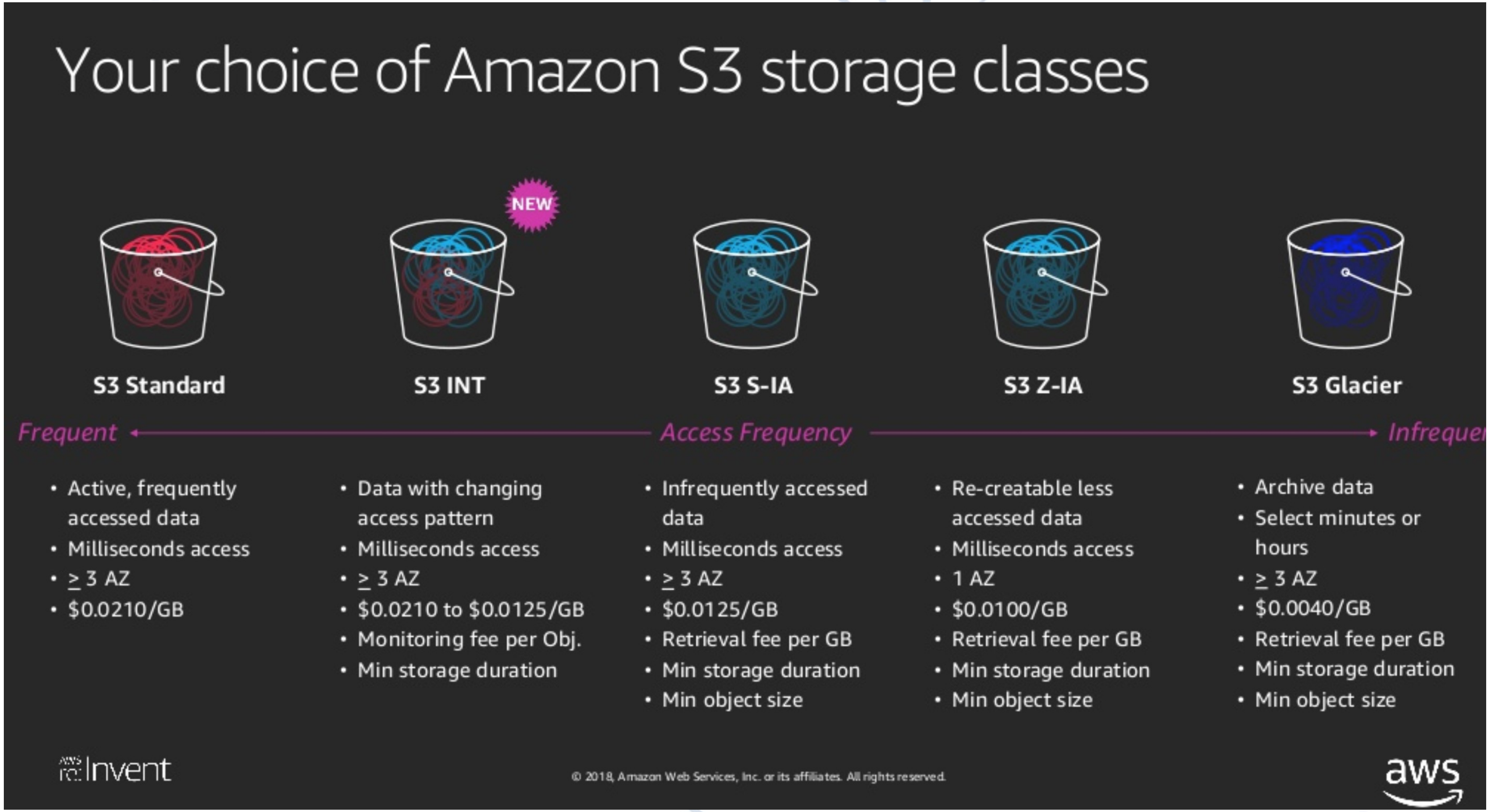
How to describe your storage?

- **Role/usage**: experiment-specific, use-case drive
- **Expected behaviour**: access time, bandwidth, durability, …
- **Cost**: money? Storage footprint? %-of-pledged resources?
- **Usage requirements**: optimal block size, minimum- or maximum lifetime, minimum/maximum file size

Similar to what is happening in the commercial world …

Funded by the European Union's
Horizon 2020 - Grant N° 824064

QoS, an introduction



Your choice of Amazon S3 storage classes

S3 Standard

- Active, frequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0210/GB

S3 INT

- Data with changing access pattern
- Milliseconds access
- ≥ 3 AZ
- $0.0210 to $0.0125/GB
- Monitoring fee per Obj.
- Min storage duration

S3 S-IA

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0125/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Z-IA

- Re-creatable less accessed data
- Milliseconds access
- 1 AZ
- $0.0100/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier

- Archive data
- Select minutes or hours
- ≥ 3 AZ
- $0.0040/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

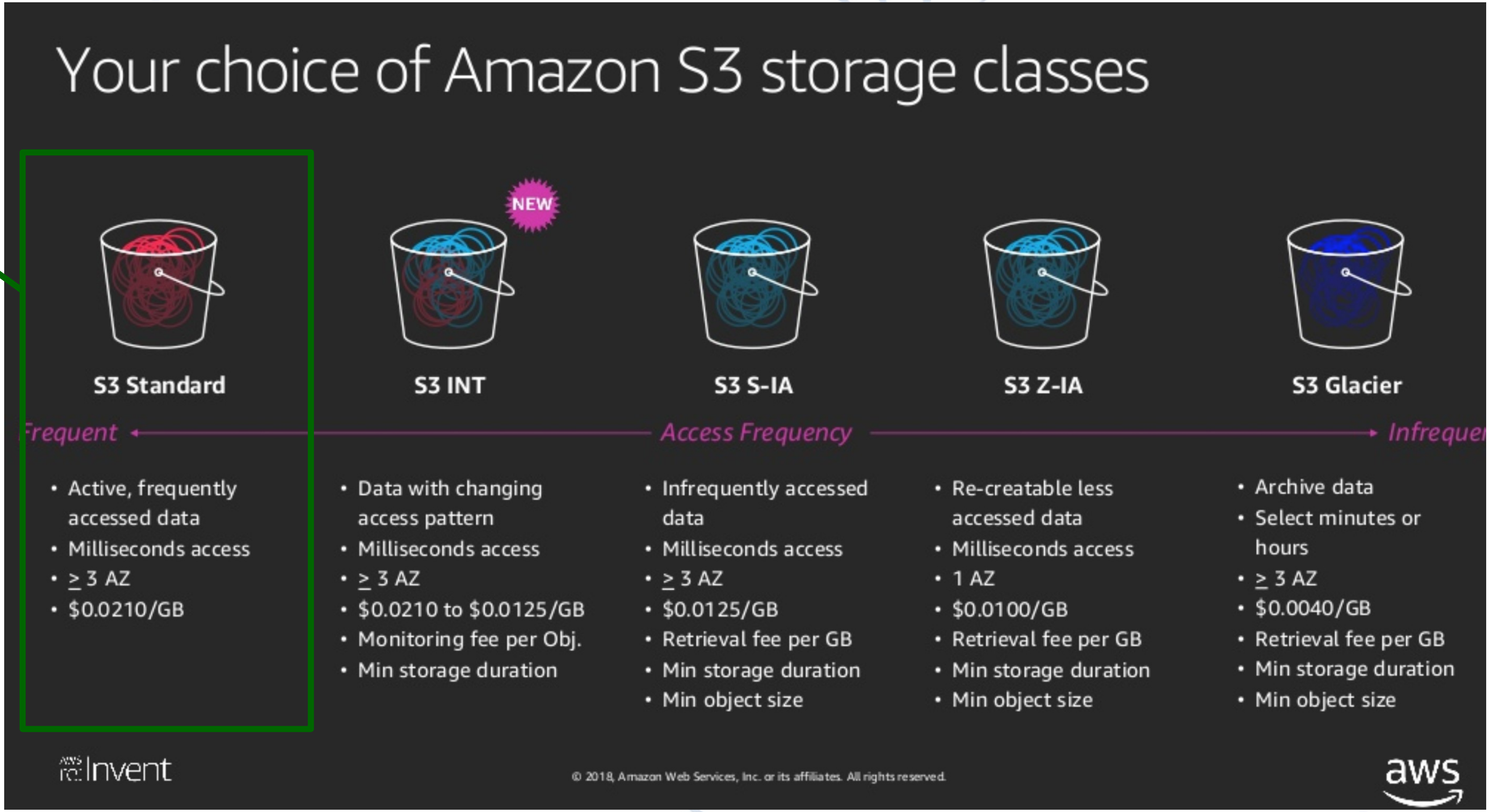
AWS re:Invent

aws

From https://medium.com/@devopslearning/aws-s3-intelligent-tiering-s3-int-a75b6e1763db

QoS, an introduction

QoS class



Your choice of Amazon S3 storage classes

Frequent ← Access Frequency → Infreque

S3 Standard

- Active, frequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0210/GB

S3 INT

- Data with changing access pattern
- Milliseconds access
- ≥ 3 AZ
- $0.0210 to $0.0125/GB
- Monitoring fee per Obj.
- Min storage duration

S3 S-IA

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0125/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

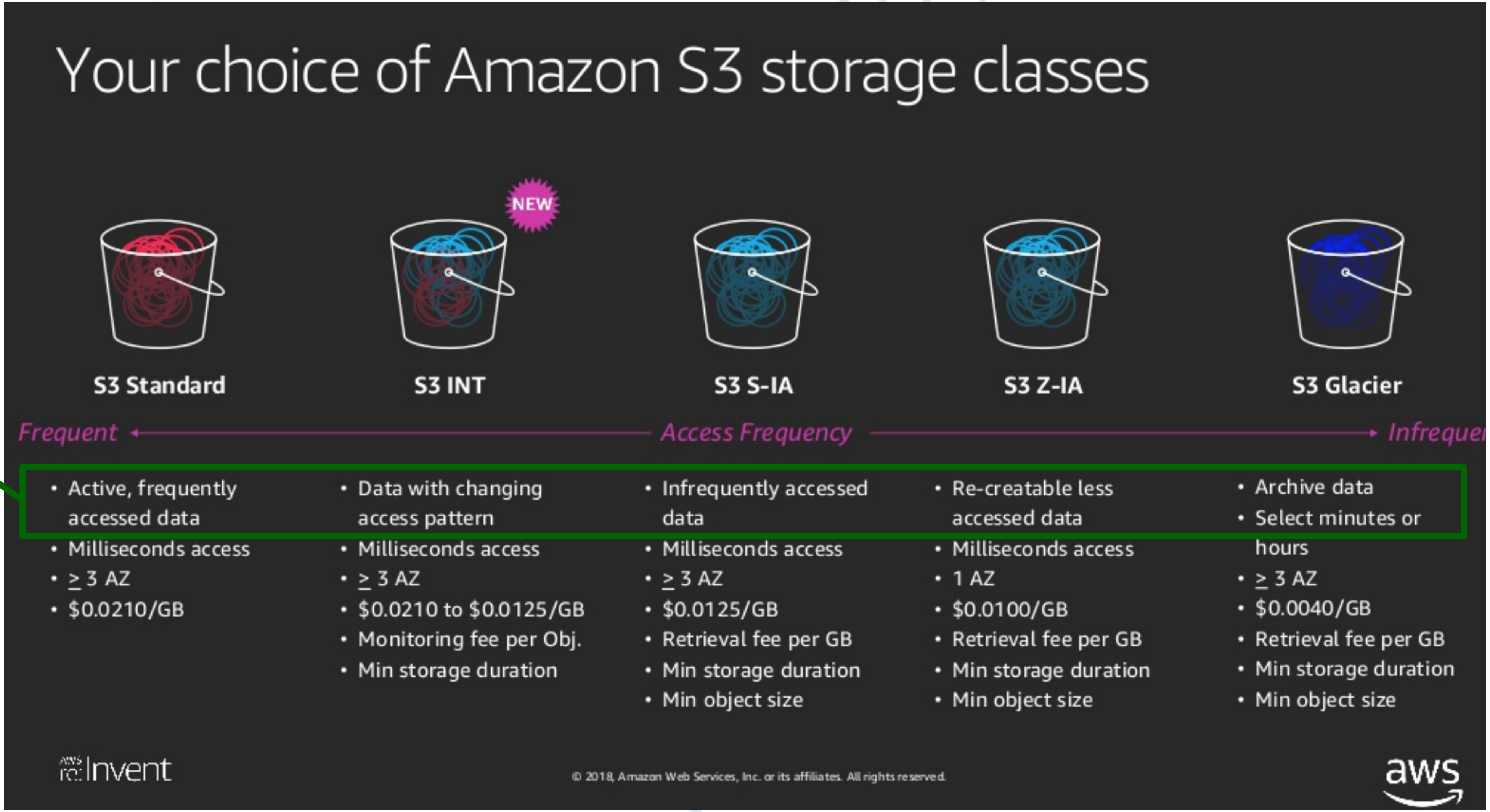
S3 Z-IA

- Re-creatable less accessed data
- Milliseconds access
- 1 AZ
- $0.0100/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier

- Archive data
- Select minutes or hours
- ≥ 3 AZ
- $0.0040/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

QoS, an introduction



Your choice of Amazon S3 storage classes

NEW

Frequent ← *Access Frequency* → *Infreque*

S3 Standard	S3 INT	S3 S-IA	S3 Z-IA	S3 Glacier
• Active, frequently accessed data	• Data with changing access pattern	• Infrequently accessed data	• Re-creatable less accessed data	• Archive data • Select minutes or hours
• Milliseconds access	• Milliseconds access	• Milliseconds access	• Milliseconds access	
• ≥ 3 AZ	• ≥ 3 AZ	• ≥ 3 AZ	• 1 AZ	• ≥ 3 AZ
• $0.0210/GB	• $0.0210 to $0.0125/GB	• $0.0125/GB	• $0.0100/GB	• $0.0040/GB
	• Monitoring fee per Obj.	• Retrieval fee per GB	• Retrieval fee per GB	• Retrieval fee per GB
	• Min storage duration	• Min storage duration	• Min storage duration	• Min storage duration
		• Min object size	• Min object size	• Min object size

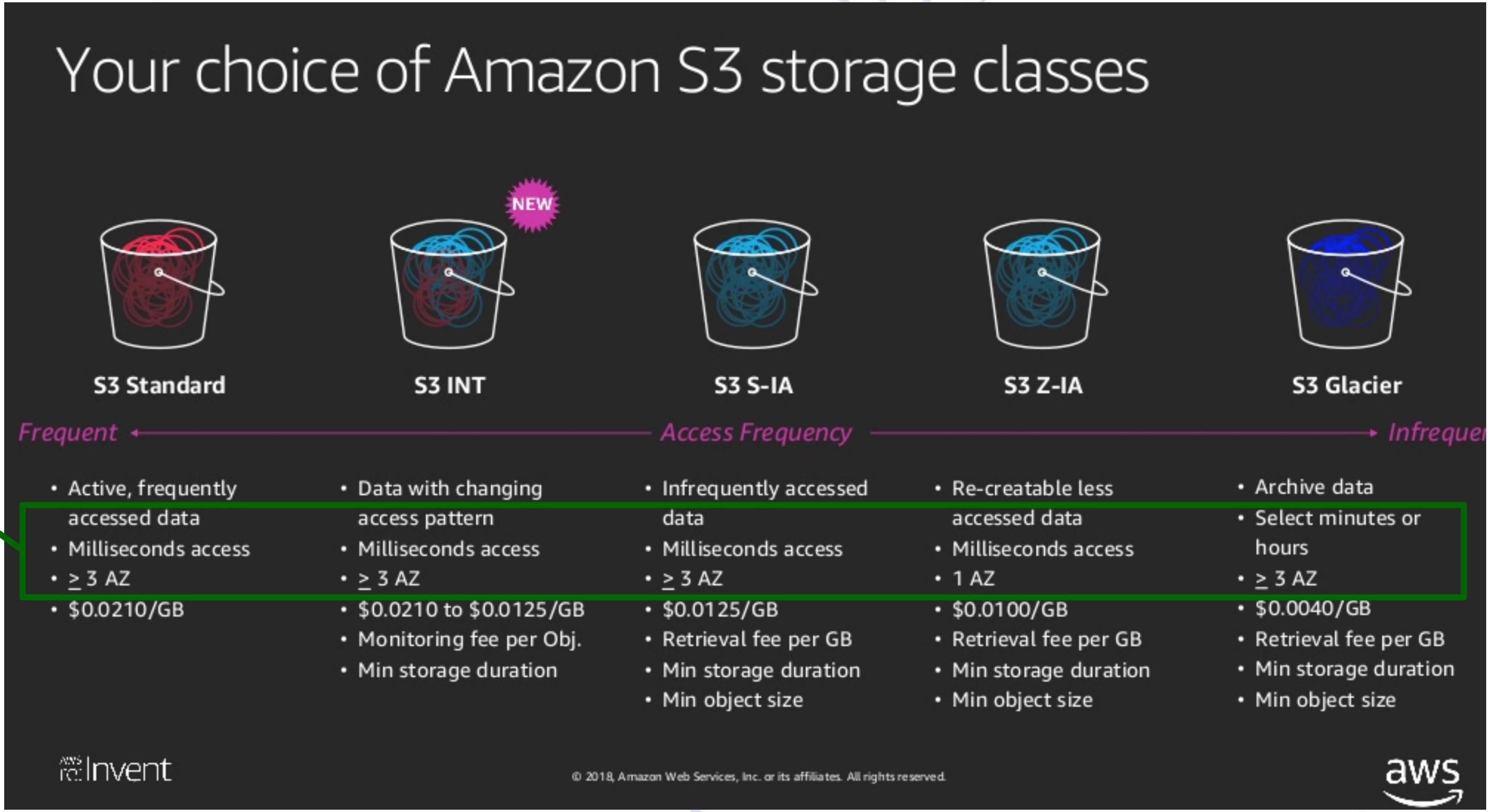
Role/usage

re:Invent

aws

Funded by the European Union's
Horizon 2020 - Grant N° 824064

QoS, an introduction



Your choice of Amazon S3 storage classes

Frequent ← *Access Frequency* → *Infreque*

S3 Standard

- Active, frequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0210/GB

S3 INT

- Data with changing access pattern
- Milliseconds access
- ≥ 3 AZ
- $0.0210 to $0.0125/GB
- Monitoring fee per Obj.
- Min storage duration

S3 S-IA

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0125/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Z-IA

- Re-creatable less accessed data
- Milliseconds access
- 1 AZ
- $0.0100/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier

- Archive data
- Select minutes or hours
- ≥ 3 AZ
- $0.0040/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

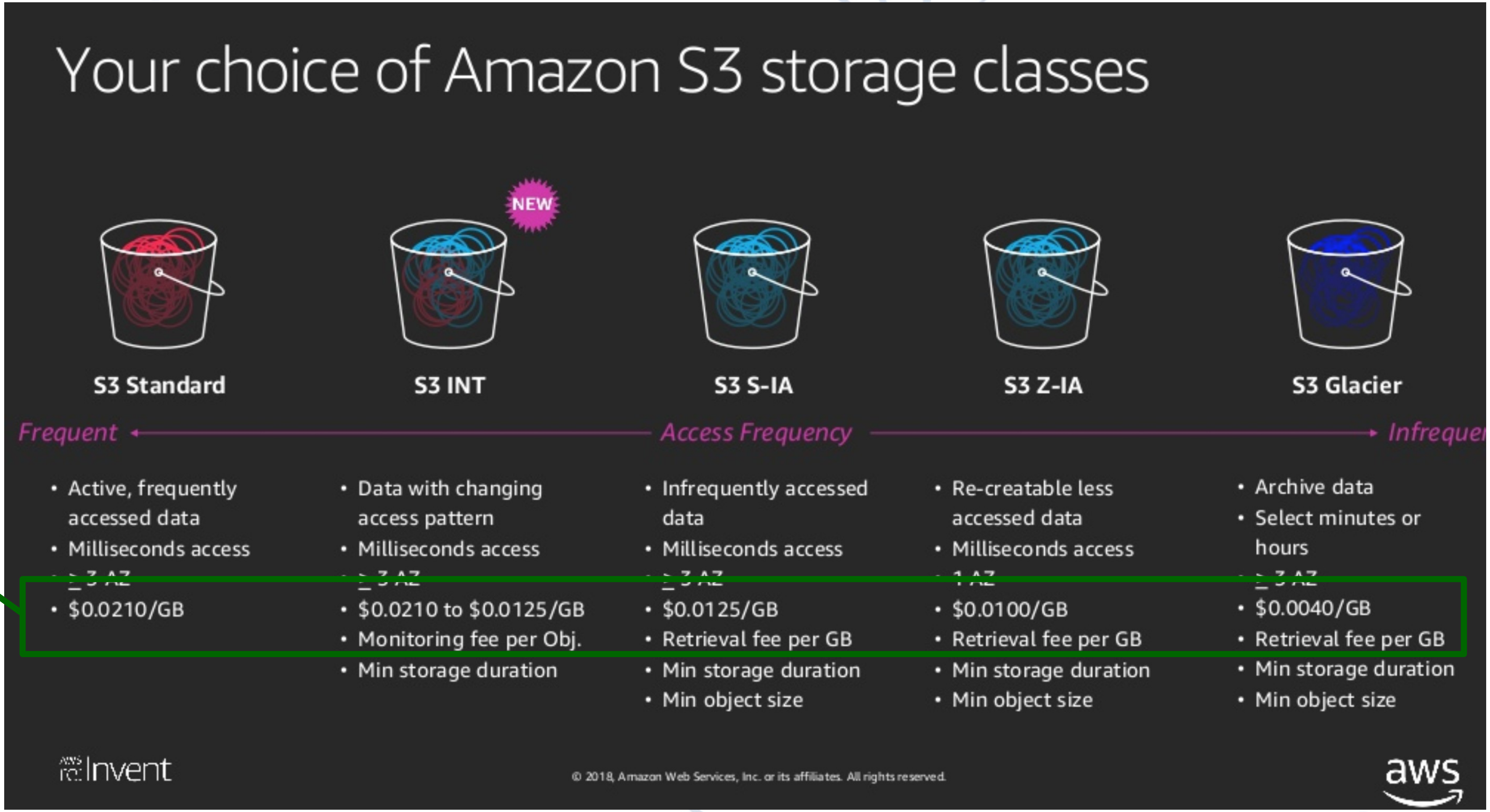
re:Invent

aws

Expected behaviour

Funded by the European Union's Horizon 2020 - Grant N° 824064

QoS, an introduction



Your choice of Amazon S3 storage classes

Frequent ← *Access Frequency* → *Infreque*

S3 Standard

- Active, frequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0210/GB

S3 INT

- Data with changing access pattern
- Milliseconds access
- ≥ 3 AZ
- $0.0210 to $0.0125/GB
- Monitoring fee per Obj.
- Min storage duration

S3 S-IA

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0125/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Z-IA

- Re-creatable less accessed data
- Milliseconds access
- 1 AZ
- $0.0100/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier

- Archive data
- Select minutes or hours
- ≥ 3 AZ
- $0.0040/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

re:Invent

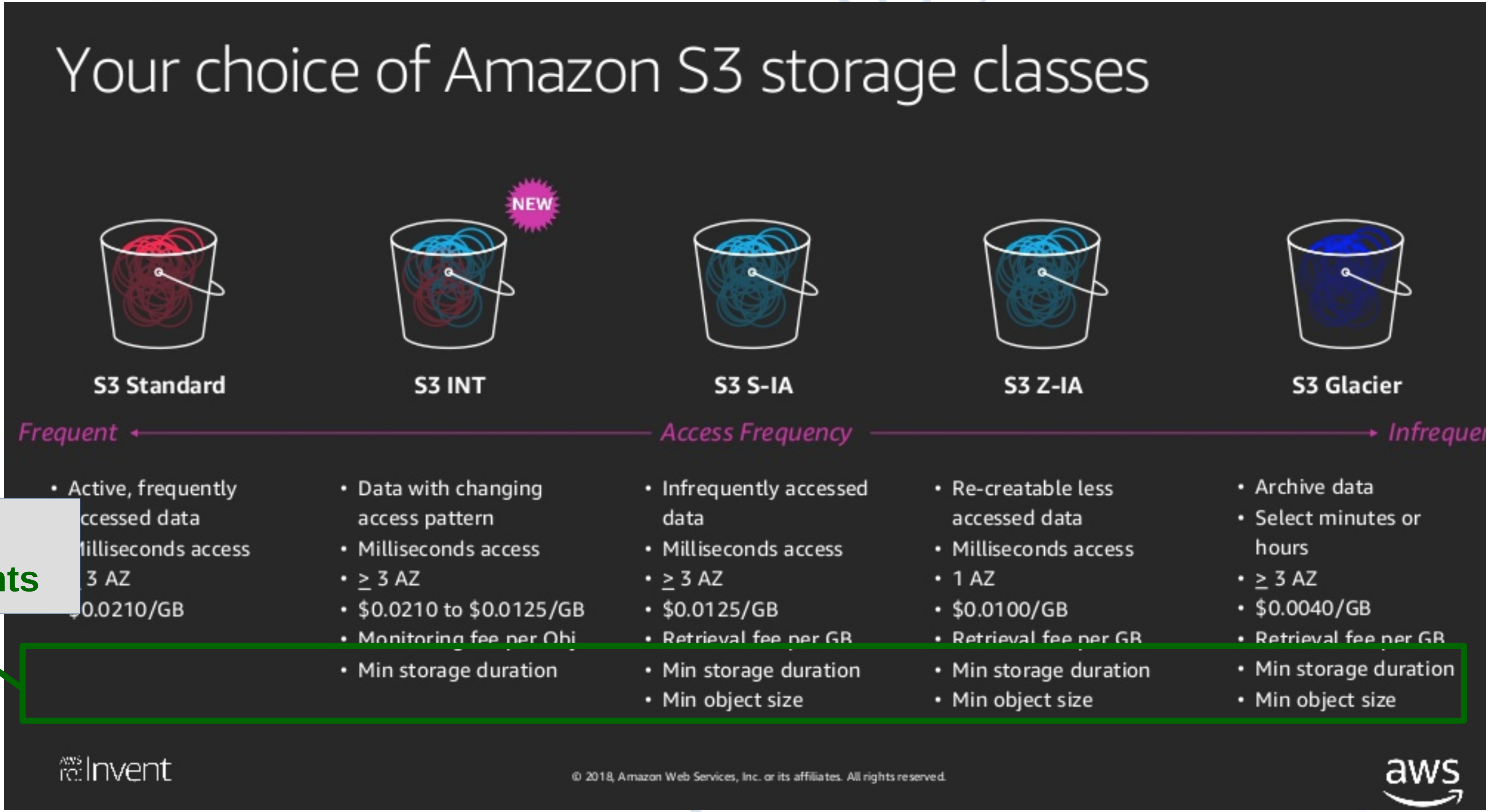
© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.

aws

Cost

Funded by the European Union's
Horizon 2020 - Grant N° 824064

QoS, an introduction



Your choice of Amazon S3 storage classes

Frequent ← *Access Frequency* → *Infreque*

S3 Standard

- Active, frequently
ccessed data
- illiseconds access
- 3 AZ
- $0.0210/GB

S3 INT

- Data with changing access pattern
- Milliseconds access
- ≥ 3 AZ
- $0.0210 to $0.0125/GB
- Monitoring fee per Obj
- Min storage duration

S3 S-IA

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0125/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Z-IA

- Re-creatable less accessed data
- Milliseconds access
- 1 AZ
- $0.0100/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier

- Archive data
- Select minutes or hours
- ≥ 3 AZ
- $0.0040/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

Usage requirements

Funded by the European Union's
Horizon 2020 - Grant N° 824064

The QoS, other projects

In WLCG

- **DOMA-QoS**: DOMA WG is developing support for the next-generation of storage technologies; QoS task is investigating QoS.
- **Data carousel**: ATLAS-led task investigating focusing on optimising tape usage.
- **MAS**: experimental caching layer in front of tape

Other projects

- **XDC**: extreme DataCloud is an EU-funded development project funding work on QoS (amongst other work).
- **ASTERICS/OBELICS**: …

Funded by the European Union's
Horizon 2020 - Grant N° 824064

DOMA-QoS site survey

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Sites – WLCG survey

DOMA-QoS and Ops-WG put together a survey of sites contributing to WLCG

- Eight QoS questions
- **~80 sites** contributed (of 165 sites)
- Answers: https://twiki.cern.ch/twiki/bin/view/LCG/QoSSurveyAnswers
- Detailed conclusions: https://twiki.cern.ch/twiki/bin/view/LCG/QoSSurveyConclusions

Information from Oliver Keeble’s GDB presentation

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

Sites – WLCG survey

Very brief summary:

- Underlying media: **mostly enterprise SAS/HDD**, but with some "consumer" drives. SSDs mostly for service use (journaling, caching).
- Media combination: **mostly 12–16 disk RAID-6**. Others are JBOD Ceph, EOS, HDFS and GPFS, all with redundancy.
- Access layer: dCache, DPM, EOS, StoRM, xrootd; also direct access (mounted fs). Most sites provide POSIX layer.
- Effort: difficult to draw conclusions. Storage is porous, with exact boundaries varying from site to site.

Information from Oliver Keeble's GDB presentation

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Sites – WLCG survey

- Non-WLCG communities: shared resources, but WLCG calling shots.
- Future directions:
  - **Ceph**:
    - RAID→ JBOD and expose via CephFS or RBD/librados
    - Provisioning of S3
    - POSIX, as a stage-in cache (ARC style)
    - Migration from Luster.
  - Some investigating **experimental media** (SMR, low-endurance SSD), but no clear direction.
  - Some densification experiments

Information from Oliver Keeble’s GDB presentation

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

ESCAPE Testbed

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Sites – ESCAPE testbed

INFN:

- Bologna: StoRM tape & GPFS-HDD, Ceph might be available in 2020
- Roma: DPM-Ceph-ErasureCoding (can play with settings)
- Napoli: distributed DPM: Napoli (POSIX), Roma (Ceph), Frascati (POSIX)

GSI: xrootd-HDD

IN2P3: “Nessie” dCache-HDD/RAID-6. Could add tape and maybe SSD (if strong enough reason)

LAPP: dCache-HDD/RAID-(6?).

RUG: iRODS as a middle layer, managing QoS. Quite sophisticated system

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

Sites – ESCAPE testbed

CERN: distributed EOS {CERN + ES, NL and AUS}.

DESY: dCache-HDD-RAID-6, distributed {Hamburg, Zeuthen, Moscow}

PIC: dCache

SurfSARA: dCache

AARNET: EOS?

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Sites – ESCAPE testbed

Summary:

A good collection of access software: dCache, DPM, EOS, iRODS, xrootd.

Some distributed deployments: dCache (DESY), DPM (Napoli), EOS (CERN), iRODS (RUG).

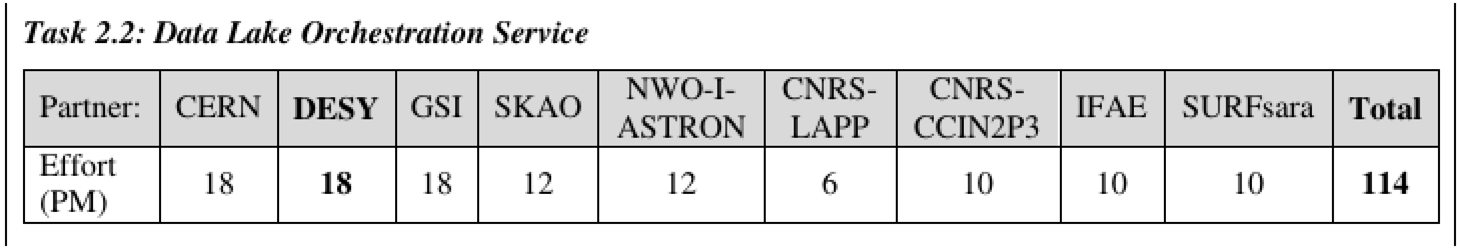
Media is mostly HDD (either RAID, Ceph), but may be able to use tape and possibly SSD.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

ESCAPE Experiments

Funded by the European Union's
Horizon 2020 - Grant N° 824064

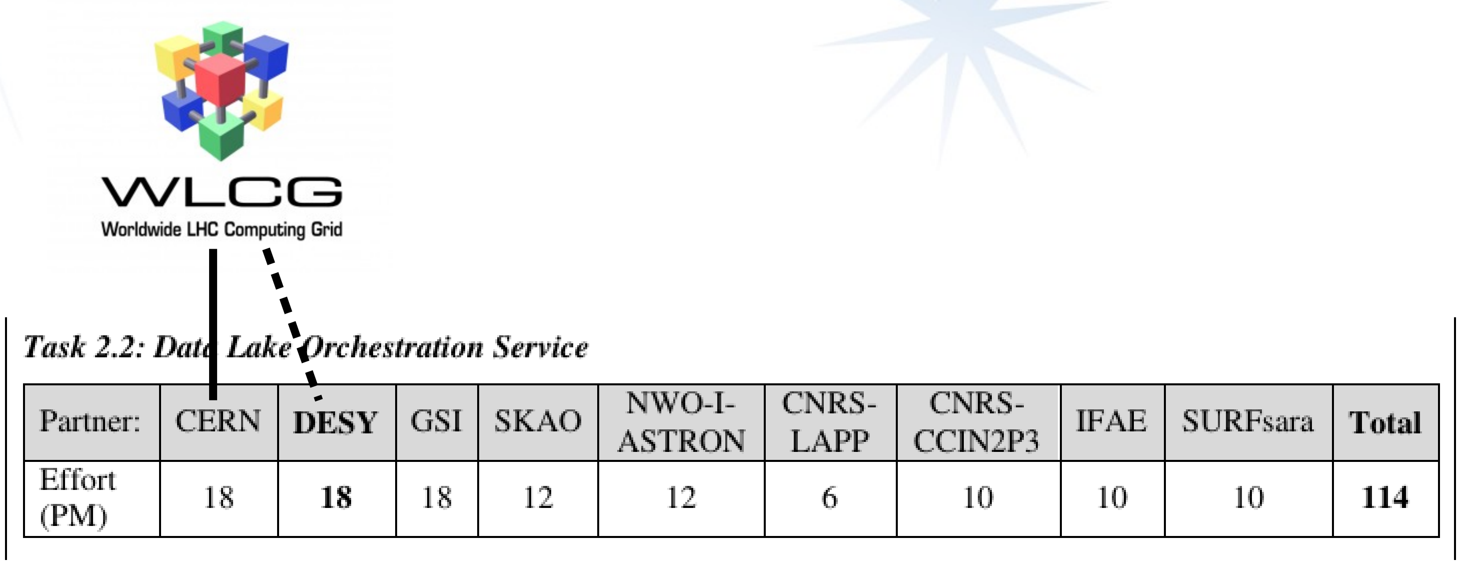
Links to the experiments



Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Links to the experiments

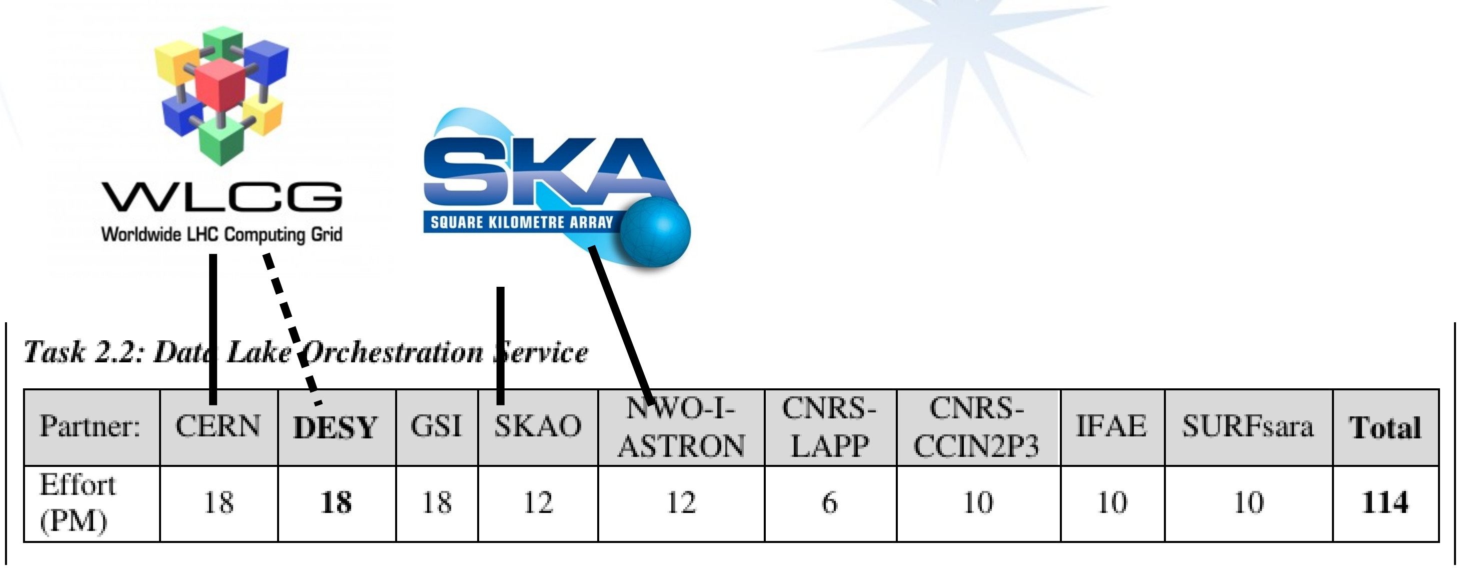

WLCG
Worldwide LHC Computing Grid

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Funded by the European Union's
Horizon 2020 - Grant N° 824064

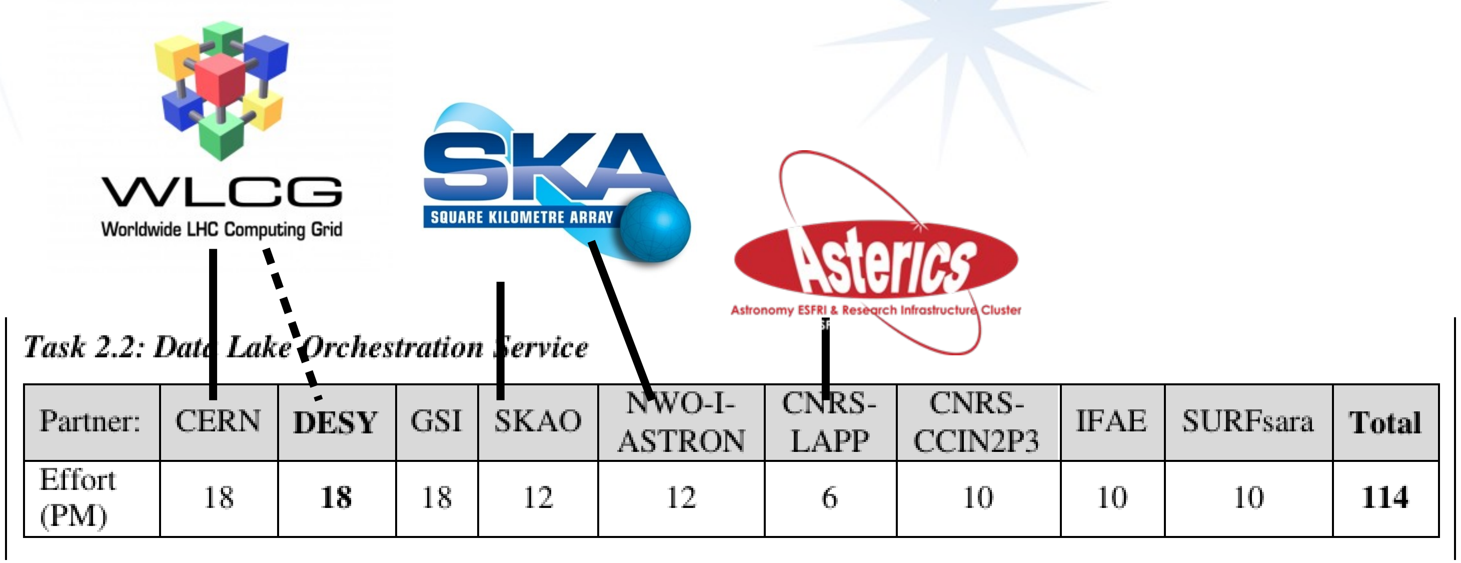
Links to the experiments



Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Links to the experiments

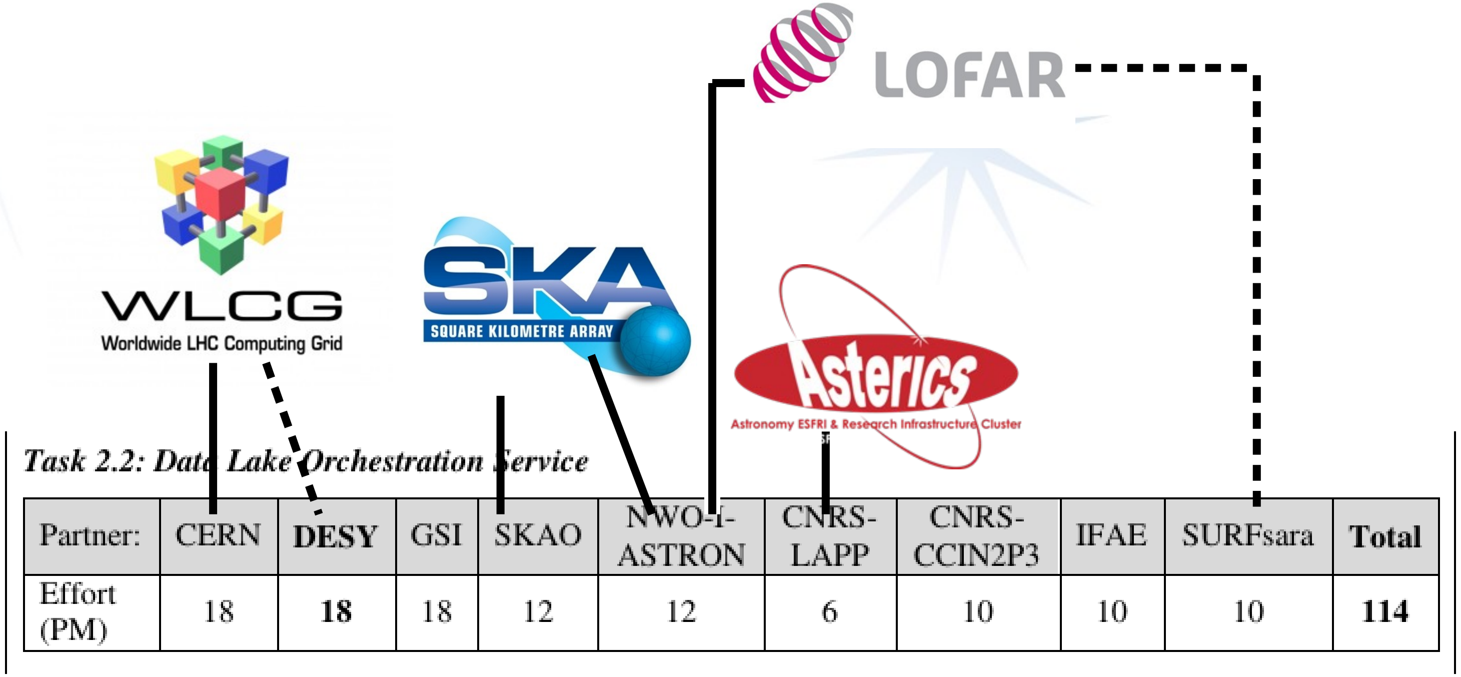

WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Links to the experiments



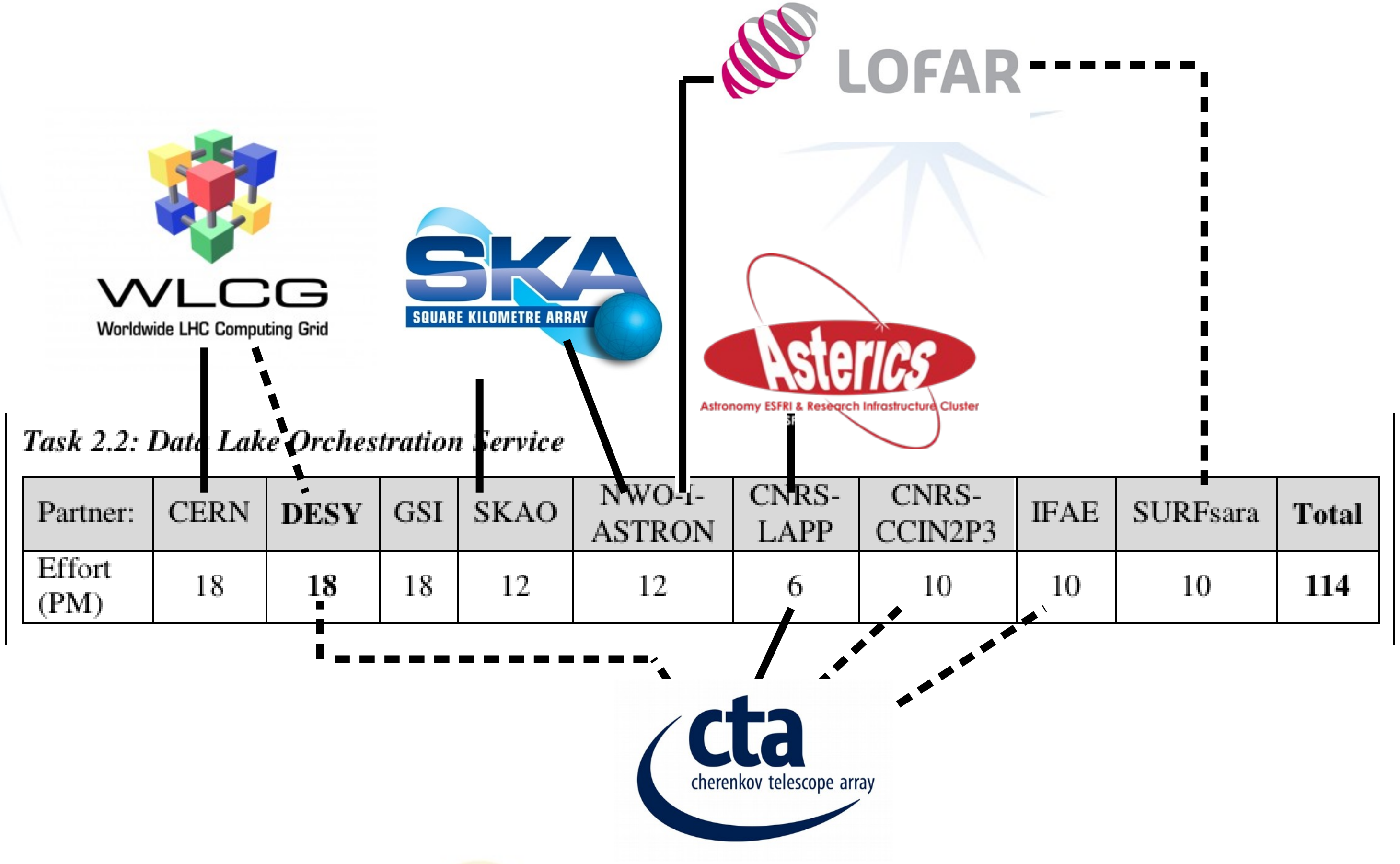
LOFAR
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Funded by the European Union's
Horizon 2020 - Grant N° 824064

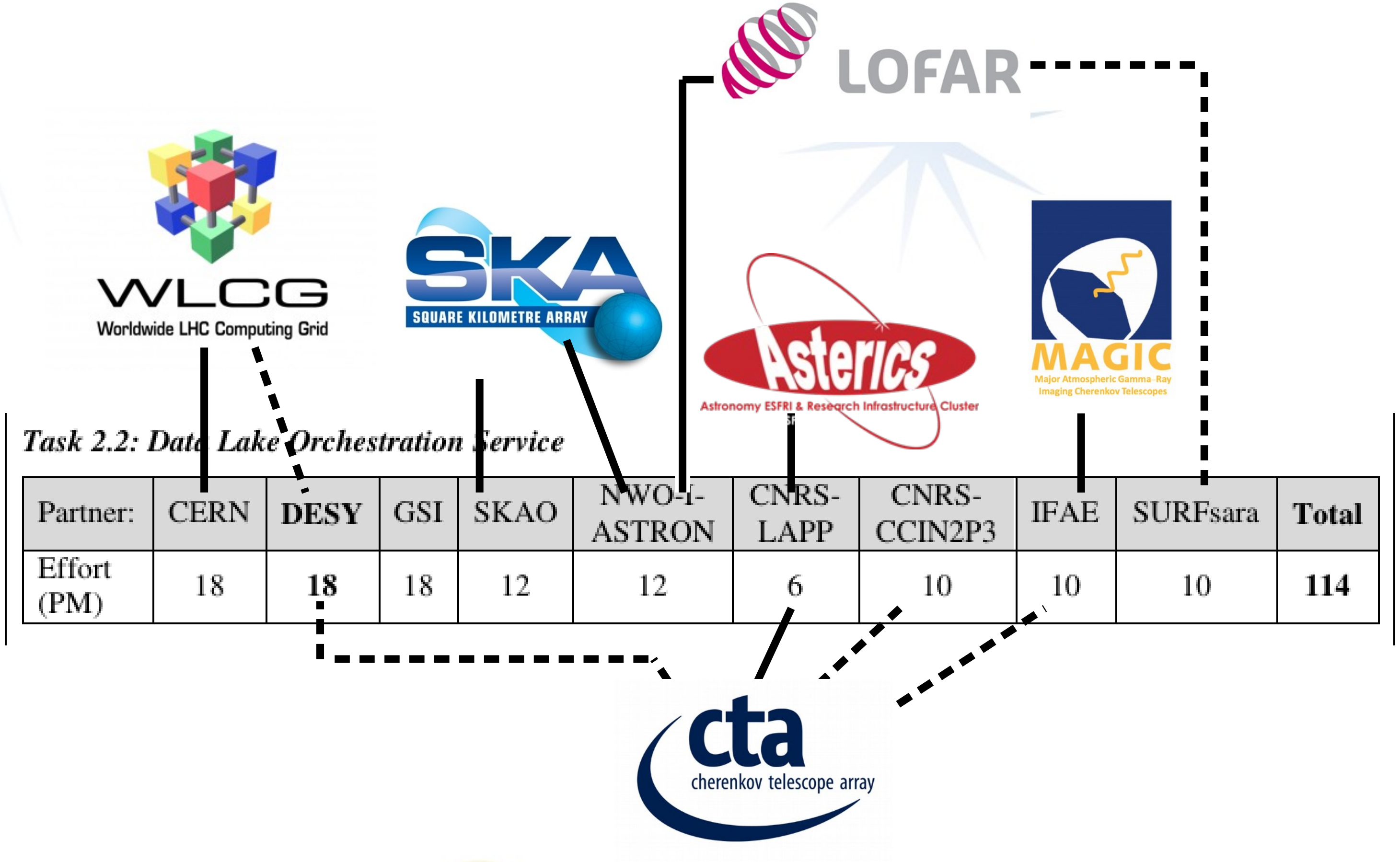
Links to the experiments


LOFAR
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster
cta
cherenkov telescope array

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Links to the experiments

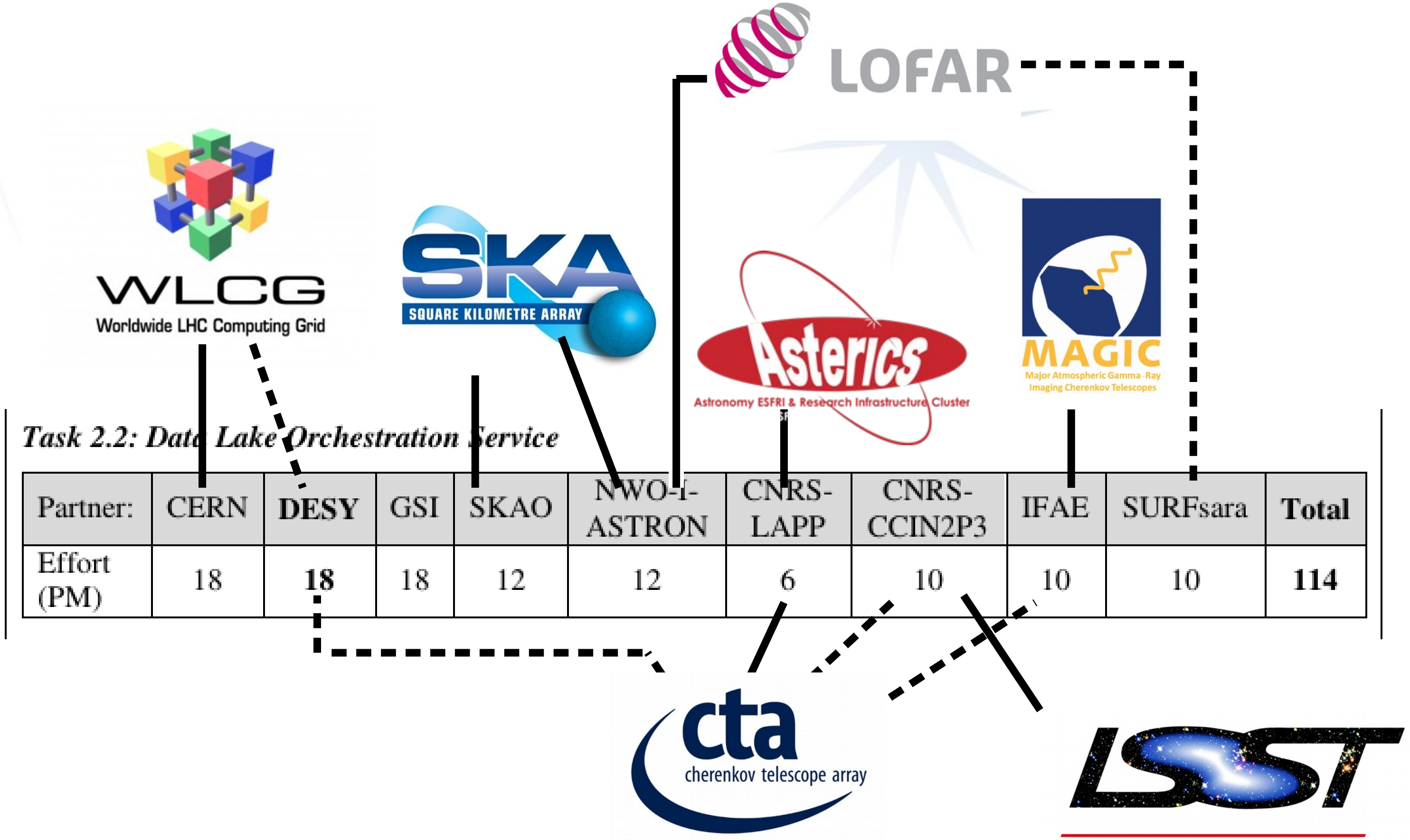

LOFAR
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster
MAGIC
Major Atmospheric Gamma-Ray Imaging Cherenkov Telescopes
cta
cherenkov telescope array

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Funded by the European Union's
Horizon 2020 - Grant N° 824064

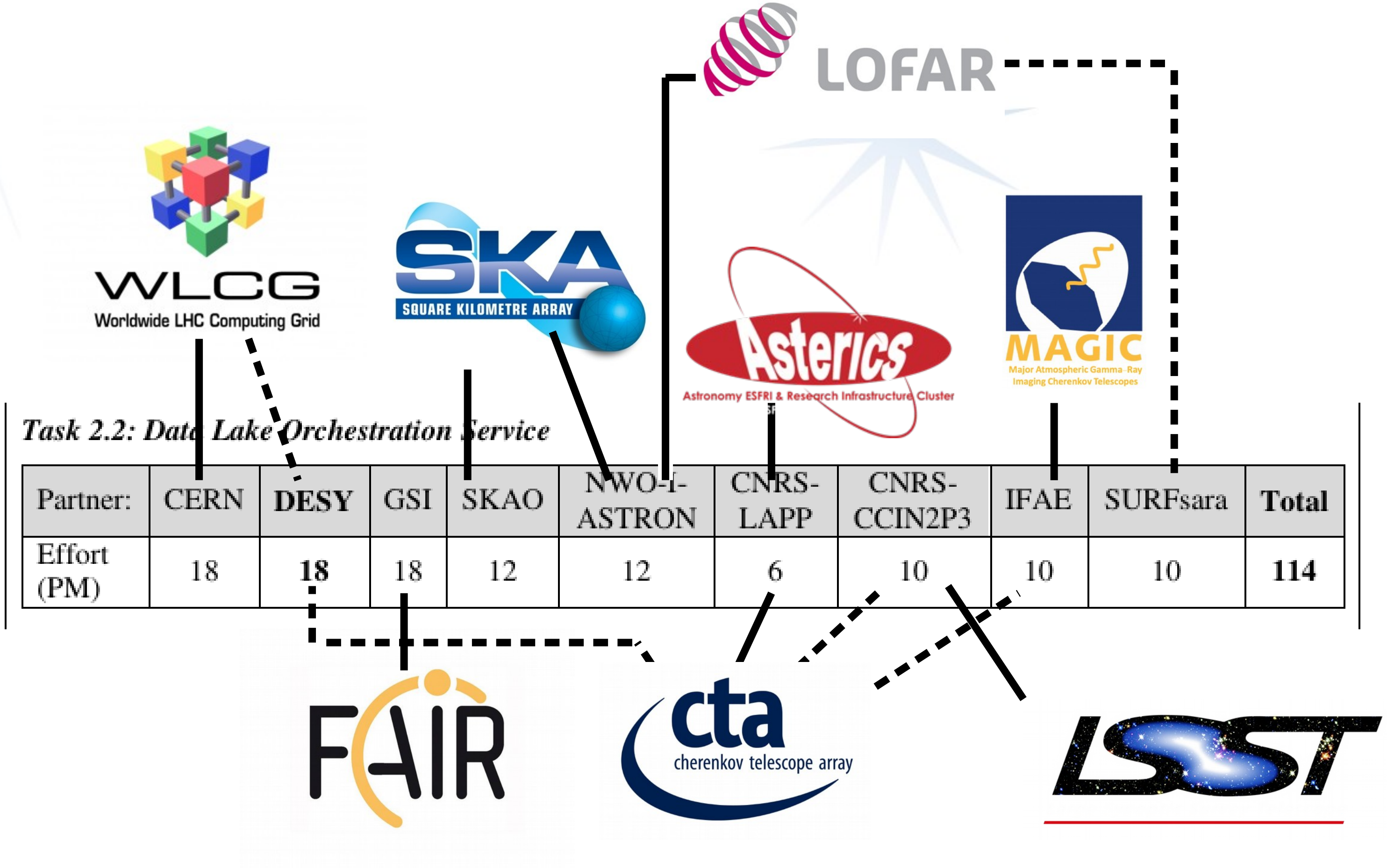
Links to the experiments


LOFAR
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster
MAGIC
Major Atmospheric Gamma-Ray Imaging Cherenkov Telescopes
cta
cherenkov telescope array
LSST

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Links to the experiments

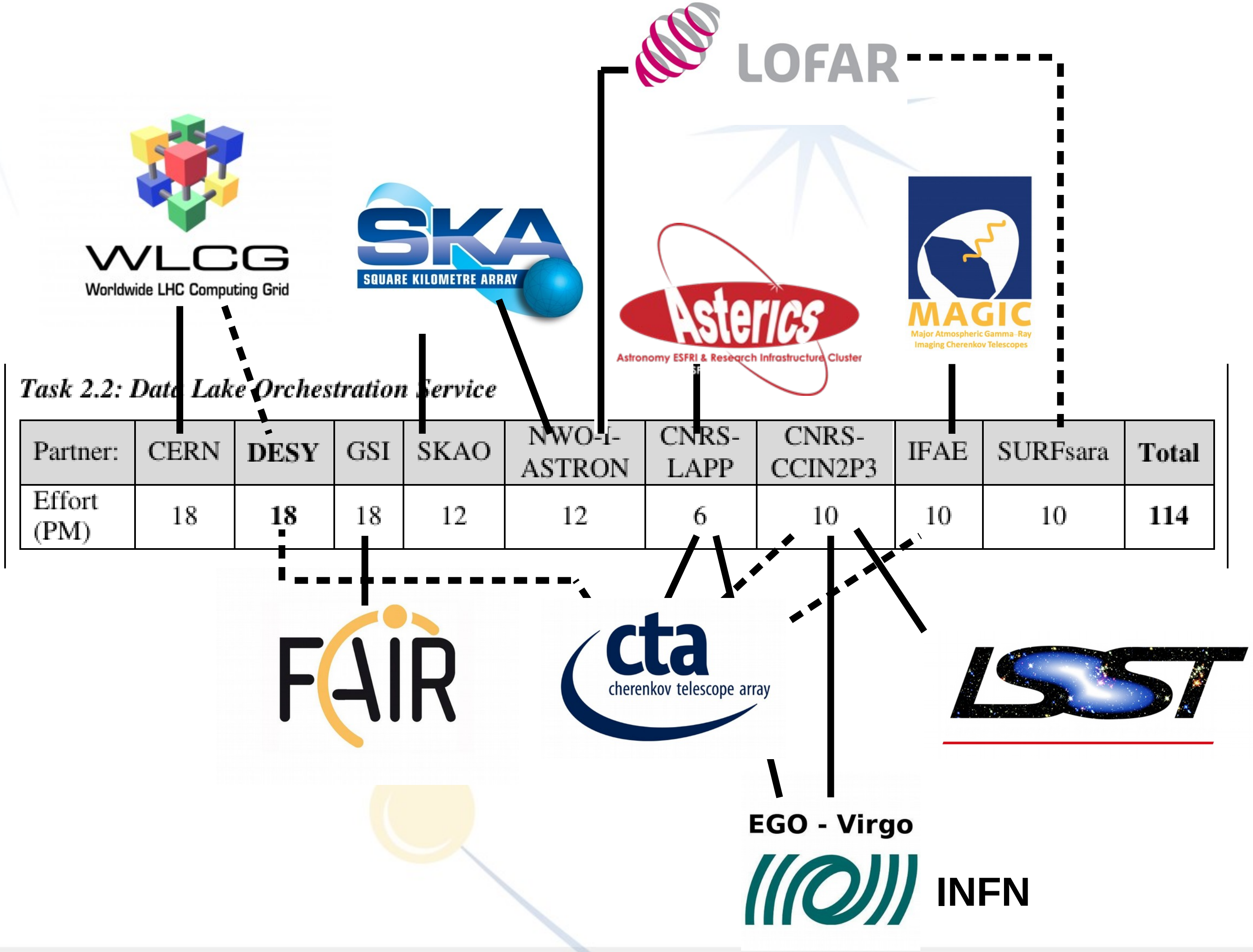

LOFAR
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster
MAGIC
Major Atmospheric Gamma-Ray Imaging Cherenkov Telescopes
FAIR
cta
cherenkov telescope array
LSST

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Funded by the European Union's Horizon 2020 - Grant N° 824064

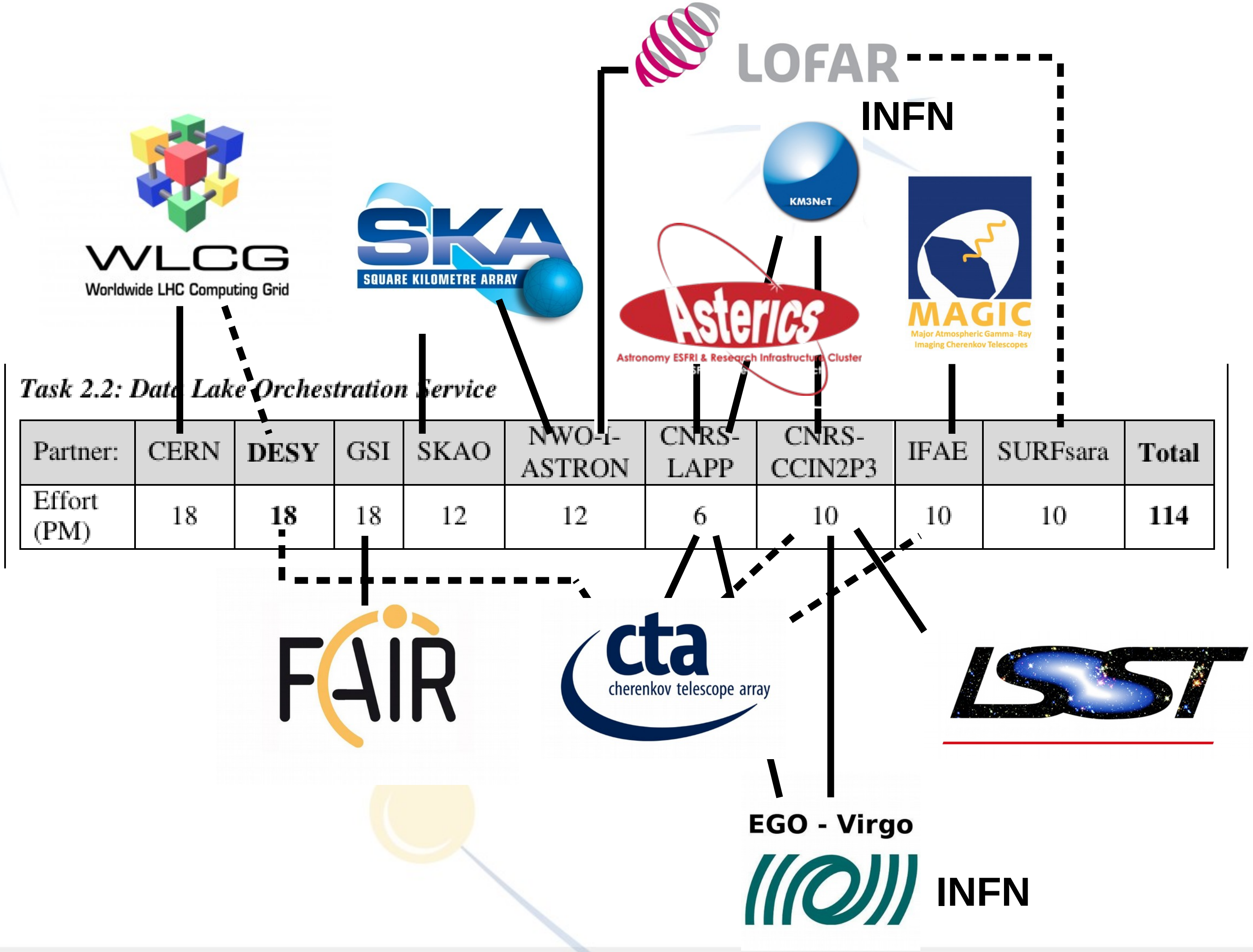
Links to the experiments


LOFAR
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster
MAGIC
Major Atmospheric Gamma-Ray Imaging Cherenkov Telescopes
FAIR
cta
cherenkov telescope array
LSST
EGO - Virgo
INFN

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

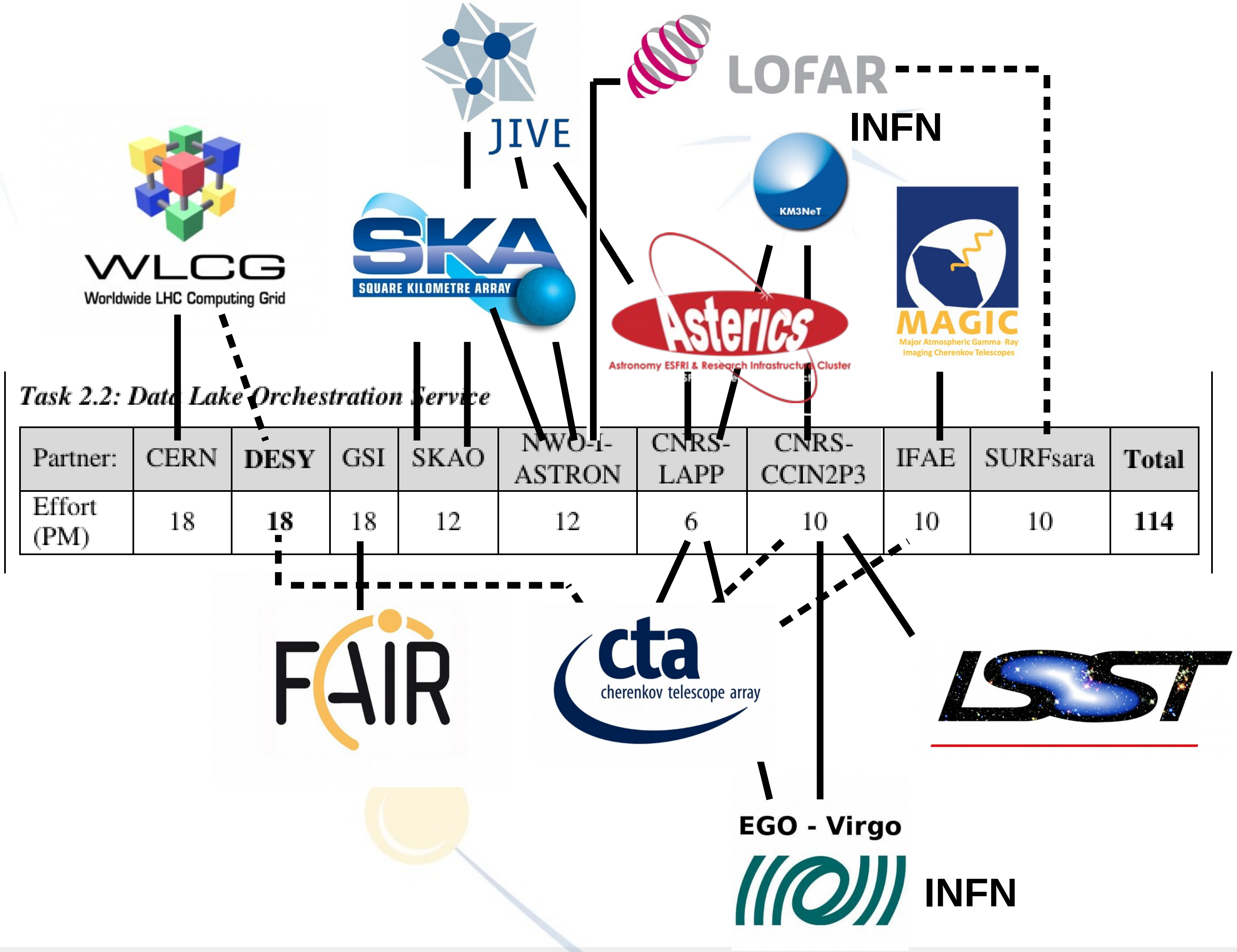
Links to the experiments


LOFAR
INFN
KM3NeT
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster
MAGIC
Major Atmospheric Gamma-Ray Imaging Cherenkov Telescopes
FAIR
cta
cherenkov telescope array
LSST
EGO - Virgo
INFN

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

Links to the experiments



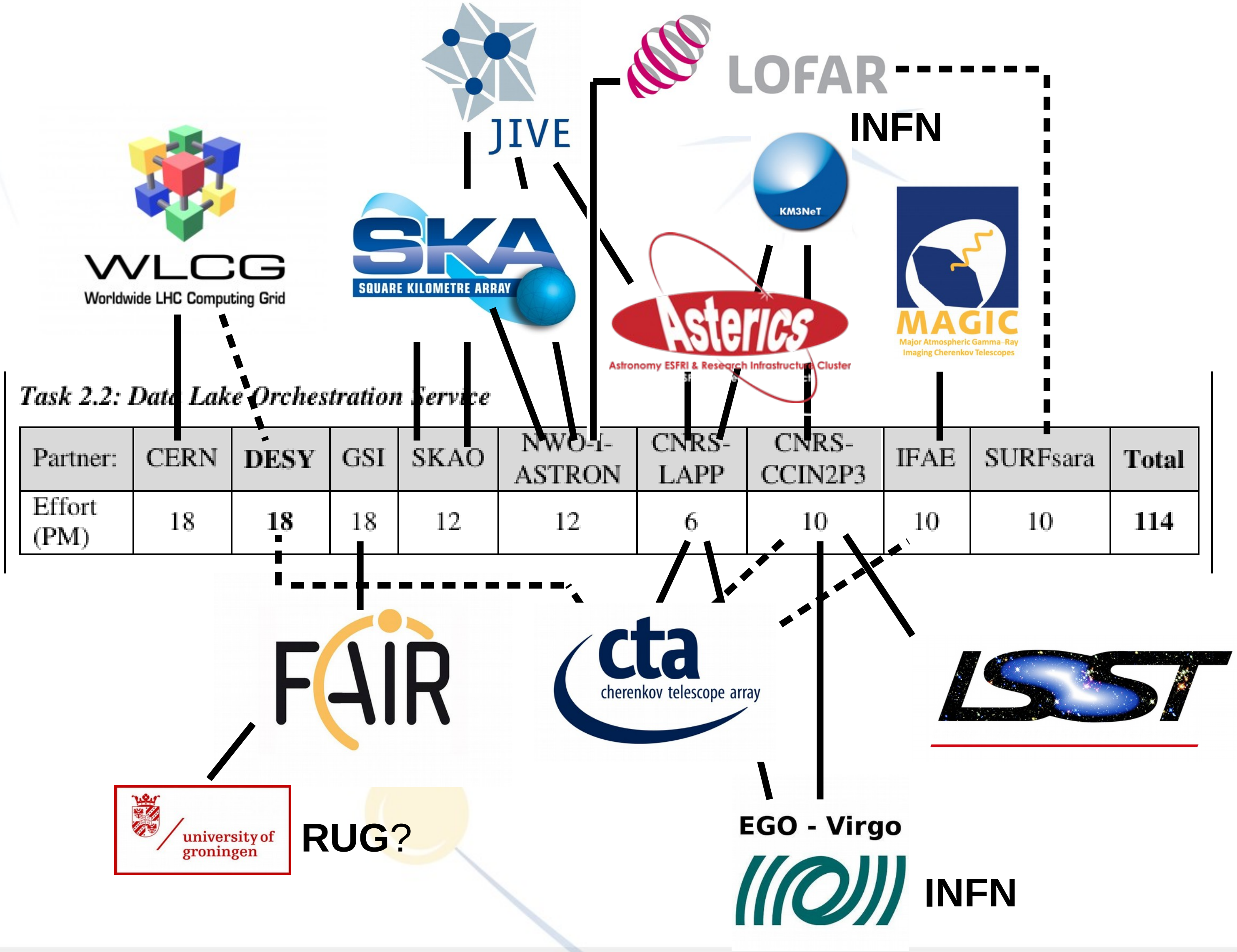
Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

JIVE
LOFAR
INFN
KM3NeT
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster
MAGIC
Major Atmospheric Gamma-Ray Imaging Cherenkov Telescopes
FAIR
cta
cherenkov telescope array
LSST
EGO - Virgo
INFN

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Links to the experiments


JIVE
LOFAR
KM3NeT
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster
MAGIC
Major Atmospheric Gamma-Ray Imaging Cherenkov Telescopes
FAIR
cta
cherenkov telescope array
LSST
university of groningen

INFN

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

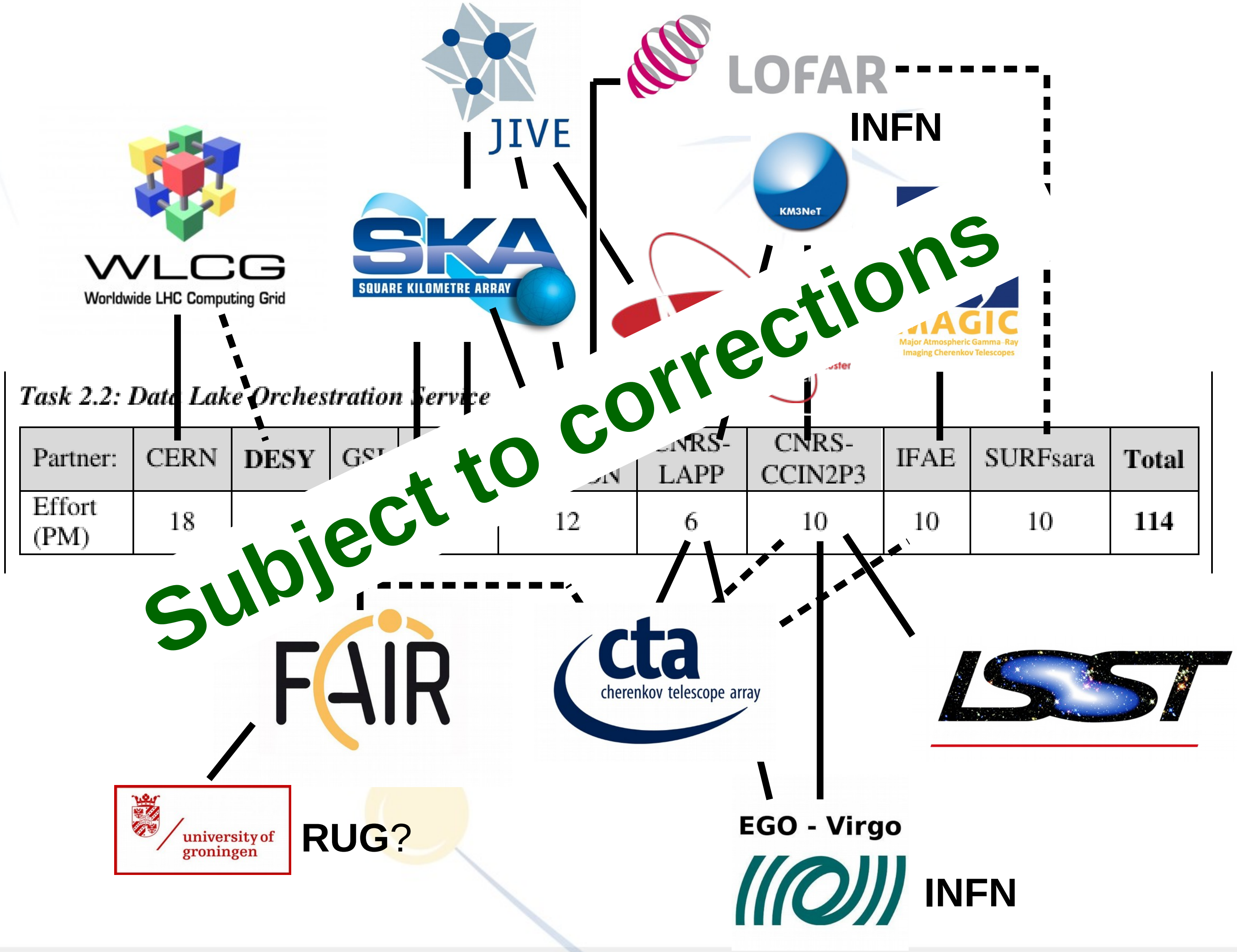
RUG?

EGO - Virgo

INFN

Funded by the European Union's Horizon 2020 - Grant N° 824064

Links to the experiments


JIVE
LOFAR
INFN
KM3NeT
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
MAGIC
Major Atmospheric Gamma-Ray Imaging Cherenkov Telescopes
FAIR
cta
cherenkov telescope array
LSST
university of groningen
RUG?
EGO - Virgo
INFN

Subject to corrections

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	DESY	GSI	[illegible]	[illegible]	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	[illegible]	[illegible]	[illegible]	12	6	10	10	10	**114**

Funded by the European Union's Horizon 2020 - Grant N° 824064

Experiments

Make contact

Organise interview(s)

Help identify places where QoS makes sense

Help translate these to specific requirements

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

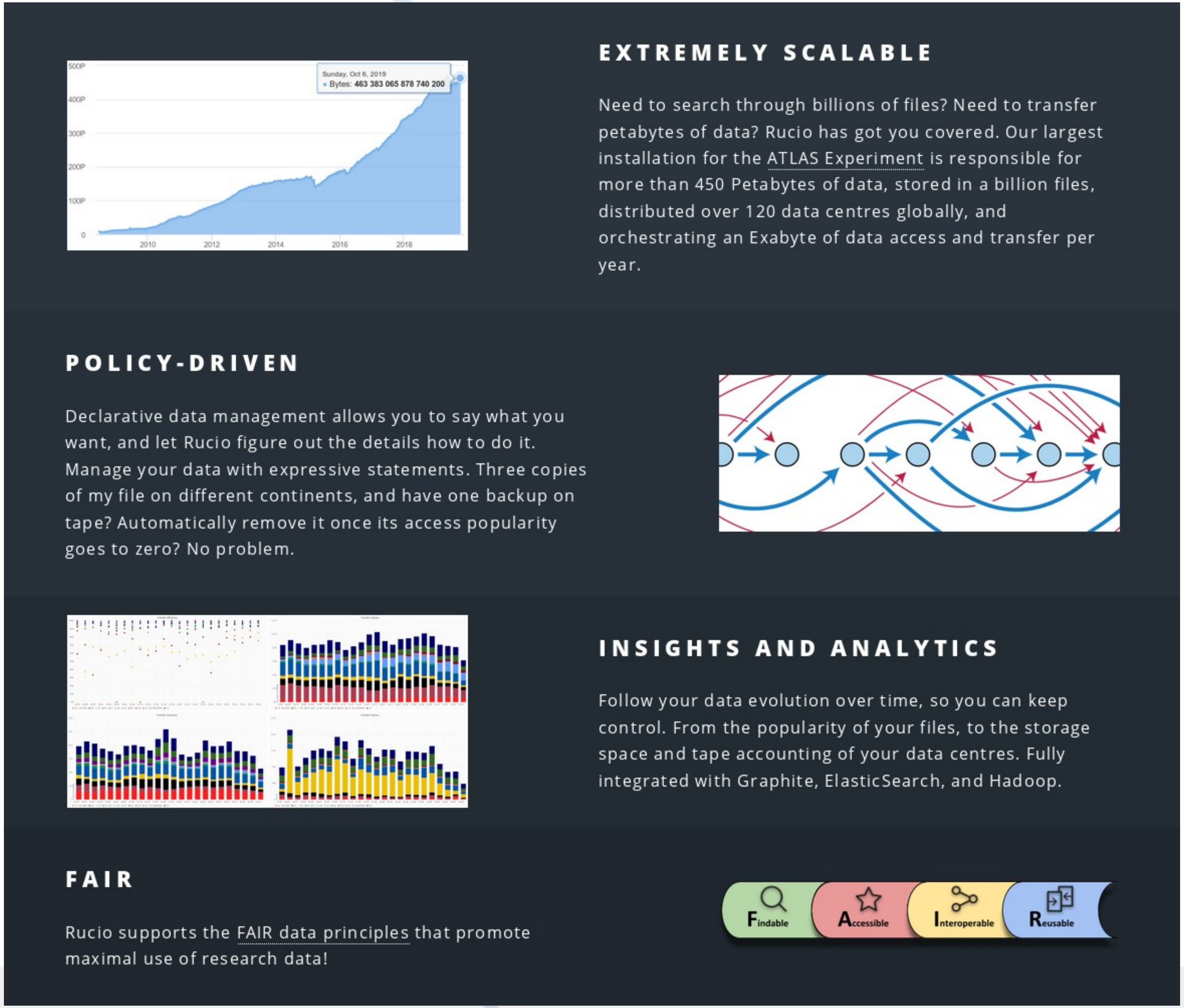
Rucio

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Rucio helps you to manage your community's data

Built on more than a decade of experience, Rucio serves the data needs of modern scientific experiments.

Large amounts of data, countless numbers of files, heterogeneous storage systems, globally distributed data centres, monitoring and analytics. All coming together in modular solution to fit your needs.



EXTREMELY SCALABLE

Need to search through billions of files? Need to transfer petabytes of data? Rucio has got you covered. Our largest installation for the ATLAS Experiment is responsible for more than 450 Petabytes of data, stored in a billion files, distributed over 120 data centres globally, and orchestrating an Exabyte of data access and transfer per year.

POLICY-DRIVEN

Declarative data management allows you to say what you want, and let Rucio figure out the details how to do it. Manage your data with expressive statements. Three copies of my file on different continents, and have one backup on tape? Automatically remove it once its access popularity goes to zero? No problem.

INSIGHTS AND ANALYTICS

Follow your data evolution over time, so you can keep control. From the popularity of your files, to the storage space and tape accounting of your data centres. Fully integrated with Graphite, ElasticSearch, and Hadoop.

FAIR

Rucio supports the FAIR data principles that promote maximal use of research data!

From https://rucio.cern.ch/

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Rucio: what can it do now?

Rucio has **RSE** to represent a storage system.

RSEs have **metadata** (a set of key-value pairs). We can add a QoS key-value pair; e.g., "qos=fast" or "qos=cold".

Rules make be written for collections/datasets to target a specific QoS, by targeting a specific metadata value. This is comparable to how data is managed generally.

QoS transitions are supported by Rucio in much the same way as regular data transfers. Rucio is told that a dataset should be present with a specific QoS by including that QoS value as a predicate in the rule. This, in turn, results in Rucio only considering RSEs that contain that metadata value (i.e., have this particular QoS). If the dataset is not already there, this triggers data replication.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Rucio: what can it do now?

Limitations:

- Each QoS class that a storage system supports must exist as separate RSE (Rucio cannot understand a storage system that supports multiple QoS).
- The path of the data must be sufficient to control the data's QoS.
- QoS transitions are only handled by transferring data, which may be inefficient if the storage can handle such changes internally.
- Mapping between what the storage system supports to RSE label is a manual process

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Status of Rucio in the testbed.

ESCAPE is running the latest version of Rucio. We anticipate being able to deploy new versions of Rucio once new features are available.

The exact Rucio configuration currently isn’t that clear to me (Paul)

- Some information is available in CRIC (e.g., list of RSEs, the transfer matrix)
- RSE metadata and rules *seem* not to be available.

Aris knows the configuration and status of the testbed. We just need to liaise with him to learn more.

Xavi (et al) are responding quickly in making information available.

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

What could we demo now?

We can set up **multiple RSEs** to represent different QoS options for the same data:

Storage with a single QoS represented with a single RSE.

The storage systems (with multiple QoS) must be able to support **different QoS for different paths**.

This option is already available in dCache, DPM and EOS, and likely the other storage systems, too.

We can demonstrate a user **simulating a QoS change** by modifying Rucio rules. These would trigger FTS to copy data, achieved using a third-party copy.

Rucio can either proactively delete the old data, or leave it as a cache copy.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

What new features could we see?

From Martin Barisits: Anticipate being able to start development work within Rucio early next year. This is with Aris finishing off his work on metadata within Rucio.

We would like to add support for:

- an RSE supporting **multiple QoS classes**,
- Rucio being able to **trigger QoS changes** by talking directly with the storage system.
- Storage systems providing a **list of QoS classes** they support, along with a list of attributes for each QoS class.
- Rucio **selecting QoS classes** based on the storage-system-provided attributes and user criteria.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

The DOMA-QoS white-paper

Funded by the European Union's
Horizon 2020 - Grant N° 824064

DOMA-QoS white paper

QoS is a **fuzzy term**: talking to user communities about it is difficult because many people talk about QoS and mean different things.

DOMA-QoS group writing a **white paper** that explains what *we* mean by QoS: what it is and how it works.

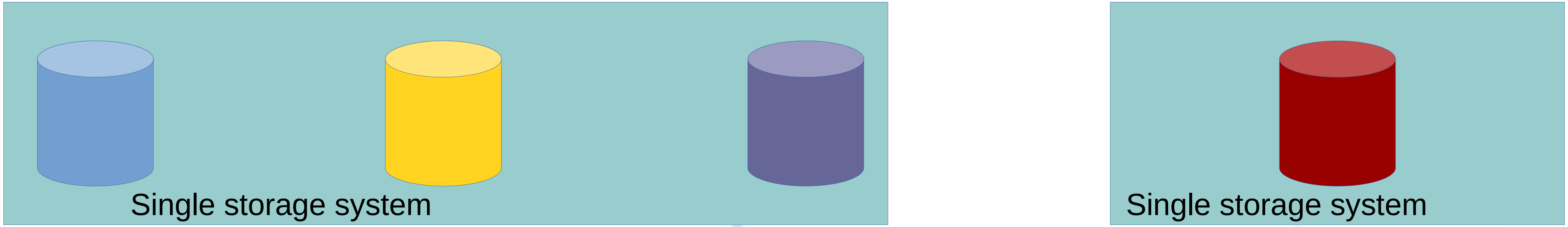
- Naturally, this isn't a definitive work, other definitions exist.
- Aimed as a primer for discussion with WLCG experiments.
- The document may evolve if feedback identifies missing information

The document is currently in first draft. We will be finished by the end of the week, and circulated for public comment.

<Insert bad pun about white paper and white Christmas here>

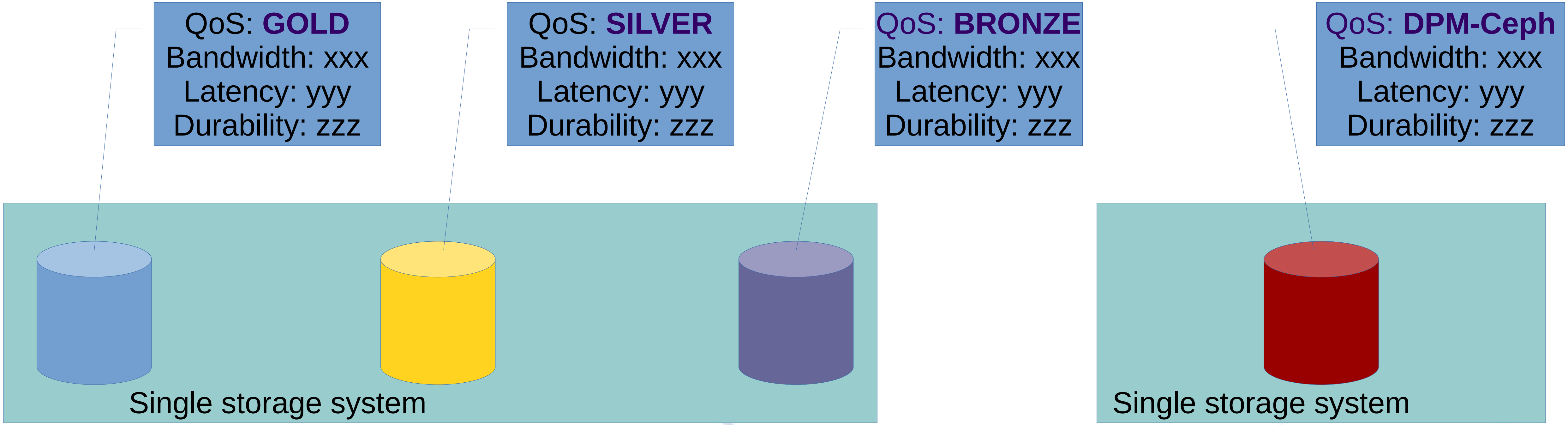
Funded by the European Union's
Horizon 2020 - Grant N° 824064

DOMA-QoS white paper



Funded by the European Union's
Horizon 2020 - Grant N° 824064

DOMA-QoS white paper



QoS: **GOLD**
Bandwidth: xxx
Latency: yyy
Durability: zzz

QoS: **SILVER**
Bandwidth: xxx
Latency: yyy
Durability: zzz

QoS: **BRONZE**
Bandwidth: xxx
Latency: yyy
Durability: zzz

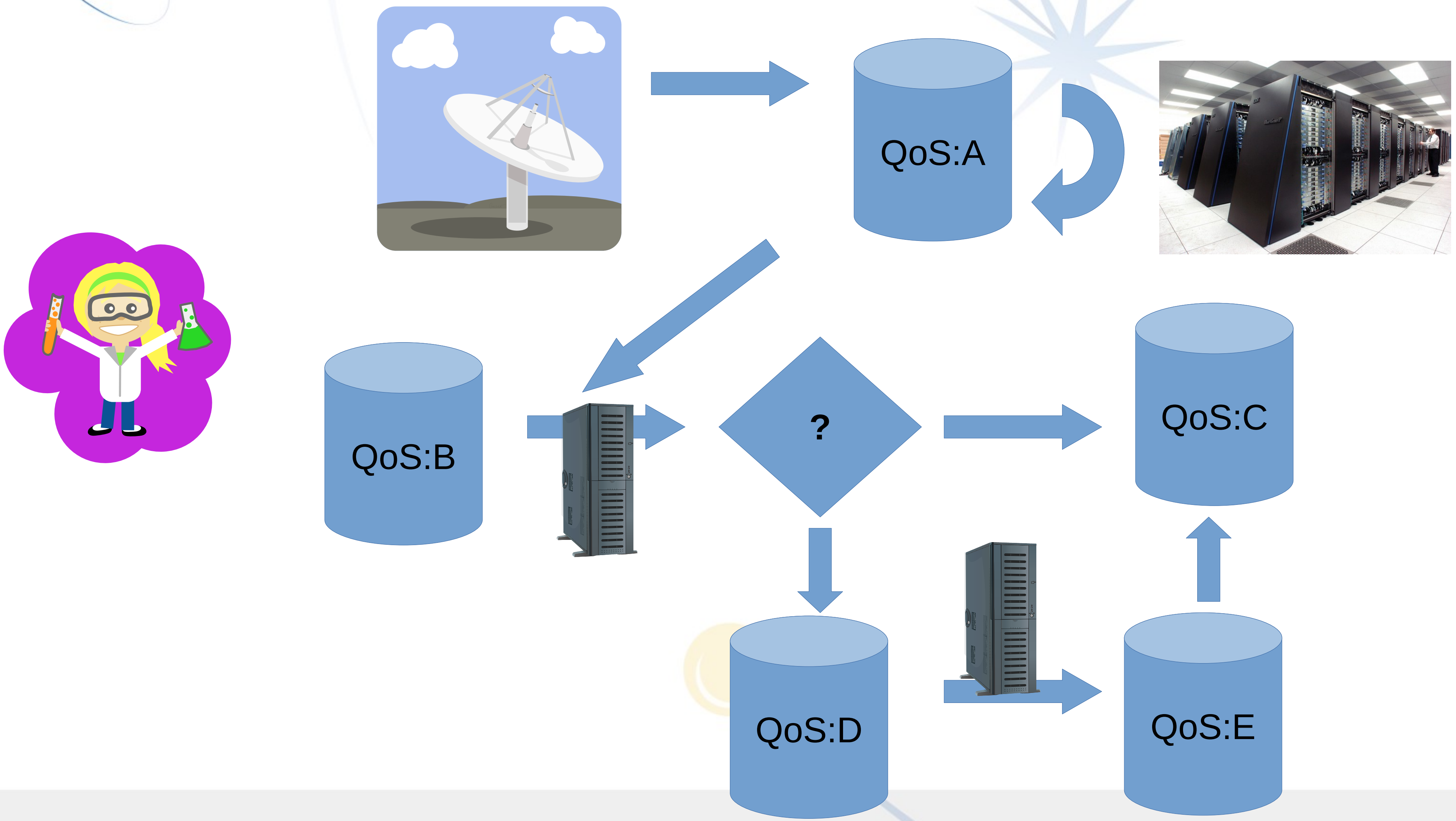
QoS: **DPM-Ceph**
Bandwidth: xxx
Latency: yyy
Durability: zzz

Single storage system

Single storage system

Funded by the European Union's
Horizon 2020 - Grant N° 824064

DOMA-QoS white paper: work-flow



QoS:A

QoS:B

?

QoS:C

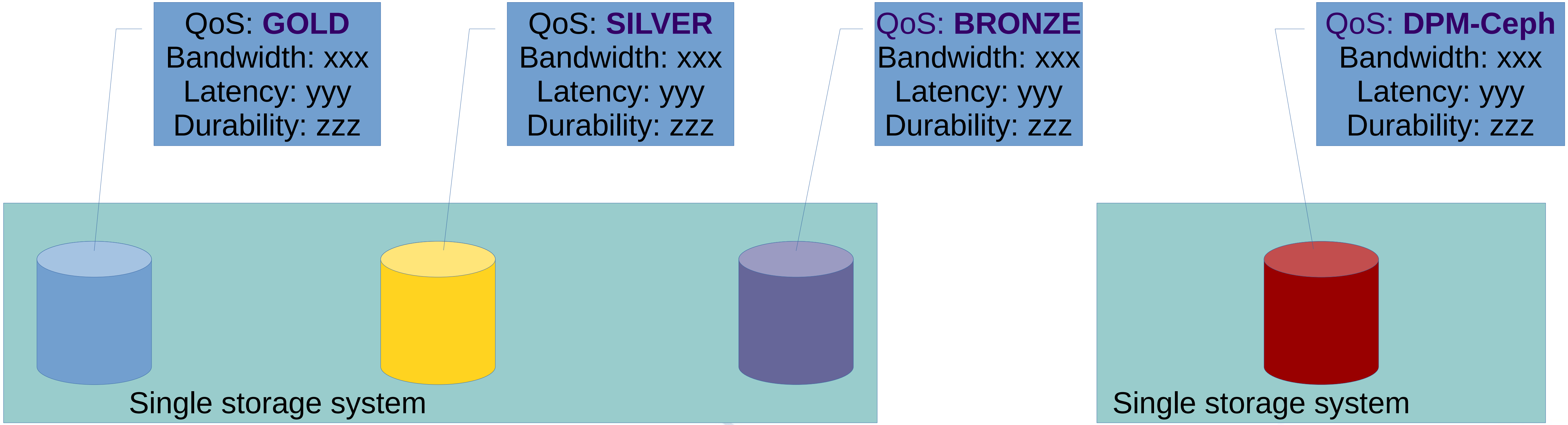
QoS:D

QoS:E

Funded by the European Union's
Horizon 2020 - Grant N° 824064

DOMA-QoS white paper

QoS:A QoS:B QoS:C QoS:D QoS:E



QoS: **GOLD**
Bandwidth: xxx
Latency: yyy
Durability: zzz

QoS: **SILVER**
Bandwidth: xxx
Latency: yyy
Durability: zzz

QoS: **BRONZE**
Bandwidth: xxx
Latency: yyy
Durability: zzz

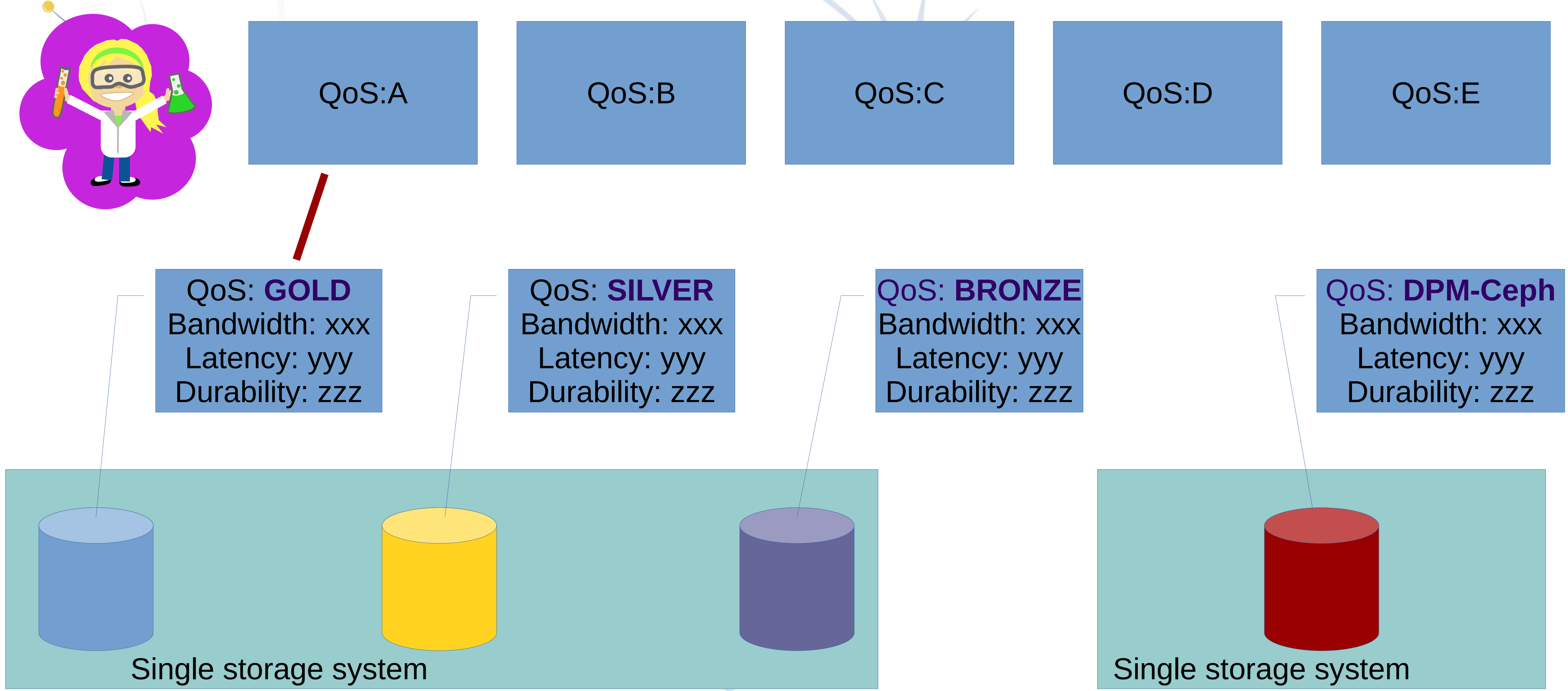
QoS: **DPM-Ceph**
Bandwidth: xxx
Latency: yyy
Durability: zzz

Single storage system

Single storage system

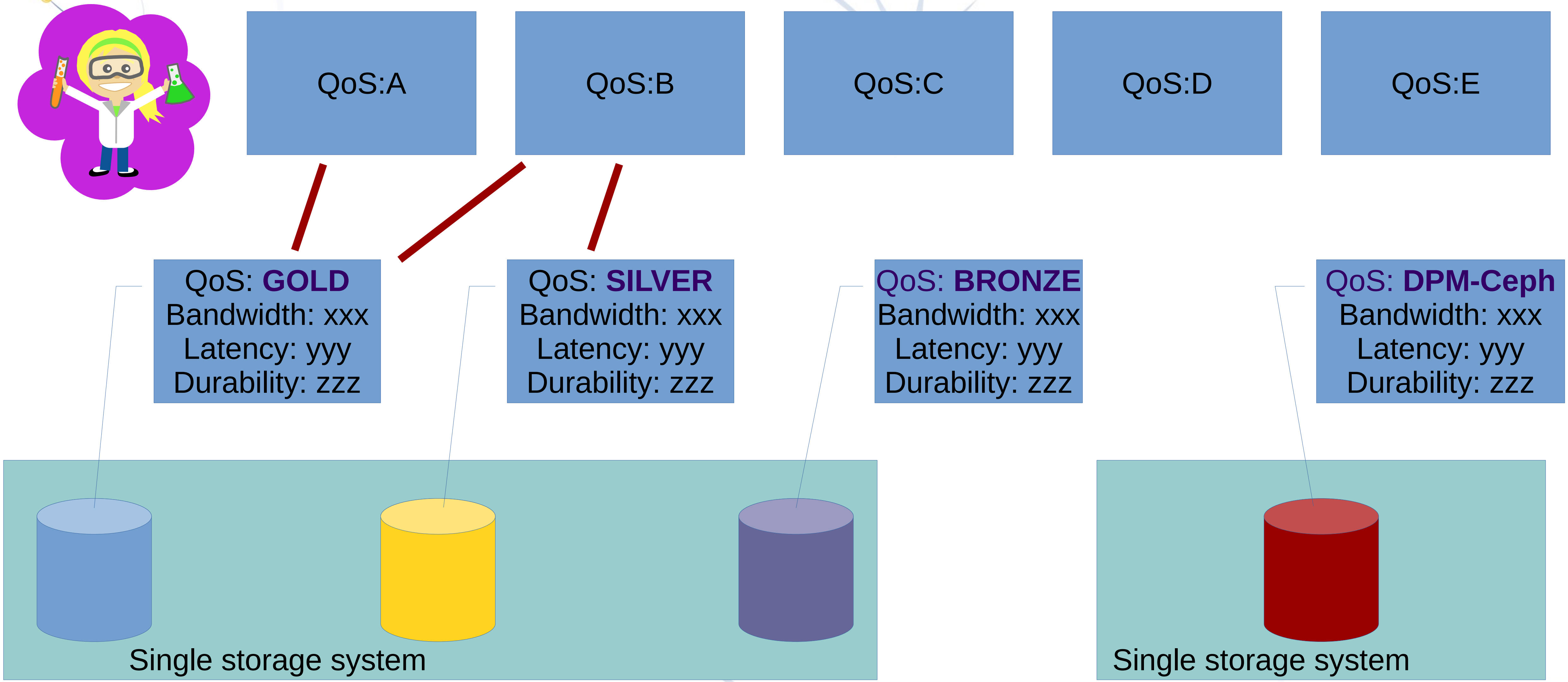
Funded by the European Union's
Horizon 2020 - Grant N° 824064

DOMA-QoS white paper



QoS:A
QoS:B
QoS:C
QoS:D
QoS:E
QoS: **GOLD** Bandwidth: xxx Latency: yyy Durability: zzz
QoS: **SILVER** Bandwidth: xxx Latency: yyy Durability: zzz
QoS: **BRONZE** Bandwidth: xxx Latency: yyy Durability: zzz
QoS: **DPM-Ceph** Bandwidth: xxx Latency: yyy Durability: zzz
Single storage system
Single storage system

DOMA-QoS white paper



QoS:A

QoS:B

QoS:C

QoS:D

QoS:E

QoS: **GOLD**
Bandwidth: xxx
Latency: yyy
Durability: zzz

QoS: **SILVER**
Bandwidth: xxx
Latency: yyy
Durability: zzz

QoS: **BRONZE**
Bandwidth: xxx
Latency: yyy
Durability: zzz

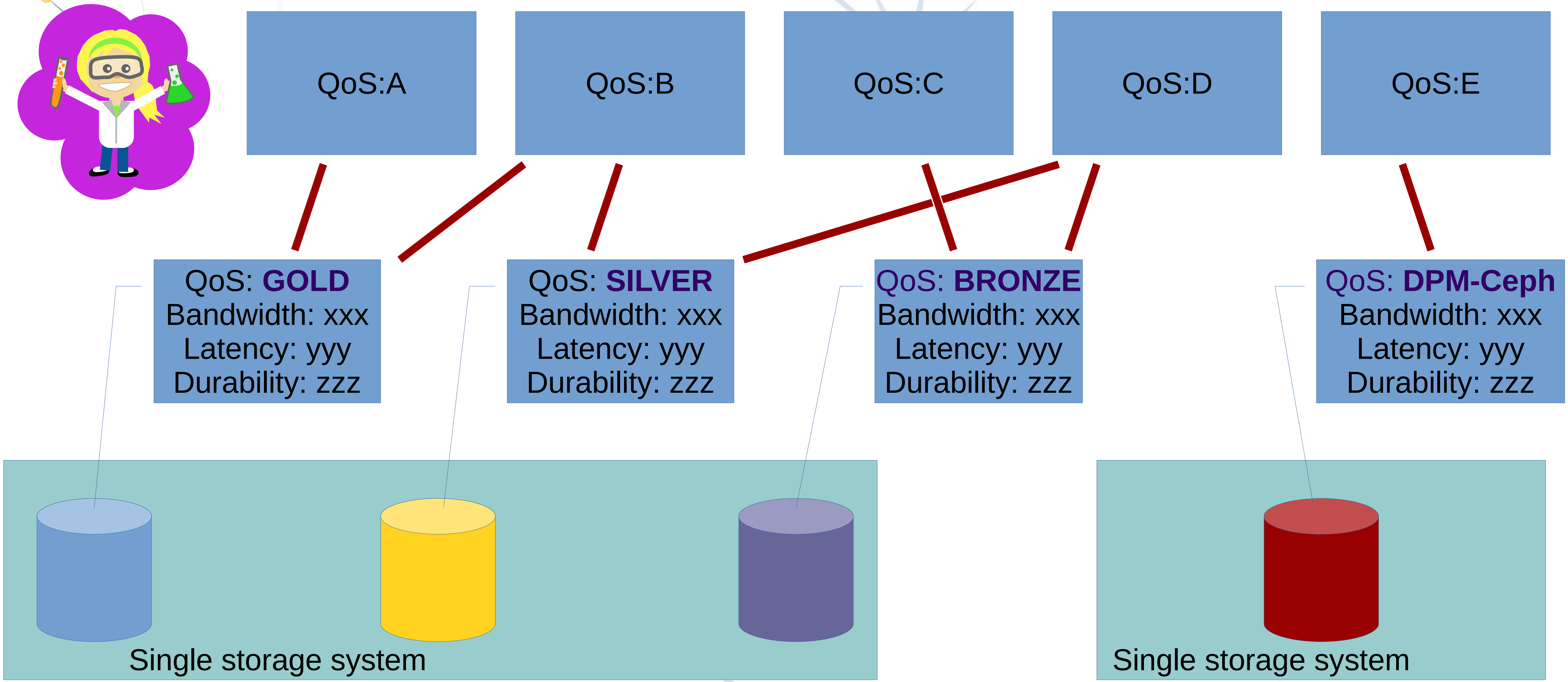
QoS: **DPM-Ceph**
Bandwidth: xxx
Latency: yyy
Durability: zzz

Single storage system

Single storage system

Funded by the European Union's
Horizon 2020 - Grant N° 824064

DOMA-QoS white paper



QoS:A

QoS:B

QoS:C

QoS:D

QoS:E

QoS: **GOLD**
Bandwidth: xxx
Latency: yyy
Durability: zzz

QoS: **SILVER**
Bandwidth: xxx
Latency: yyy
Durability: zzz

QoS: **BRONZE**
Bandwidth: xxx
Latency: yyy
Durability: zzz

QoS: **DPM-Ceph**
Bandwidth: xxx
Latency: yyy
Durability: zzz

Single storage system

Single storage system

Funded by the European Union's
Horizon 2020 - Grant N° 824064

DOMA-QoS white paper

New VO QoS class driving new deployments:

- Work-flow updated, identifing new VO QoS Class.
- Sites try to optimise for this new VO QoS Class by introducing new Storage QoS Class.

New Storage QoS class driving work-flows:

- New technology deployed at site → new Storage QoS Class
- Experiments identify new VO QoS Classes in workflow.

Alternative Storage QoS class to replace existing class:

- Site free to install new (cheaper) service that satisfies minimum requirement.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Plans for the future

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Plans for the future

Propose tasks are split into three "streams". The streams are more-or-less independent of each other, but the tasks within the streams are completed in order.

Stream A: prototyping and demonstration.

Stream B: engagement with experiments.

Stream C: software development

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Stream-A: prototyping & demo

A.1 Collect information about what QoS options are currently available in the ESCAPE testbed. This is by a simple survey.

A.2 Build prototype QoS in ESCAPE testbed by adding QoS-specific RSEs and QoS-specific labels.

A.3 Provide simple proof-of-concept demonstration to show Rucio rules may be used to drive QoS transitions.

A.4 Deploy updated versions of Rucio, FTS and storage software as they become available.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Stream-B: engagement with experiments

B.1 Build initial contact with experiments data-management expert through ESCAPE sites.

B.2 Conduct one-on-one interviews with experiments to work through their intended work-flows, to identity desired QoS and places where QoS transitions may be beneficial.

B.3 Map abstract experiment-specific QoS concepts into what Rucio and storage software can provide. This may trigger software development if a limitation is discovered.

B.4 Update testbed to match desired pattern. This steps may be delayed, waiting for new hardware, reconfigure their storage, or install updates. It may also block on Rucio development, if new features are required.

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Stream-C: software development

C.1 Put together an architecture/design paper.

Note, this distinct from the white-paper since the target audience is different (white paper: VOs, arch./design paper: software developers) and the two are different levels of abstraction.

C.2 Identify limitations of current approach, compared to the architecture.

C.3 Implement missing functionality. During this process, we anticipate fairly frequent deployment/updates to the ESCAPE testbed to allow for integration testing and to support B.4 and B.5.

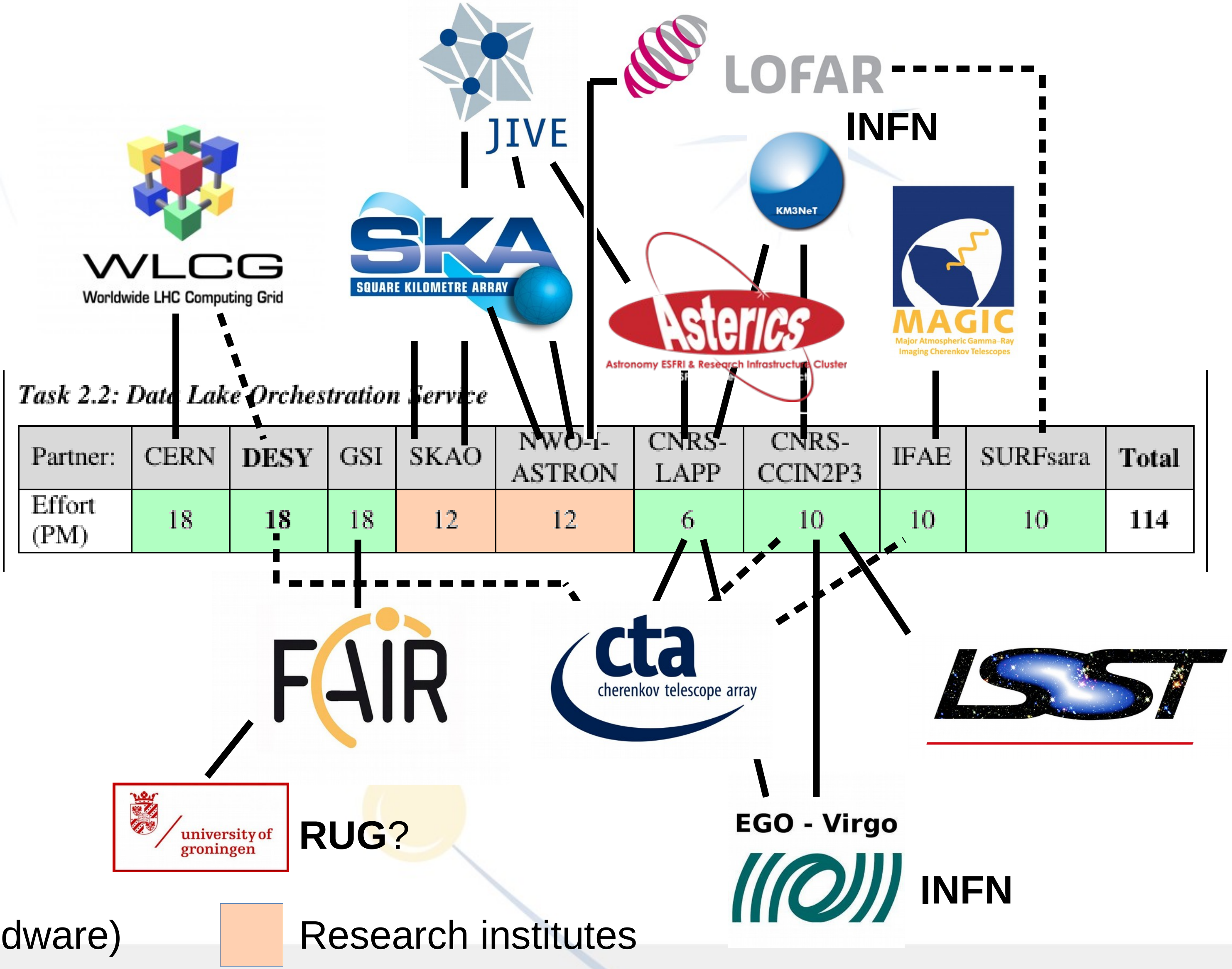
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Thanks for listening

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Links to the experiments

Subject to corrections


JIVE
LOFAR
INFN
KM3NeT
WLCG
Worldwide LHC Computing Grid
SKA
SQUARE KILOMETRE ARRAY
Asterics
Astronomy ESFRI & Research Infrastructure Cluster
MAGIC
Major Atmospheric Gamma-Ray Imaging Cherenkov Telescopes
FAIR
cta
cherenkov telescope array
LSST
university of groningen
RUG?
EGO - Virgo
INFN

Task 2.2: Data Lake Orchestration Service

Partner:	CERN	**DESY**	GSI	SKAO	NWO-I-ASTRON	CNRS-LAPP	CNRS-CCIN2P3	IFAE	SURFsara	**Total**
Effort (PM)	18	**18**	18	12	12	6	10	10	10	**114**

🟩 Sites (providing hardware)

🟧 Research institutes

Funded by the European Union's Horizon 2020 - Grant N° 824064