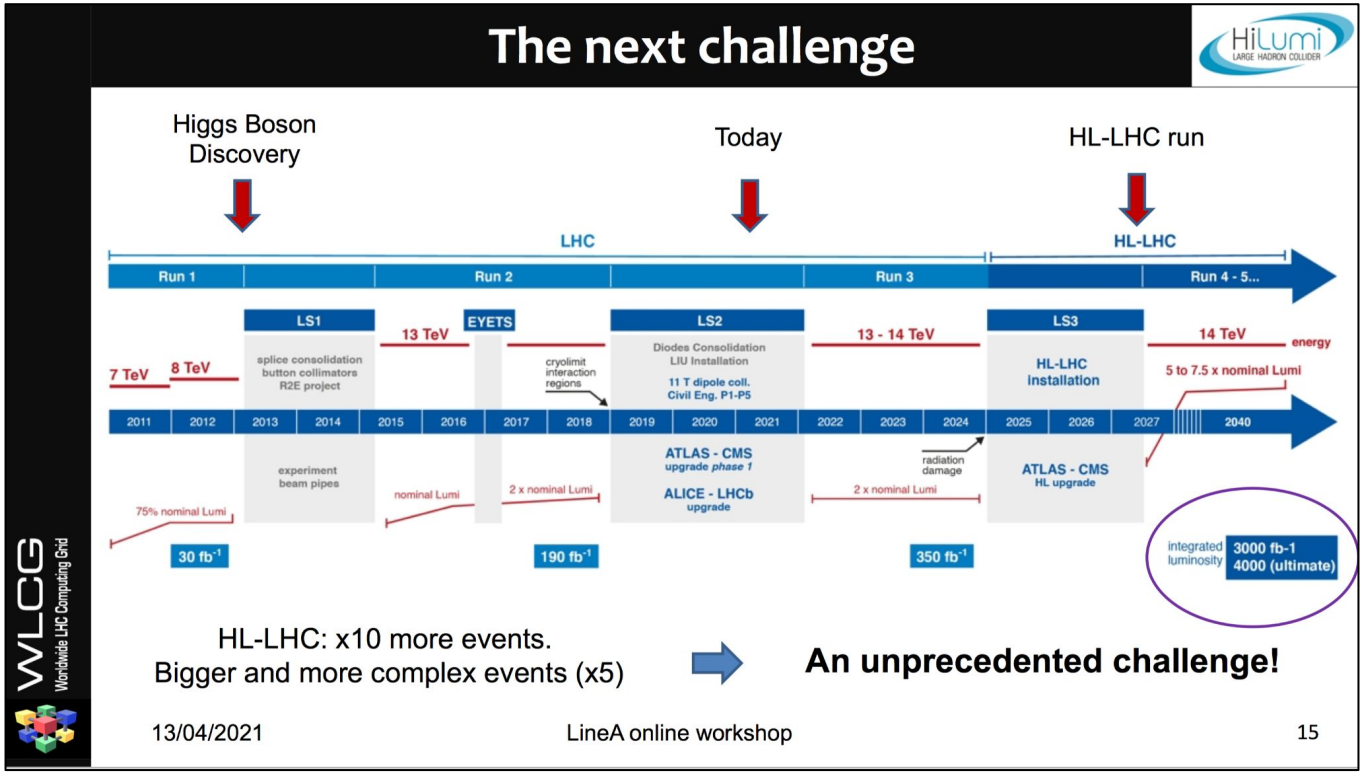
Data Management

J. Flix, J. M. Hernández

I Workshop de Computing y Software de la Red Española de LHC

The problem's face: Run3 → HL-LHC



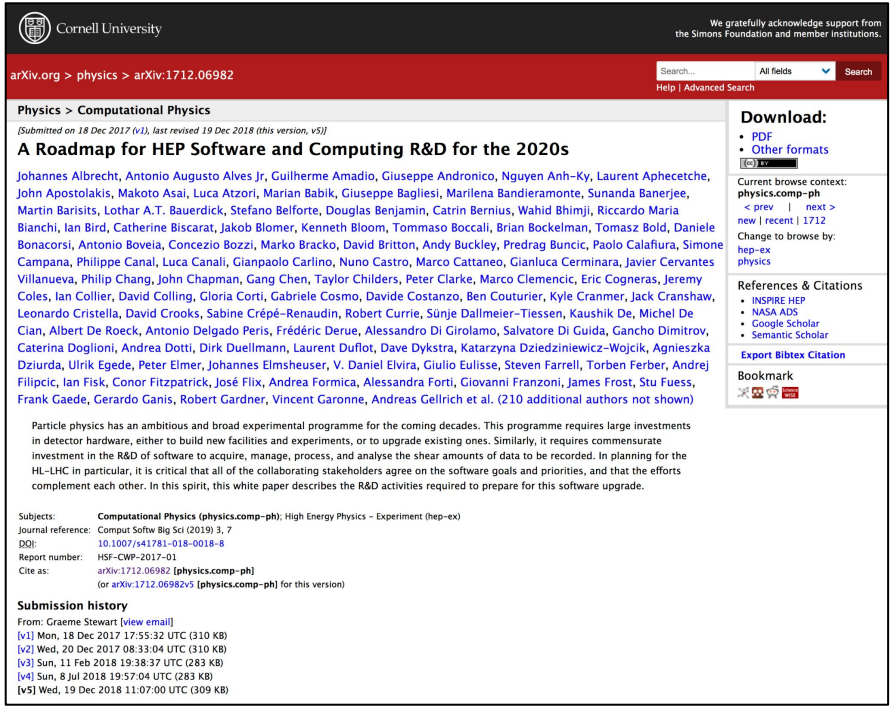
HL-LHC Computing Roadmap

In 2017 the **HEP Software Foundation**, charged by the WLCG, produced a roadmap white paper on the software and computing challenges that will be faced during the next decade

Community White Paper Reports

The roadmap summarised reports from fourteen working groups who studied the challenges in their sub-domains. All of the reports produced during the Community White Paper process are listed below. Working groups are in the process of finalising and uploading their work to arXiv.

Paper	*Report Number*	*Link*
CWP Roadmap	HSF-CWP-2017-01	arXiv
Careers & Training	HSF-CWP-2017-02	arXiv
Conditions Data	HSF-CWP-2017-03	arXiv
Data Organisation, Management and Access	HSF-CWP-2017-04	arXiv
Data Analysis and Interpretation	HSF-CWP-2017-05	arXiv
Data and Software Preservation	HSF-CWP-2017-06	arXiv
Detector Simulation	HSF-CWP-2017-07	arXiv
Event/Data Processing Frameworks	HSF-CWP-2017-08	arXiv
Facilities and Distributed Computing	HSF-CWP-2017-09	Google Doc
Machine Learning	HSF-CWP-2017-10	arXiv
Physics Generators	-	No separate paper, see CWP Roadmap, section 3.1
Security	-	No separate paper, see CWP Roadmap, section 3.13
Software Development, Deployment and Validation	HSF-CWP-2017-13	arXiv
Software Trigger and Event Reconstruction	HSF-CWP-2017-14	arXiv - Executive Summary; arXiv - full document
Visualisation	HSF-CWP-2017-15	arXiv



Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv.org > physics > arXiv:1712.06982

Search... All fields Search

Help | Advanced Search

Physics > Computational Physics

[Submitted on 18 Dec 2017 (v1), last revised 19 Dec 2018 (this version, v5)]

A Roadmap for HEP Software and Computing R&D for the 2020s

Johannes Albrecht, Antonio Augusto Alves Jr, Guilherme Amadio, Giuseppe Andronico, Nguyen Anh-Ky, Laurent Aphecetche, John Apostolakis, Makoto Asai, Luca Atzori, Marian Babik, Giuseppe Bagliesi, Marilena Bandieramonte, Sunanda Banerjee, Martin Barisits, Lothar A.T. Bauerdick, Stefano Belforte, Douglas Benjamin, Catrin Bernius, Wahid Bhimji, Riccardo Maria Bianchi, Ian Bird, Catherine Biscarat, Jakob Blomer, Kenneth Bloom, Tommaso Boccali, Brian Bockelman, Tomasz Bold, Daniele Bonacorsi, Antonio Boveia, Concezio Bozzi, Marko Bracko, David Britton, Andy Buckley, Predrag Buncic, Paolo Calafiura, Simone Campana, Philippe Canal, Luca Canali, Gianpaolo Carlino, Nuno Castro, Marco Cattaneo, Gianluca Cerminara, Javier Cervantes Villanueva, Philip Chang, John Chapman, Gang Chen, Taylor Childers, Peter Clarke, Marco Clemencic, Eric Cogneras, Jeremy Coles, Ian Collier, David Colling, Gloria Corti, Gabriele Cosmo, Davide Costanzo, Ben Couturier, Kyle Cranmer, Jack Cranshaw, Leonardo Cristella, David Crooks, Sabine Crépé-Renaudin, Robert Currie, Sünje Dallmeier-Tiessen, Kaushik De, Michel De Cian, Albert De Roeck, Antonio Delgado Peris, Frédéric Derue, Alessandro Di Girolamo, Salvatore Di Guida, Gancho Dimitrov, Caterina Doglioni, Andrea Dotti, Dirk Duellmann, Laurent Duflot, Dave Dykstra, Katarzyna Dziedziniewicz-Wojcik, Agnieszka Dziurda, Ulrik Egede, Peter Elmer, Johannes Elmsheuser, V. Daniel Elvira, Giulio Eulisse, Steven Farrell, Torben Ferber, Andrej Filipcic, Ian Fisk, Conor Fitzpatrick, José Flix, Andrea Formica, Alessandra Forti, Giovanni Franzoni, James Frost, Stu Fuess, Frank Gaede, Gerardo Ganis, Robert Gardner, Vincent Garonne, Andreas Gellrich et al. (210 additional authors not shown)

Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders agree on the software goals and priorities, and that the efforts complement each other. In this spirit, this white paper describes the R&D activities required to prepare for this software upgrade.

Subjects: **Computational Physics (physics.comp-ph)**; High Energy Physics - Experiment (hep-ex)
Journal reference: Comput Softw Big Sci (2019) 3, 7
DOI: 10.1007/s41781-018-0018-8
Report number: HSF-CWP-2017-01
Cite as: arXiv:1712.06982 **[physics.comp-ph]**
(or arXiv:1712.06982v5 **[physics.comp-ph]** for this version)

Submission history
From: Graeme Stewart [view email]
[v1] Mon, 18 Dec 2017 17:55:32 UTC (310 KB)
[v2] Wed, 20 Dec 2017 08:33:04 UTC (310 KB)
[v3] Sun, 11 Feb 2018 19:38:37 UTC (283 KB)
[v4] Sun, 8 Jul 2018 19:57:04 UTC (283 KB)
[v5] Wed, 19 Dec 2018 11:07:00 UTC (309 KB)

Download:
- PDF
- Other formats

Current browse context: **physics.comp-ph**
< prev | next >
new | recent | 1712
Change to browse by:
hep-ex
physics

References & Citations
- INSPIRE HEP
- NASA ADS
- Google Scholar
- Semantic Scholar

Export Bibtex Citation

Bookmark

DOMA Working Group

In 2018 the WLCG DOMA R&D Working Group was launched, covering several activities in the area of **Data Organization, Management and Access**, with a focus on the medium/long term evolution

Twiki → https://twiki.cern.ch/twiki/bin/view/LCG/DomaActivities
Indico area → https://indico.cern.ch/category/10360/

3 active groups on:

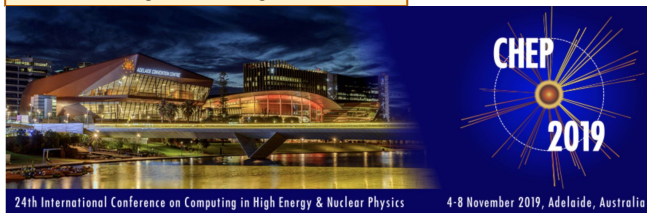
- *Data Access, Content Delivery and Caching*
- *Third Party Copy*
- *Quality of Service*

DOMA oriented events

Since then, many meetings, workshops, sessions, papers, reviews, etc...
lot of progress! very active area!

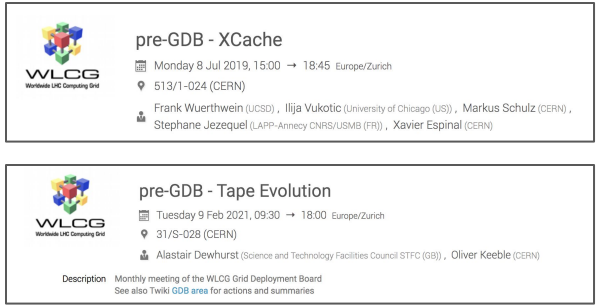
BoF session

Track 4 – Data Organisation, Management and Access

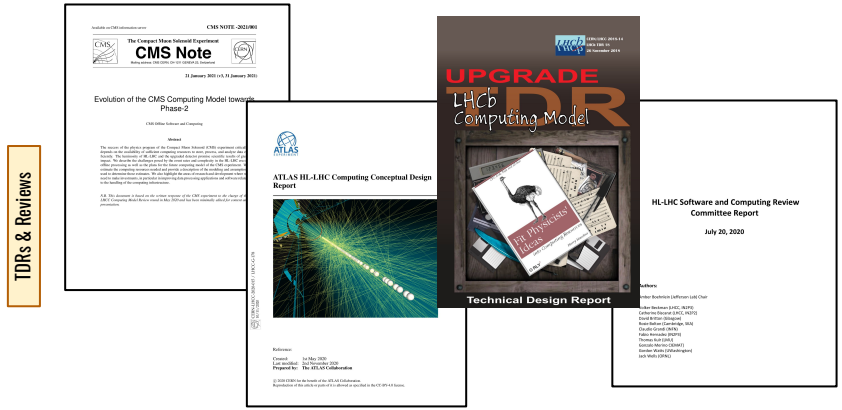


Computing sessions focused on storage and its evolution towards the HL-LHC needs

pre-GDBs



TDRs & Reviews



DOMA Working Group - new phase

The **first phase** of DOMA has concluded

The main goals over two and half years were:

a) creating a forum to discuss and foster ideas (and issues) on data management across experiments
b) evaluate and evolve technologies toward future data management models, as defined in the HFS Community White Paper and the WLCG HL-LHC computing strategy documents

The **second phase** of DOMA will focus on defining prototypes based on those technologies, measuring their effectiveness and commissioning them progressively toward HL-LHC through a series of **data challenges**

Of course, continuing offering a **forum to discuss** and **foster ideas** (and raise issues) remains as a primary goal

DOMA Working Group - new phase

Opening for two new co-coordinators for DOMA phase-2 ← tell me (or Simone) if interested!

Mandate

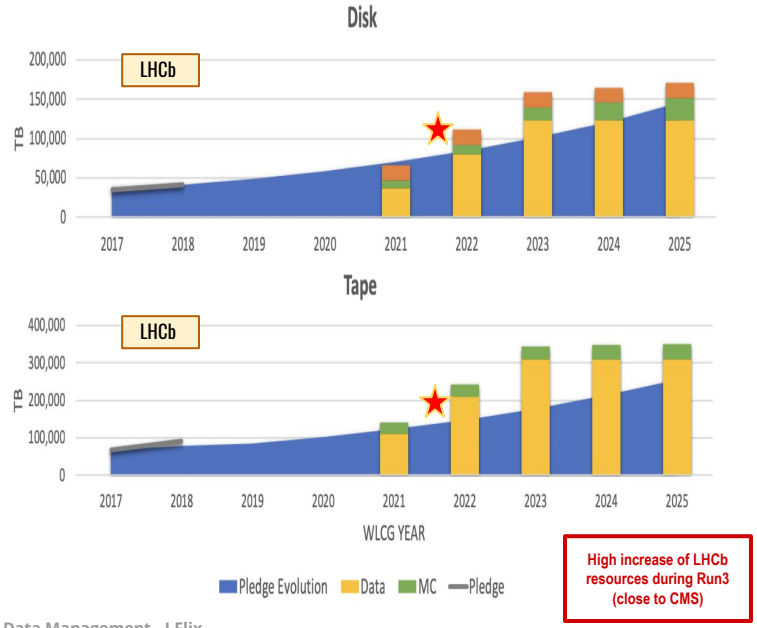
The WLCG DOMA sub-project provides a forum to discuss and foster ideas on all aspects of data management across experiments. It facilitates the process of evaluating and evolving data management technologies towards future models addressing the HL-LHC computing challenge. The DOMA co-coordinators are mandated to oversee the activities of the WLCG DOMA sub-project and in particular the R&D process leading to the implementation of novel solutions. This activity consists of collecting and documenting the different solutions and what they intend to achieve, what are the boundary conditions that led to defining a solution and how it is foreseen to integrate it with an experiment workflow and data management system. Defining the metrics to be measured to determine success and on which timescale is part of this process. DOMA as a sub-project should follow the evolution of the different implementations and facilitate the exchange of information and know-how. It should facilitate interactions between experiments and sites and support common definitions, goals and processes. The DOMA co-coordinators are appointed by and report to the WLCG Management Board, after an open call for nominations. Their appointment normally lasts 2 years

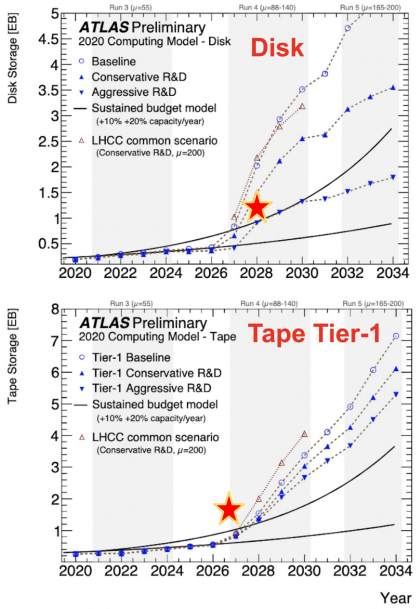
Storage projections towards HL-LHC

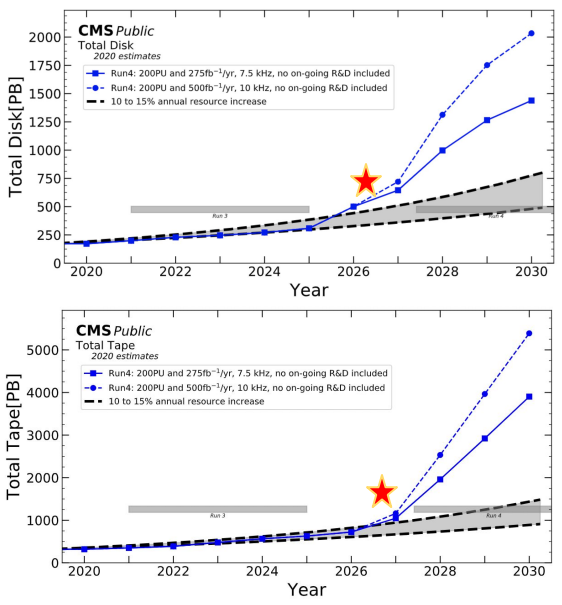
Storage is the most expensive resource to deploy and operate in WLCG

These tasks resulted in **developments** to reduce needs (costs) of storage infrastructure

→ Data lakes, reduce data size and duplication, analysis via caches, central processing through buffers, …





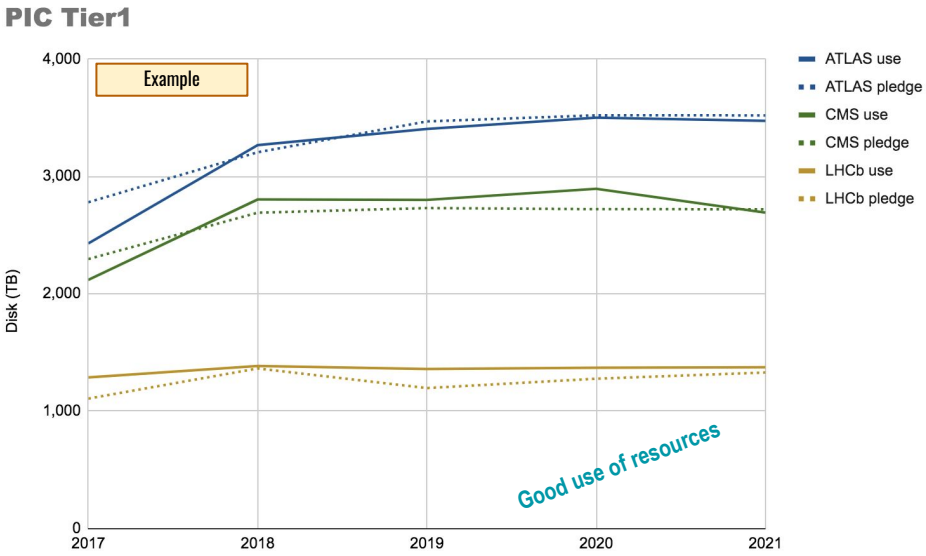


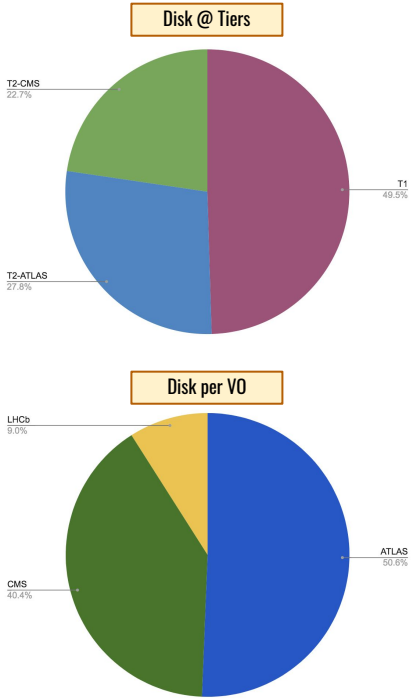
Flat-funding deviations

Current storage resources in Spain

Storage resources are provided for ATLAS, CMS, and LHCb communities, both local and remote users

Spain provides **~4% of Tier1** resources and **~3% of Tier2** resources, worldwide





2021

~15 PB Disk
~20 PB Tape

Current storage resources in Spain

Complex data management systems deployed at sites, operated at higher levels of reliability

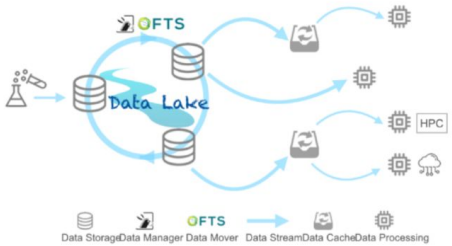
For example @IFIC:

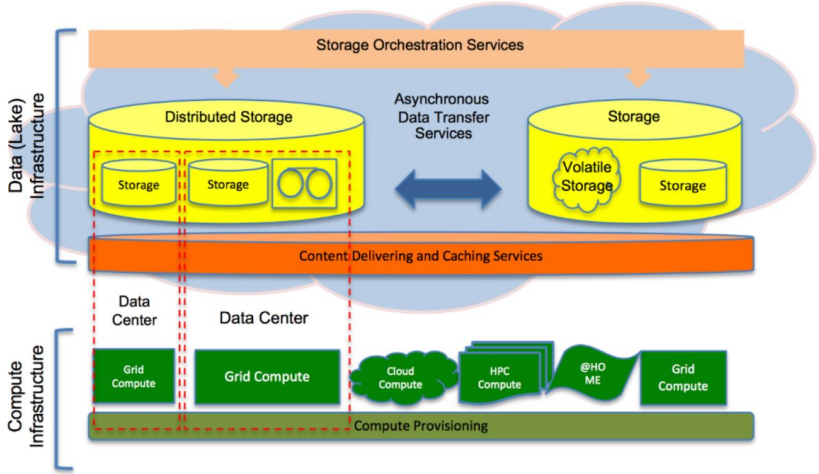
Link to talk

Moving to a network-centric model

Datalake model:

Fewer number of facilities operating storage services, less data replication





CPUs and storage not necessarily co-located: need to deliver the content over the WAN and/or cache it

Moving to a network-centric model

Definition of “**datalake**” in our context: an infrastructure where CPU and storage capacity are loosely coupled (not necessarily co-located)

- Storages in the datalake needs to be connected by a **fast and reliable network**
- Storage might offer **different QoS classes**
- A **content delivery system** might help serving data from storage to CPUs

Mixed statements and experiences from the experiments about **caching**, **latency hiding** and **remote data access** (work in progress to draw solid conclusions)

All experiments declare interest in **QoS** but today this is **prototyped very little** (except for disk/tape)

Regional datalake implementation ideas have been presented at DOMA meetings. Focus in 2021 is to prototype those ideas. DOMA ACCESS and QoS WGs merged into a **DOMA Datalakes WG**

Moving to a network-centric model

I Workshop de Computing y Software de la Red Española de LHC

28 Apr 2021, 09:00 → 29 Apr 2021, 16:00 Europe/Zurich

Andreu Pacheco Pages (Institut de Física d'Altes Energies - Barcelona (ES)), Jose Hernandez (CIEMAT), Jose Salt (Univ. of Valencia and CSIC (ES)), Luca Fiorini (Univ. of Valencia and CSIC (ES))

09:55 **A Spanish data cache service for the CMS experiment**

Speaker: Jose Flix Molina (Centro de Investigaciones Energéti cas Medioambientales y Tecno)

XCache: test instances (OSG repo)

In CIEMAT (XRootD 4.11):

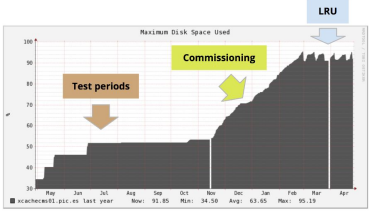
→ 22 TB, with RAID10, 8 cores, 8 GB RAM

→ Used by 2 Worker Nodes for all data tiers but only for TFC fall-back (data is not at CIEMAT)

In PIC (XRootD 5.1.1-1.3):

→ Distributed 4TBx36 [130 TB], no-RAID, 16 cores L5630 (HT enabled), 48 GB RAM, 10 Gbps - 90-95% occupancy

→ attached to one unique Worker Node that "caches all data tiers" as of today



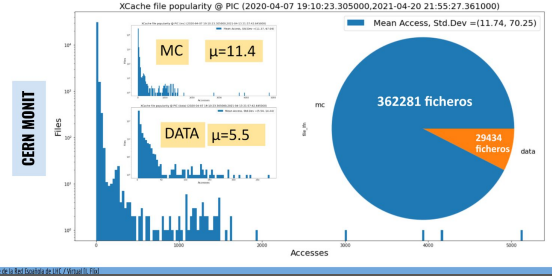
150 TB of data copied in the period

The XCache applies LRU algorithm to flush the cache (95% → 90%)

XCache: current status

Monitors and **proper settings** being deployed and/or investigated

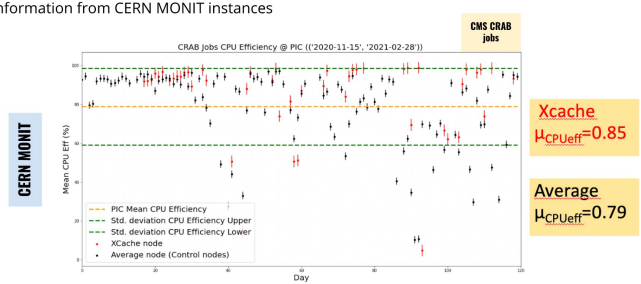
- Local monitor (based on XRootD cinfo files)
- Information from CERN MONIT instances



XCache: current status

Monitors and **proper settings** being deployed and/or investigated

- Local monitor (based on XRootD cinfo files)
- Information from CERN MONIT instances



Moving to a network-centric model

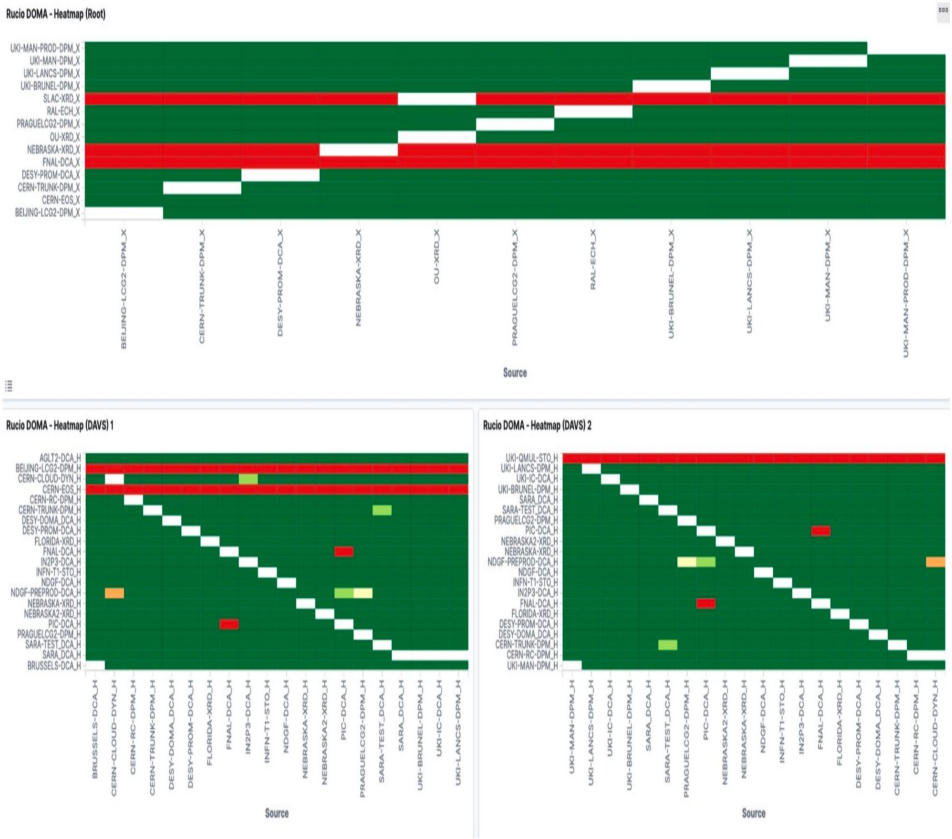
The decisions about storage deployment in a region will impact the **network**

- Distributed storage or storage-less sites will require more network
- Deployment of caches might reduce the network needs, depending on the use cases and the workflows

Defining regional plans for storage and the corresponding network needs will be the obvious next step. The two aspects should be discussed in the same context

Third Party Copies

HTTP and XRootD protocol testbeds for TPC



Third Party Copies

It is agreed to consider **HTTP as the WLCG baseline protocol for TPC**. Every storage solution should implement it and every site should deploy it

- Alice has its own scenario based on xrootd protocol which already works in production at the ALICE sites

The timescale is tight: it is desirable to be gridFTP-free by end of 2021 (linked to the Globus decommissioning)

Need to **make it work at the required scale** across the WLCG infrastructure. Will be followed up in the DOMA TPC task force and WLCG Ops

- Data challenges in Q3 2021 to convince ourselves we can replace gridFTP with HTTP in full scale production

Authentication & Authorisation Infrastructure (AAI)

Two aspects here (related but not the same):

- Decommissioning of Globus (in OSG by end of 2021)
- Wish to progress toward x509-free infrastructure (toward **token-based AAI**)

Globus-free does not imply X509-free (e.g. dCache supports X509 w/o Globus)

Very hard to have the whole of WLCG independent from Globus in ~1 year, while it is well possible for OSG. Need to interoperate

Fully X509-free WLCG is a LS3 (tight) target - **testbeds** in place for token-based TPC

- Tokens and X509 need to interoperate in the next years

WLCG Ops Coord and OSG have drafted a **multi-year plan**, focusing initially into the Globus decommissioning. It is being followed up at the GDB:

→ https://twiki.cern.ch/twiki/bin/view/LCG/WLCGTokensGlobusWG

Archive storage

Tape Storage:

- It is not a ‘cold storage’ for LHC
- Three frontend solutions in WLCG: EOS/CTA, dCache, StoRM
- In the short term **SRM** will continue playing a role. FTS should hide the complexity of “stage+transfer” via SRM(dCache,StoRM) or xrootd(CTA)+HTTP
- In the medium term, harmonise the tape access through a **common REST API**. The dCache bulk request API seems a good way to standardize - being followed up in DOMA general meeting /presented at GDB

Archive storage does not need to be tape

- Disk-based solution presented by KISTI
- Storage TCO needs to be considered, particularly if the usage will increase (e.g. tape carousels) - BNL studies
- The Archive Storage working group has been revamped - pre-GDB organized

Archive storage

Expansion of the new Tape Library

PIC port d'informació científica

NEW

IBM TS4500: 2 frames (L55+D55) + 8 LT08 drives
- → 4.8 PB capacity installed with cartridges LT07 M8
- → 750 TB capacity installed with cartridges LT08

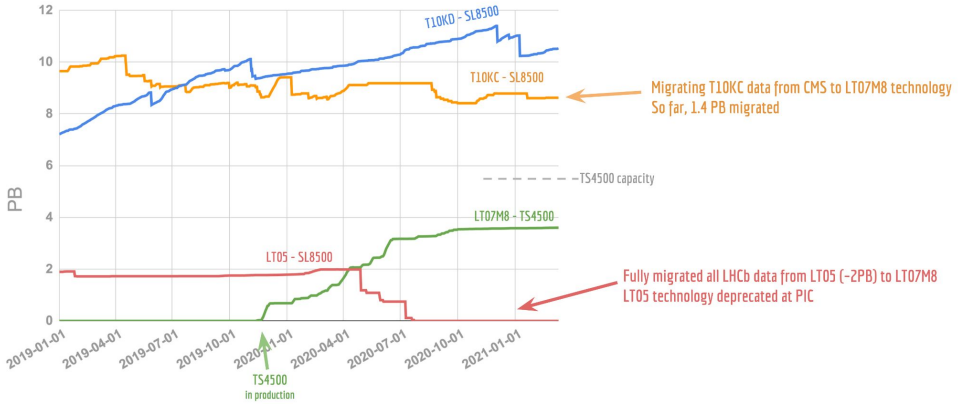
This library is expected to grow to host future data

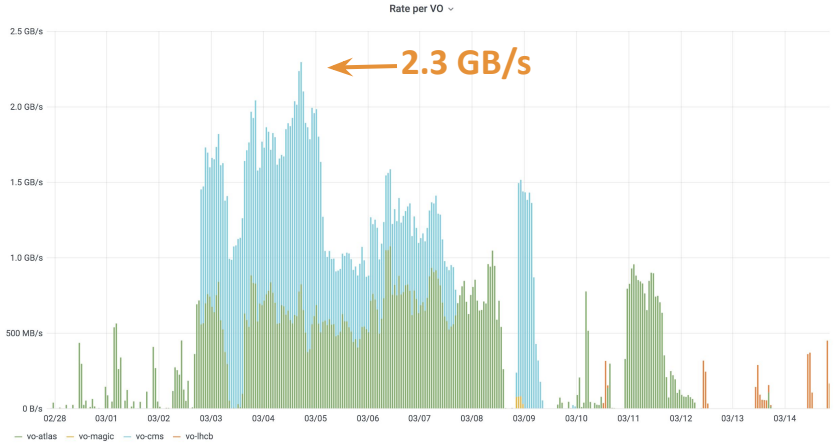
- → It will host new data and data migrated from SL8500 library (ongoing)
- → Dedicated drives, frames and cartridges installed to handle this

All new **CMS**, **LHCb** and **MAGIC** data go to the IBM

PIC currently runs **Enstore 6.3.4-2** (CentOS7)

All LHC data (20 PB) being migrated to the new system (for the next 2 years)





Recent massive read/write tests from tape @PIC for both ATLAS and CMS

Erasure Coding

There is a growing interest in the community towards **Erasure Coding***

- Avoids running RAID5/6 arrays. Error recovery in RAID5/6 becomes more problematic as size of storage and disks increases
- Avoids running JBOD w/o redundancy. No site is willing to run w/o redundancy today because of the operational effort
- Avoids full mirroring in JBOD (raw storage = 2 * available storage)

Some storage (e.g. **EOS**) can provide EC on top of JBOD directly. Most can also rely on an dedicated ”external component (e.g. **CEPH**) to provide EC.

Some experience in the community running EC on storage backends and many “warnings” were highlighted: possible internal network congestion, larger frontend memory needs, ...

A **HEPIX WG was created** on this topic: share ideas, experience, ...

* Erasure coding (EC) is a method of data protection in which data is broken into fragments, expanded and encoded with redundant data pieces and stored across a set of different locations or storage media

Reducing Operational effort

Several ideas in DOMA aim at **reducing the effort needed in operations**

- The infrastructure and the experiment activities will grow in size and complexity. The available effort in operations probably will not
- Rather than reducing effort, the aim is to **operate at a large scale with the same effort** → we will need to be **more efficient**

These new ideas have not been modeled yer very well

The impact of many new ideas in DOMA on the operational effort needs to be understood further

For the time being we should continue sharing the experience of sites prototyping different solutions and explaining where they saw a benefit (or not)

Reducing Operational effort

Handling of **data loss** is still a labor intensive task

- Sites need to collect the list from the storage and re-create consistency inside the storage itself (disk/namespace). This is storage technology specific, but not site specific
- The information needs to be transmitted to the experiments. No common way of doing this at the moment
- The experiments need to consume the information and manage the loss (re-replicate, re-create, ignore...)

Any chance to share tools and have some level of automation there?

Periodic **consistency checks** between storage and experiment catalogs are needed

- No common way to expose storage content (list of files). Harmonise?

It is agreed WLCG operations coordination will look into opportunities in those areas

- Check operational intelligence from Panos at the GDB:
  → https://indico.cern.ch/event/876788/

Special Facilities

Analysis facilities might focus on distinct aspects **[details at the next talk]**

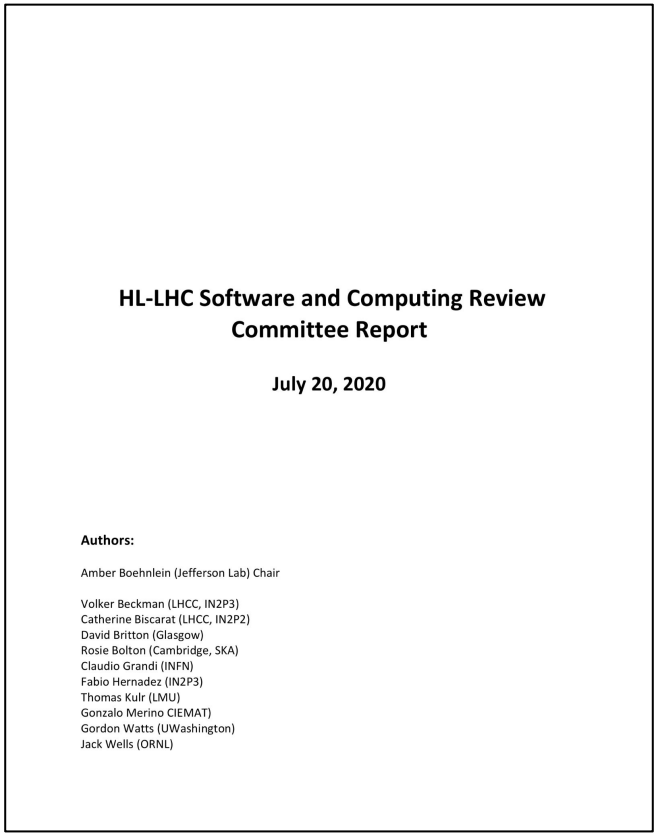
- Fast turnaround time (e.g. Alice GSI facility, fully fledged Grid site)
- User friendly access to data (no Grid services except for e.g. xCache)

Both ideas are being prototyped and the experience should be shared

HPCs present the known challenges related to data access **[see today's short-talk]**

- Many times the challenges are of non technical nature and can be addressed by the good collaboration between the HPC center and the WLCG community (e.g. the CNAF-CINECA integration)
- Tsocks or HTCondor communications via a shared file systems seem interesting solutions to overcome issues with external connectivity

LHCC HL-LHC SW & Computing review



HL-LHC Software and Computing Review
Committee Report

July 20, 2020

Authors:

Amber Boehnlein (Jefferson Lab) Chair

Volker Beckman (LHCC, IN2P3)
Catherine Biscarat (LHCC, IN2P2)
David Britton (Glasgow)
Rosie Bolton (Cambridge, SKA)
Claudio Grandi (INFN)
Fabio Hernadez (IN2P3)
Thomas Kulr (LMU)
Gonzalo Merino CIEMAT)
Gordon Watts (UWashington)
Jack Wells (ORNL)

LHCC HL-LHC SW & Computing review

First episode (May 19th 2020) of a process leading to the HL-LHC computing TDR(s) in ~2025
- Owned by the LHCC

5 documents submitted: ATLAS, CMS, WLCG, **DOMA**, software

Focus on the **R&D plans** to:
- address the resource needs in a flat budget scenario
- ensure long term sustainability of tools, services and infrastructure for HL-LHC and beyond

The review report is publicly available here:
→ https://cds.cern.ch/record/2725487/files/LHCC-G-177.pdf

A promising and **prioritized list of R&D activities** was presented. For the next review a more formal assessment of the gains is expected

LHCC HL-LHC SW & Computing review

Storage still emerges as the main HL-LHC resource challenge. Many recommendations for ATLAS and CMS go in this direction:

- **Data carousel**: Data processing (reco) through buffers, orchestrating a workflow with tape carrousel, where a quantity of data is taken from the tape to the buffer, processed, erased and then more data is taken from the tape. Implications on sites (so, engage them further) quantify savings, think about mitigations for potential risks (e.g. unclear future of tape)
- Assess how access to **analysis data through caches** might help reducing storage needs
- Full adoption of **reduced data formats** (MiniAOD/DAOD_PHYS, NanoAOD/DAOD_PHYSLITE) → ~kB/event
- Elaborate on the model of **“data parking”** to balance the cost over the years

The committee encourages for a better definition of **LHC running conditions** for HL-LHC (2028-2030)

WLCG should leverage on the collaboration with **EGI** and **OSG**, and **coordination with other sciences** with similar approaches

LHCC HL-LHC SW & Computing review

For **DOMA**: focus in defining milestones, quantifying the potential reductions in cost (equipment and manpower), mitigate risks

Quantify the **potential gains** and **evaluate risks of the data-lake model**: implications of the specialized role of sites, impact of content delivery networks on cost and efficiency, person-power needs for operations

Testing networks in real scenarios

Therefore, the **DOMA Data Challenges** (network, datalakes) will be the key in the next couple of years

Sites and experiments are expected continue remaining fully engaged in those activities, in preparation for the next phase of the HL-LHC computing review

Data Challenges towards HL-LHC

As part of the planning for HL-LHC **Data Challenges** are foreseen from this year every 2 years until the start of HL-LHC

Increasing transfer rates until we reach the required level for HL-LHC → **2021, 2023, 2025, 2027**

The aim is to **prepare** WLCG infrastructure and **demonstrate** that the bandwidth is used effectively

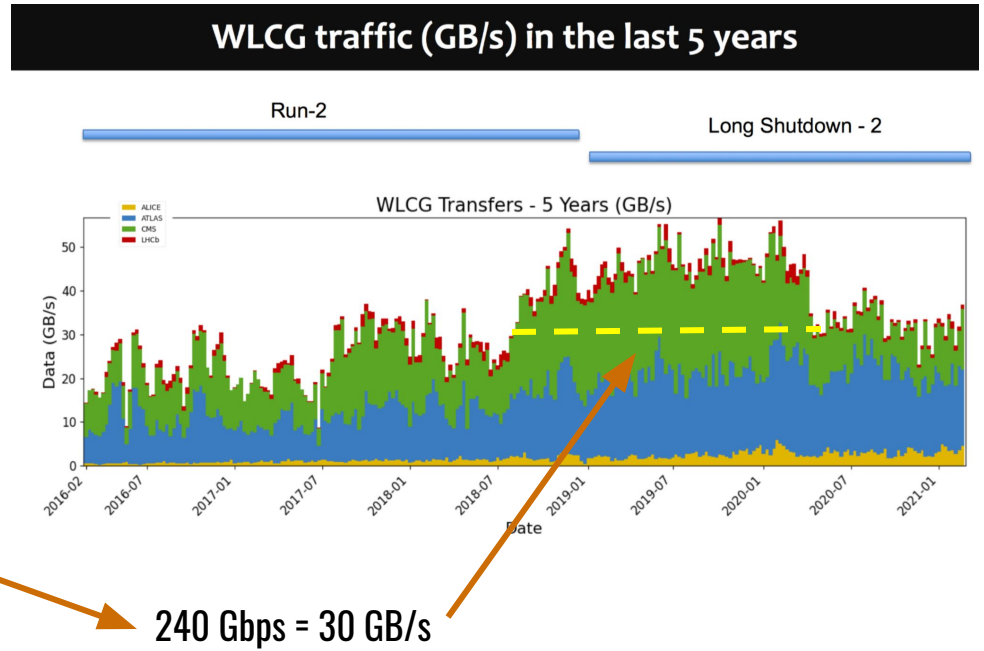
The Data Challenges should follow and drive the **expansion of the network capacity**

All 4 experiments in parallel and **all sites**, using experiments production infrastructure, **in parallel with experiments activities**

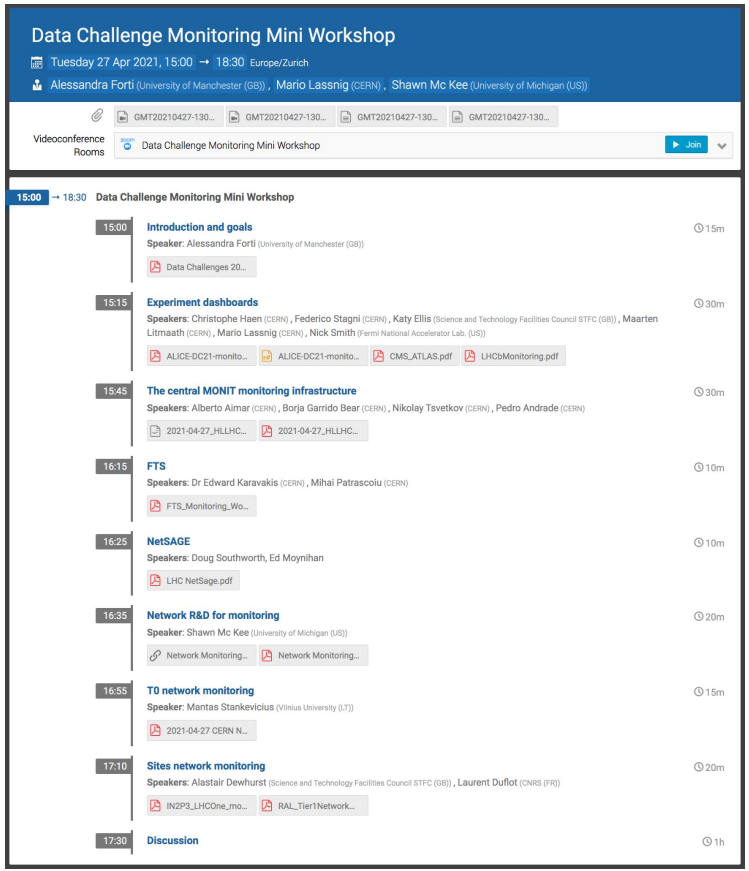
Data Challenges are by definition a **common effort** that requires the involvement of experiments, network people, monitoring and sites - **How to monitor these activities is a crucial part**

Data Challenges towards HL-LHC

T1	Data Challenge target 2027 (Gbps)	Data Challenge target 2025 (Gbps)	Data Challenge target 2023 (Gbps)	Data Challenge target 2021 (Gbps)
CA-TRIUMF	98	59	29	10
DE-KIT	312	187	94	31
ES-PIC	93	56	28	9
FR-CCIN2P3	281	169	84	28
IT-INFN-CNAF	336	202	101	34
KR-KISTI-GSDC	25	15	7	2
NDGF	71	43	21	7
NL-T1	94	56	28	9
NRC-KI-T1	62	37	19	6
UK-T1-RAL	296	177	89	30
RU-JINR-T1	52	31	15	5
US-T1-BNL	227	136	68	23
US-FNAL-CMS	454	273	136	45
(atlantic link)	681	408	204	68
Sum	2400	1440	720	240



Data Challenges towards HL-LHC



Data Challenge Monitoring Mini Workshop

Tuesday 27 Apr 2021, 15:00 → 18:30 Europe/Zurich

Alessandra Forti (University of Manchester (GB)) , Mario Lassnig (CERN) , Shawn Mc Kee (University of Michigan (US))

GMT20210427-130... GMT20210427-130... GMT20210427-130... GMT20210427-130...

Videoconference Rooms: Data Challenge Monitoring Mini Workshop — Join

15:00 → 18:30 Data Challenge Monitoring Mini Workshop

15:00 **Introduction and goals** — 15m
Speaker: Alessandra Forti (University of Manchester (GB))
Data Challenges 20...

15:15 **Experiment dashboards** — 30m
Speakers: Christophe Haen (CERN) , Federico Stagni (CERN) , Katy Ellis (Science and Technology Facilities Council STFC (GB)) , Maarten Litmaath (CERN) , Mario Lassnig (CERN) , Nick Smith (Fermi National Accelerator Lab. (US))
ALICE-DC21-monito... ALICE-DC21-monito... CMS_ATLAS.pdf LHCbMonitoring.pdf

15:45 **The central MONIT monitoring infrastructure** — 30m
Speakers: Alberto Aimar (CERN) , Borja Garrido Bear (CERN) , Nikolay Tsvetkov (CERN) , Pedro Andrade (CERN)
2021-04-27_HLLHC... 2021-04-27_HLLHC...

16:15 **FTS** — 10m
Speakers: Dr Edward Karavakis (CERN) , Mihai Patrascoiu (CERN)
FTS_Monitoring_Wo...

16:25 **NetSAGE** — 10m
Speakers: Doug Southworth, Ed Moynihan
LHC NetSage.pdf

16:35 **Network R&D for monitoring** — 20m
Speaker: Shawn Mc Kee (University of Michigan (US))
Network Monitoring... Network Monitoring...

16:55 **T0 network monitoring** — 15m
Speaker: Mantas Stankevicius (Vilnius University (LT))
2021-04-27 CERN N...

17:10 **Sites network monitoring** — 20m
Speakers: Alastair Dewhurst (Science and Technology Facilities Council STFC (GB)) , Laurent Duflot (CNRS (FR))
IN2P3_LHCOne_mo... RAL_Tier1Network...

17:30 **Discussion** — 1h

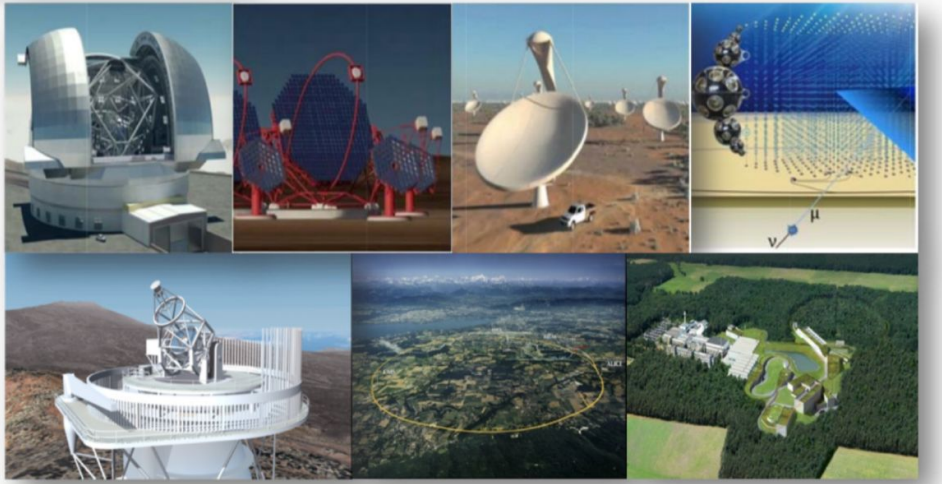
Future Data Challenges Workshops to happen

Toward a sustainable, open and shared infrastructure

ESCAPE
European Science Cluster of Astronomy & Particle physics
ESFRI research infrastructures

Horizon 2020
funded project

EUROPEAN OPEN SCIENCE CLOUD



Goals:

Prototype an infrastructure adapted to the Exabyte-scale needs of the large science projects.

Driven by the sciences

Address *FAIR* data management

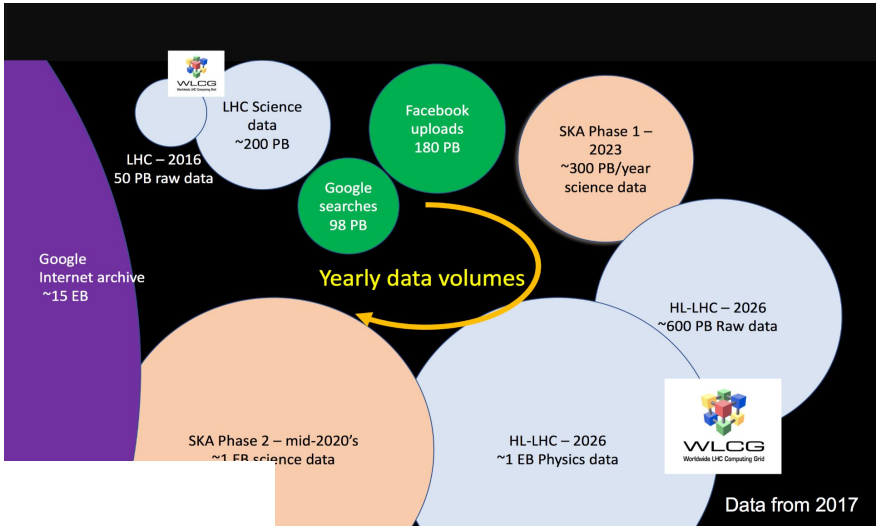
Science Projects

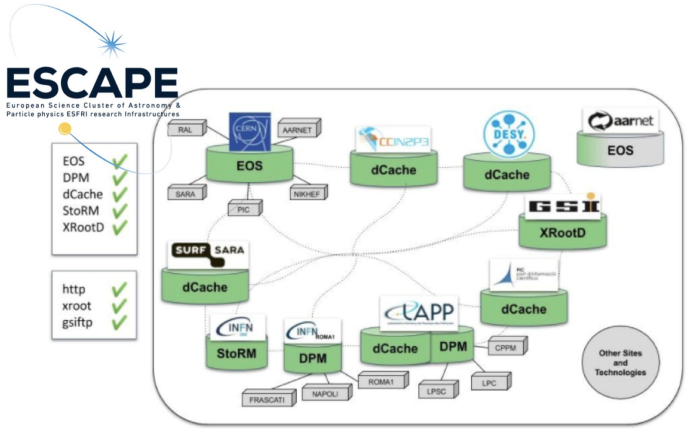
HL-LHC	SKA
FAIR	CTA
KM3Net	JIVE-ERIC
ELT	EST
EURO-VO	EGO-VIRGO
(LSST)	(CERN,ESO)

Data centres: CERN, INFN, DESY, GSI, Nikhef, SURFSara, RUG, CCIN2P3, PIC, LAPP, INAF

Toward a sustainable, open and shared infrastructure

HPCs: CERN, SKA, PRACE and GEANT signed a collaboration agreement in July 2020. Complement the work done by the experiments





e.g.

cta cherenkov telescope array

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

WLCG

Conclusions

LHC Computing models together with the infrastructure and services changed with time adapting to the evolving landscape (experience and funding). The HL-LHC will be an **unprecedented challenge** for us in terms of scale and sustainability, in particular in the Data Management arena

Various **R&D projects to prototype/test** such a data management infrastructure & associated tools

- Reduce the global cost of storage (hw and operations)
- Enable a more effective use of existing storage
- Efficient and scalable data delivery to large, remote, heterogeneous, compute resources

Building a **common set of DM tools** based on open and standard protocols that can be used by a broad set of scientific experiments

Evolution of the **AAI towards federated identities and token-based systems** in line with most modern network services

Thoughts for discussion

In Spain we have ~15 years experience in providing computing resources and services to WLCG - very **satisfactory results** in terms of availability and performance

However, we need to be **proactive in these new R&D DM activities**, since they are based on a very intense R&D program: we need to draft a sustainable model in Spain towards HL-LHC for Data Management → it implies an evolution of the infrastructure and network

- Storage consolidation, federation, inclusion of caches, interoperability...

We do **not foresee an increase in funding** for LHC Computing in the next years in Spain. The RES data can help us to get additional resources and this would need to be explored more

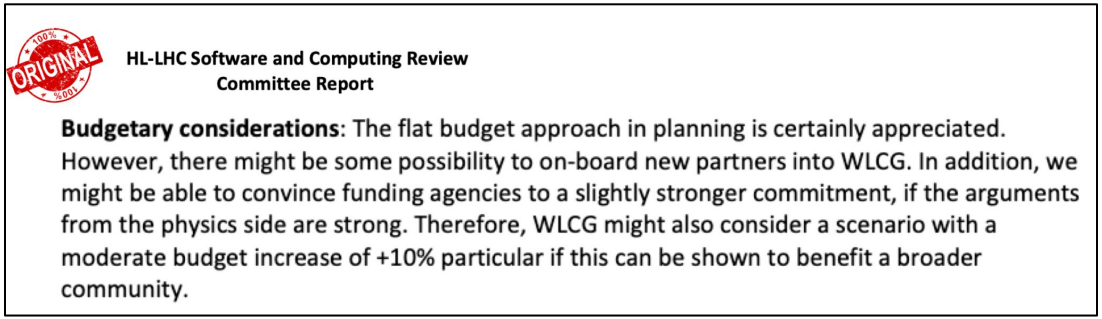
- EU-projects are useful to develop new mechanisms and tools (but not for resources)
- Getting additional resources from Spanish infrastructure calls (co-funded)
- We cannot get/rely on opportunistic disk usage!

LHCb has a step up on resources during Run3 → how to cope with it with tight funding in Spain?

Thoughts for discussion

We need to strength to our funding agency that the investments they make in LHC Computing has a positive **impact on other sciences** (articulated by ESCAPE, ARCHIVER, EOSC, ...)

One of the outcomes of the **HL-LHC Software and Computing review** focus on:



WLCG to start this discussion with the Funding Agencies. It is important to demonstrate that there is a benefit for a broader set of communities ← We should get organized with the next HEP national program manager **[community - WLCG mgt - agency]**

Thoughts for discussion

Last but not least, we need to **engage young people** to join our activities. The LH-LHC challenge presents an opportunity and opens the door for many activities and a career perspective

We need to explore all of the possibilities to get funds for **PhD students** to join our activities, and in particular enroll them in the Data Management challenges

Thanks!

Questions?