



ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

ESCAPE DIOS: towards the prototype

Xavier Espinal (CERN)


ESCAPE WP2/DIOS Workshop, 10 December 2020



Work Package Deliverables



Deliverables

Nr	Description (type)	Task	Lead participant	Month
D2.1	Implementation plan and design of pilot; R&D questions/metrics that will be addressed in the pilot and prototype. (R)	2.1, 2.2, 2.3, 2.4, 2.5, 2.6	CERN	8 
D2.2	Assessment and analysis of the performance of the first pilot data lake (R)	2.1, 2.2, 2.6	SKAO	24
D2.3	Final assessment and analysis of the full prototype, outlook for further development and deployment towards full production services within EOSC (R)	2.1, 2.2, 2.4, 2.6	CERN	40

Current main focus



Project Title: European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructure

Project Acronym: ESCAPE

Grant Agreement No: 824064

Instrument: Research and Innovation Action (RIA)

Topic: Connecting ESFRI infrastructures through Cluster projects (INFRA-EOSC-4-2018)

Start Date of Project: 01.02.2019

Duration of Project: 42 Months

Project Website: www.projectescape.eu

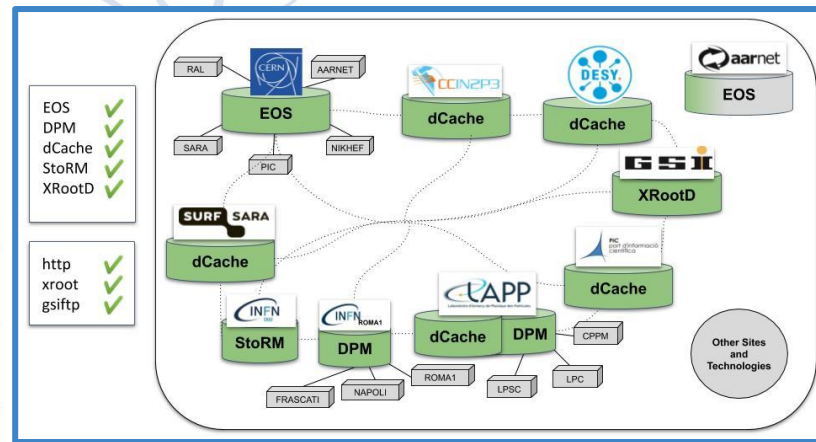
D2.1 - IMPLEMENTATION PLAN AND DESIGN OF PILOT



Work Package Milestones

- Well ahead of M2.2, functional pilot data lake with 10 storage endpoints

M2.1	✓	First WP2 workshop on the initial design and goals of the first pilot data lake, prepare D2.1	WP2	M6	Workshop summary report
M2.2	✓	Initial pilot data lake with at least 3 core data centres	WP2	M18	Progress report; Active monitoring of activity (web site)
M2.3	✓	Second WP2 workshop to analyse the performance of the pilot, prepare D2.2	WP2	M22	Workshop summary report
M2.4	✓	Expanded prototype – more data centres including 3rd party centres, demonstrate integrated data management tools, verify RI data accessibility from compute platforms including commercial clouds	WP2	M24	Review of D2.2; Monitoring web site
M2.5		Extension of the data lake to efficiently serve data to external compute resources providers	WP2	M30	Progress report; Monitoring web site
M2.6		ISO 16363 certification process underway in core data centres	WP2	M32	Progress report; core data centres finished self-certification audit and ready to submit to external audit.
M2.7		Third WP2 workshop to review performance of the full prototypes, and to explore future directions, prepare D2.3	WP2	M38	Workshop summary report



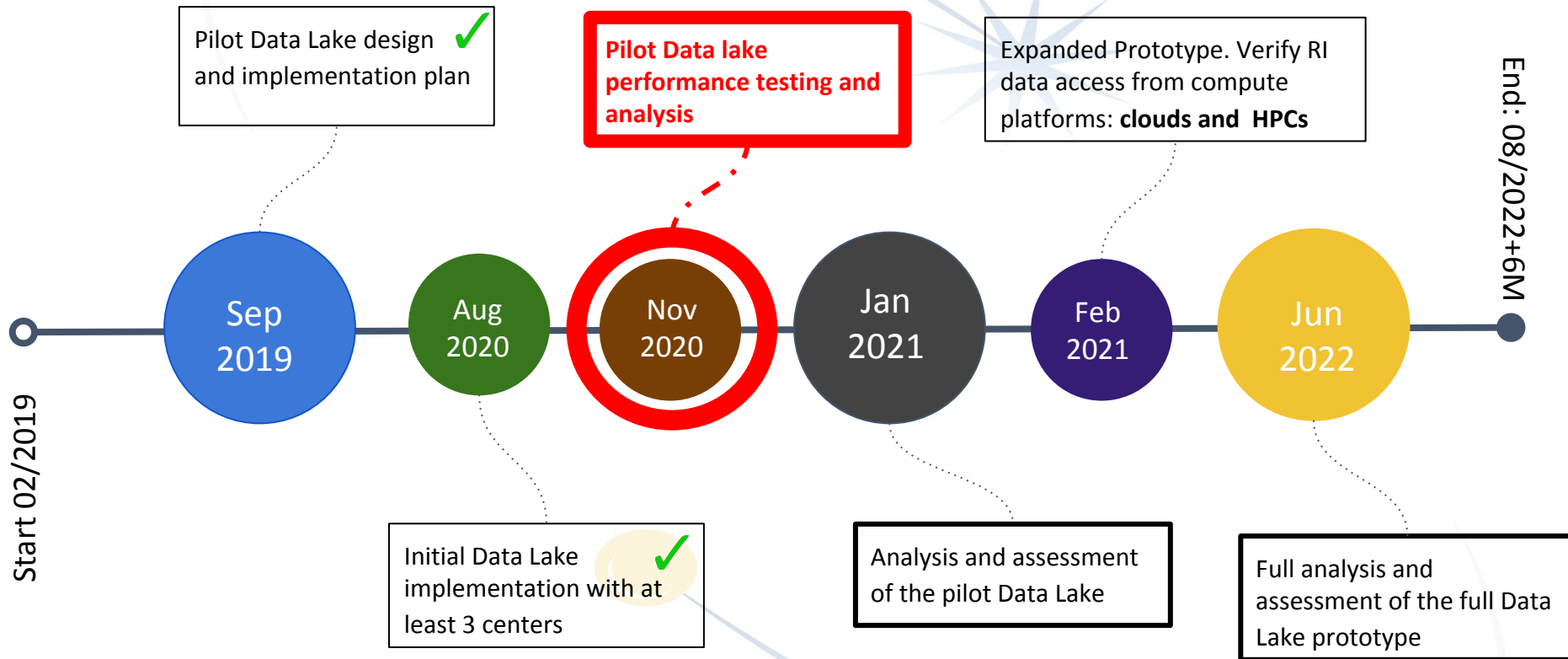
Successful Transfers Percentage (davs) ↓

Source \ Destination	ccdcalltest10.in2p3.fr	dcache-door-doma01.desy.de	dclxwp2dids1.gsi.de	door05.pic.es	eoseulake.cern.ch	lapp-dcache01.in2p3.fr
ccdcalltest10.in2p3.fr	-	100.00%	99.76%	100.00%	100.00%	99.95%
dcache-door-doma01.desy.de	100.00%	-	99.69%	99.99%	100.00%	99.48%
dclxwp2dids1.gsi.de	99.24%	99.55%	-	99.19%	73.65%	98.94%
door05.pic.es	99.97%	100.00%	99.69%	-	100.00%	99.90%
eoseulake.cern.ch	100.00%	100.00%	98.97%	100.00%	-	99.87%
lapp-dcache01.in2p3.fr	99.69%	99.63%	98.89%	99.53%	99.94%	-
lapp-esc02.in2p3.fr	100.00%	99.82%	99.81%	100.00%	100.00%	100.00%
lapp-test01.in2p3.fr	96.16%	96.17%	94.39%	96.35%	94.88%	95.16%
t2-dpm-dome.na.infn.it	97.44%	100.00%	91.01%	98.18%	92.86%	96.67%
webdav.grid.sara.nl	100.00%	100.00%	99.85%	100.00%	100.00%	99.79%
xfer.cr.cnaf.infn.it	100.00%	100.00%	78.27%	100.00%	95.31%	99.96%

Already Partially achieved



Summary WP2 Roadmap



Start 02/2019

End: 08/2022+6M





Risk review - from the GA

Performant and scalable data lake not being able to serve multiple RIs simultaneously. (Low)	WP2	Build on many years' experience in WLCG in data and storage federations, with existing and demonstrated solutions, integrating these with a coordination and service layer.
Lack of maturity and uptake of underlying technologies required; such as European-wide AAI services. (Low)	WP2	Collaborate with and set requirements for ongoing actions and projects: for example, AARC, EOSC-hub, EOSCpilot, GEANT etc.
Difficulty in integrating ad-hoc storage solutions at data centres into a coherent federated service. (Medium)	WP2	The core data centres selected bring several different and widely used storage solutions to the project; previous expertise of the partners (CERN, DESY, INFN) in building large-scale storage will address this.
Access to external HPC and cloud resources to support and validate data-heavy workflows is essential, missing HPC and cloud environments to support exascale data processing capabilities. (Low)	WP2	Collaborate strongly with PRACE for HPC, HNSciCloud partners for cloud, and GEANT for data access.

Working pilot data lake in place, strong indications it can serve multiple RIs

Successful interactions and experience: common AAI in production (indigo IAM)

Different storage technologies with a variety of storage backends have been successfully integrated

Ongoing collaboration with the GEANT, PRACE, CERN, SKA collaboration agreement on HPC. Participated in the [kick-off workshop Sep-2020](#)





Risk review - new possible risks

- **Risk:** Difficulty to endeavour partner data centers into the certification as trustworthy digital repositories. ISO16363: long term data preservation (low risk)
- **Mitigation:** Collaboration ongoing with ARCHIVER project, join efforts, some common stakeholders in both projects (CERN, DESY and PIC). This does not compromise the project as it is a data center specific certification

- **Risk:** Difficulty to procure and commission external compute resources for a meaningful data lake access demonstrator (low risk)
- **Mitigation:** Joint proposal with an ICT call was rejected, keep an eye on possible cooperation with current/new EU projects, otherwise start a dedicated task during the next period





Next steps

- Pilot datalake **assessment exercise** (November/December)
 - Very successful Full Dress Rehearsal on the 17th Nov M2.3 (M22)
 - **Second WP2/DIOS Workshop** 8th and 9th December
- Deliverable D2.2 (M24): **Performance assessment and analysis** of the pilot datalake
- From M24: Pilot datalake **evolution towards the prototype** phase
 - Increase storage size, allowing to scale-up data ingestion and throughput challenges
 - Demonstrate data access from compute platforms (including commercial clouds), evaluating the possibility to include HPCs
 - Trustworthy digital repositories for long term archive (ISO16363 certification)
 - WP5/ESAP intensify collaboration focusing to prepare full workflows (planing a mini-workshop)
- A very rich work program for 2021:
 - Collaborating with European Projects: ARCHIVER, CS2MESH4EOSC, (EOSC-Future)
 - Engagement with Fellow Cluster Projects: PANOSC /EXPAND
 - And technical challenges: Tokens, Embargo data, FTS/network throttling multi-VO, RUCIO event notification mechanism, Downtime/ ticketing, TSP, Immediate use-cases (EGO)...





Summary

- For the first time flagship ESFRIs are collaborating together since the real beginning to implement the standards of data FAIRness and Open Access in a common scientific data management infrastructure
- Breaking historic sociological barriers among different scientific communities in the search for commonalities in their computing models and potential synergies
- A very encouraging scientific atmosphere. The engagement from from **ALL OF YOU** have been overwhelming. We should be proud of what we achieved so far and use this momentum for the 2nd phase of the project. I am convinced we will deliver value that will persist after the end of the project.



