



Our Digital Future Multidisciplinary Perspectives on Long Term Data Preservation and Access

Data Preservation project at the LHCb experiment at CERN

ANA TRISOVIC TRISOVIC@HEP.PHY.CAM.AC.UK 15 MARCH 2016 CAMBRIDGE

Agenda

LHCb experiment at CERN

Data preservation project and motivation

CERN Analysis Preservation portal

LHCb data and software

Data dependency database

LHCb experiment at CERN

One of the four particle detectors at the Large Hadron Collider at CERN

Investigate asymmetry between matter and antimatter

840 people from 60 scientific institutes



Data Preservation project

- The experimental & simulated data
- Software & documentation
- Analysis & publications

Motivation for Data Preservation

- Research reproducibility reanalysing data in search of new signals or to improve current measurements.
- 2. Scientific potential Analyse old data to search for a signal predicted by a new theory.
- 3. Social reasons CERN is funded by the world community; therefore data should be preserved and made available to general public.

Different approach for short term and long term data preservation







Web portal for physicists, fellows, interns, doctoral and summer students to log information about their analyses

 Input data, n-tuples, code, papers, other publication, peer review and Q&A

Team: Sünje Dallmeier-Tiessen, Anxhela Dani, Tibor Sinko, Javier D. Fernandez, Pamfilos Fokianos, Patricia S. Herterich

Adapta Name Service Serv			Basic Info JSC	DN Permissions			
	Basic Information						
Addition in Production Construction Con	Ana	alysis Name	CPV in D0 -> KS KS				
Of V & D(p) - St S	Analysis Number CPV in D+ -> phi pi and Ds+ -> KSpi decays						
Distancian CV/L RD > 25 B(5 p) Sales a migring like And indice displayed bit Sales a migring like Date Sales a migring like Sales a migring Sales a migring <t< td=""><td></td><td></td><td>CPV in D(s)+ -> KS h</td><td>,</td><td></td><td></td></t<>			CPV in D(s)+ -> KS h	,			
Auforde analysis of 0 > 2.5 ° µ Solid cating long long Auforde analysis of 0 > 2.5 ° µ Note of additional data with the data wit	DST selection		CPV in D0 -> KS KS T-odd moments in D+ -> KS K ni ni				
Bob c antiping like Second antiping like Descond antiping like Togor Intervent Intervent Bibly Dia Second antiping like Second antiping like Bibly Dia Second antiping like Intervent Bibly Dia Second antiping like Second antiping like Bibly Dia Second antiping li			Amplitude analysis of D0 -> KS K pi				
Langer de la construit de la c	Select a stripping line	dantan Una					
Inger Dag 2.454/second Imger//Indice.cem.di/eeord/301842/opdil Keyer Dag Widen on Charm WG, Mixing and CP violation Charm WG, Mixing and CP violation Code Widen on 2.556/second Mixing//Indice.cem.di/eeord/311842/opdil Kayer on Widen on 2.556/second Charm WG, Mixing and CP violation Low Dag 2.566/second Charm WG, Mixing and CP violation Charm WG, Mixing and CP violation Migner//Indice.cem.di/eeord/31182/contribution/1/individ/ildide/0.pdf Charm WG, Mixing and CP violation Charm WG, Mixing and CP violation Migner//Indice.cem.di/eeord/31182/contribution/1/individ/ildide/0.pdf Charm WG, Mixing and CP violation Charm WG, Mixing and CP violation Low Dag Low Dag Low Dag Low Dag Charm WG, Mixing and CP violation Low Dag Mixing Charm WG, Mixing and CP violation Charm WG, Mixing and CP violation Low Dag Mixing Charm WG, Mixing and CP violation Charm WG, Mixing and CP violation Low Dag Mixing Charm WG, Mixing and CP violation Charm WG, Mixing and CP violation Low Dag Mixing Charm WG, Mixing and CP violation Charm WG, Mixing and CP violation Low Dag Mixing	5m				Documentations		
Imper dam Dem # & & & & & & & & & & & & & & & & & & &		Ingger					
AC Des AC		Inpur Dara	Data + Add	New Item		URL https://indice.com.ch/event/361842/contribution/3/material/slides/0.ndf	
Code Plafam UKC code Plafam UKC code Plafam UKC code Plafam UKC code Plafam User code Plafam UKC code Plafam User code Sconton User code Plafam User code P			MC Data + Add New Item			Keyword classical and the second states of the seco	
Code Fadom UCb code Add how ten UCb code Add how ten UCb code Lossienin Code Refer UCb code Add how ten Ucb cod						Comment	
High Date High Date High Date Date Listering Listering		Code	Platform			Start of WG review: time-integrated CP asymmetry in D0->KSKS decays	
Use code Add Now boar Use code Add Now boar Date Location MC Date Location Location Use code MC Date Location Use code Reference Location Use code MC Date Location Use code Reference Location Use code MC Date Add Now brain Use code Reference Location Use code MC Date Date MC Date MC Date MC Date MC Date MC Date MC Date MC Date Add Now brain Lich Date Add Now brain Lich Date Add Now brain <td></td> <td>LHCb code + Add</td> <td>New Item</td> <td></td> <td></td>			LHCb code + Add	New Item			
Verified of work work Werk work work Level Date Level Date Level Date Level Date Licktie werk Licktie werk <td></td> <td></td> <td>lleer ende</td> <td></td> <td></td> <td></td>			lleer ende				
Keyword			User code + Add	New Item		URL https://indico.cern.ch/event/311253/contribution/1/material/slides/0.pdf	
Inspire dots Dots Location MC Date Location Location URL Location URL Internal Discussions URL https://twiki.cem.dh/wiki/bin/view/LHCbPhysic/DDKSKS Code Reform LiStie out Add Now Inter Output Date Date Acc Date Date Acc Date Add Now Inter URL https://cds.cem.ch/record/2053739 URL https://cds.cem.ch/record/2037647 URL https://cds.cem.ch/record/2037647 URL https://cds.cem.ch/record/2037647		Input D	ta .			Keyword Charm WG, Mixing and CP violation	
MC Date Location Location MR Date Location URL https://twiki.cem.d/twiki/bin/view/LHCbPhysics/DDKSKS Code Paterne LiCkb code > Add Nave Ins Location URL https://cds.cem.d/twiki/bin/view/LHCbPhysics/DDKSKS Code Paterne LiCkb code > Add Nave Ins		inpur Dai	Data	Location		Comment Update on D0->KS KS	
Internal Discussions Location Location Location Code Plafere LCHester of Add New Iwn Date Add New Iwn Code Pesentations URL https://cds.cem.ch/record/2053739 URL https://cds.cem.ch/record/2037647			MC Data	location			
Location Internal Discussions Code Plaform LCh code + Add New Imm Code Date MC Date Date Presentations URL https://wiki.cem.ch/wiki/bin/view/LHCbPhysica/D0KSKS Presentations URL https://dis.cem.ch/wiki/bin/view/LHCbPhysica/D0KSKS Presentations URL https://dis.cem.ch/reicord/2053739 URL https://dis.cem.ch/record/2037647 Output Date Locations URL Locations URL https://dis.cem.ch/record/2037647 Publications Journal Title							
Locotion URL Locotion URL Code Hafern LCH-bode +.4dd New Nex User code +.4dd New Nex Code Dots Mc Date			Location		Internal Discussions		
Code Platform Code Platform User code Add New Ises Output Dates Dates MC Dates URL MC Dates MC Dates MC Dates Journal Title Platform LUFt No code Platform LUFt No code Platform Platform Platform Platform Platform Platform Platform Platform Platform Platform Platform Platform Platform Platform Platform				Location		URL Lung (/ with some of / with / his / his / his / his / his / box size / box size	
Code Plafform LCHc code 4.dd New Nam User code 4.dd New Nam Output Dates Date MC Date MC Date MC Date Add New Nam MC Date MC Date Add New Nam Journal Title						https://twiki.cern.cn/twiki/bin/view/LttCbrnysics/D0K3K3	
Input Data Code Input Data Code Input Data Code Persentations URL Input Data Code Publications Publications Publications Input Data Add New Item Input Data Input Data Input Data Input Data Input Data In		Co	de				
Vier cole Add New Item Output Date Date MC Date MC Date Input Date Date MC Date Add New Item Journal Title Journal Title	Output D		LCHb code + Add	New Item	Presentations		
Output Data Output Data Output Data Data MC Data URL https://cds.cern.ch/record/2037647 URL https://cds.cern.ch/record/2037647 Publications Input Data MC Data Add New Item User code Mc Data MC Data Add New Item User code User code User code User code User code <td>liter code</td> <td></td> <td></td> <td></td>			liter code				
Output Date MC Date MC Date Input Date Input Date MC Date<				New Item		URL https://cds.cern.ch/record/2053739	
Ling MC Data MC Data <td>ata Di C</td> <td></td> <td></td> <td></td>			ata Di C				
Input Data Input Data <td>Data</td> <td></td> <td></td> <td>URL https://cds.cern.ch/record/2037647</td>			Data			URL https://cds.cern.ch/record/2037647	
Input Data Data Add New Item MC Data + Add New Item						E. E. M. Cardina and S. Barradia A. Sarada and S. Sarada	
Input Data Code Platform LCHb code + Add New Item							
Input Data Add New Item MC Data Add New Item Code Platform LCHb code + Add New Item User code + Add New Item					Publications		
MC Data + Add New Item Code Platform LCHb code + Add New Item User code + Add New Item	Input		Data + Add New Item			laureal Titla	
Code Platform LCHb code + Add New Item User code + Add New Item			MC Data + Add 1	New Item			
Code Platform LCHb code +Add New Item User code +Add New Item							
LCHb code + Add New Item		Co	de Platform				
			LCHb code + Add 1	New Item			
			User code	New Item			

CERN Open data portal

Access point to data produced by the research and experiments conducted at CERN

Provides the data, software and documentation

Data is going to be preserved if it is available online and used by scientists worldwide





CERN launches Open Data Portal to make public the data of LHC experiments cern.ch /go/tN15T #cernopendata

CERN Open data portal



Database of the data and software dependencies

Y L author -

the second second

Experimental data

During the run, there are 40 million collisions per second

The mechanism called the trigger identifies interesting events and saves them, discarding the other 99.9% of the data

Experimental data

Elementary particles collide creating unstable particles that decay quickly

Necessary to reconstruct an "image" of the event



Simulated data

Simulation mimic what happens in the LHCb detector

Comparing the simulated with the real data helps us interpret the results

The volume of the simulated data is bigger than the real data

LHCb data management

Flow of the real and simulated data:

Reconstruction

- From raw data format to readable data
- Heavy software for processing

Data reduction

Producing streams corresponding to activity of the working groups

User analysis





Size of the LHCb data

What do we save?

- O(10) PB raw data
- O(100) TB processed data
- O(1) TB users' data

LHCb software

Gaudi framework provides interfaces and services for event data processing applications

- DaVinci application Particles manipulation and measuring physics processes
- Brunel application Event reconstruction: particle tracks, particle IDs

• Etc.

Data are compatible with different software versions

Data and software dependences

The database with:

- information about the software, the versions released and their relationships (e.g. what do I need if I want to run DaVinci X)
- information about the data lifecycle, from primitive data files to processed data and their compatibility with the LHCb software

Use cases

- 1. Short term future: Software needed to analyse the data from 2012
- 2. Automatically determining tests that have to be run to guarantee we can still (re)analyse the data
- 3. Identifying legacy software versions
- 4. CERN Analysis Preservation (CAP) portal
- 5. LHCb web pages

Implementation

Implemented in the Graph database Neo4j



Examples: easy lookup for the data



Examples:

Data with particular detector conditions

Thank you for your attention!

