



The JADE long term data preservation projects in Max-Planck Institute für Physik

Andrii Verbytskyi on behalf of JADE Resurrection Group

2nd DPHEP Collaboration Meeting
Geneva,
March 14, 2017

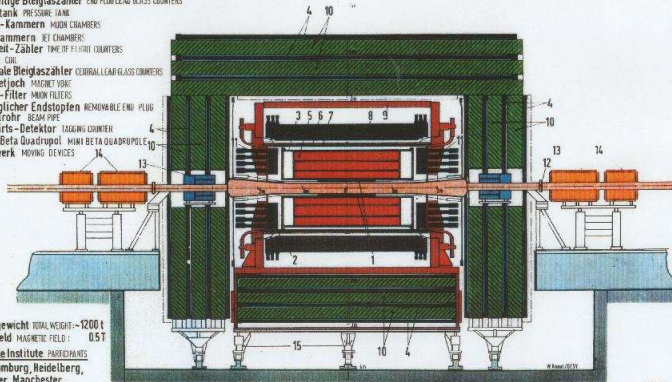
PETRA

- 
- Electron-positron collider, started in 1979 in DESY;
 - 2.3km ring, beam energies up to 23GeV;
 - Experiments: TASSO, MARK-J, CELLO, **JADE**.

JADE

MAGNETDETEKTOR MAGNET DETECTOR JADE

- 1 Strahlrohrzähler BEAM PIPE COUNTERS
- 2 Endseitige Bleiglaszähler END PLUG LEAD GLASS COUNTERS
- 3 Drucktank PRESSURE TANK
- 4 Myon-Kammern MUON CHAMBERS
- 5 Jet-Kammern JET CHAMBERS
- 6 Flugzeit-Zähler TIME OF FLIGHT COUNTERS
- 7 Spule COIL
- 8 Zentrale Bleiglaszähler CENTRAL LEAD GLASS COUNTERS
- 9 Magnetjoch MAGNET YOKE
- 10 Myon-Filterer MUON FILTERERS
- 11 Beweglicher Endstopfen REMOVABLE END PLUG
- 12 Strahlrohr BEAM PIPE
- 13 Vorwärts-Detektor TRACKING COUNTER
- 14 Mini-Beta Quadrupol MINI BETA QUADRUPOLE
- 15 Fahrwerk MOVING DEVICES



Gesamtgewicht TOTAL WEIGHT: ~1200 t
Magnetfeld MAGNETIC FIELD: 0.5 T

Beteiligte Institute PARTICIPANTS
DESY, Hamburg, Heidelberg,
Lancaster, Manchester,
Rutherford Lab., Tokio

33188

- multipurpose detector
- large solid angle coverage
- digital readout
- advanced tracking system
- calorimeter
- muon chambers

35 years later: still unique energy range coverage!

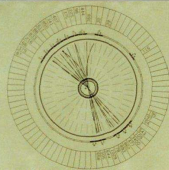
JADE and data preservation

JADE@PETRA reminder:

- One of key experiments for QCD: discovery of gluon, α_s measurements.
- **The oldest and most successful Data Preservation effort!**

Motivation for data preservation:

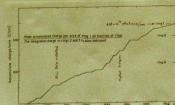
- Future data (re-)analysis with new models and new approaches.
- Modelling for the future experiments.



Records of the
JADE
Inner Detector
at PETRA

Operation 14.Jan.1979–3.Nov.1986

Data Taking	21 100 h
Electronics on	36 700 h
Integrated Luminosity	216 pb ⁻¹
Clean Multihadrons	43 100
CM-Energy Range	12–46.78 GeV
ON/OFF Switchings	~11 500
Emergency Switch-offs	2 000
Gasfillings	22
Mean Accumulated Charge on the Wires of Ring 1	1,45·10 ¹⁷ <u>electrons</u> cm



JADE data preservation history

- 1986: End of data taking.
- 1995: Private initiative to rescue data (J. Olsson)
- 1995-2003: Preservation in MPP
 - Physical transfer of data to MPP;
 - Software port to AIX4.3@IBM RS6000;
 - Interface to Pythia6 and Herwig MC;
 - PAW output;
 - Preservation of paper documentation.
- 1996-2013: Physics
 - 11 papers, O(40) conference talks, thesis, JADE notes.
- 2016-2017: Update of preservation:
 - Data is available online;
 - Software port to Linux(Mac)@x86_64;
 - Virtualisation;
 - Interface to HepMC3 (enables most modern MC generator);
 - ROOT output;
 - Preparation of digital documentation on computing notes.

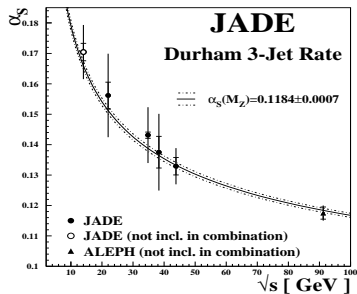
Yes, 11 papers!

Data preservation is about new and interested results with old data. In addition the Data Preservation experience with JADE has an extreme importance on itself.

In our model we describe ingredients and tools:

- Data bits
- Software
- But in the end we are interested in **physics** .
- Experiment documentation
- DP policies and documentation

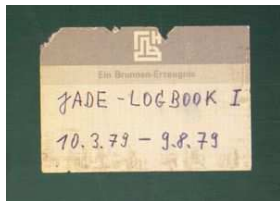
Main idea: enable physics and make it doable with modern methods in modern environments with minimal effort.



EPJ C73 (2013) 3, 2332

MPP DP model for documentation and policies

- JADE publications are available in InSpire, journals, arXiv or scanned by KEK.
- The non-digital documentation is preserved in DESY and MPP.
- Logbooks included!
- Some available online as well, see details at <https://wwwjade.mpp.mpg.de/>



MPP DP model for JADE data bits

- JADE data are stored in MPCDF on locally accessible tapes and in disk pool.
- Access via multiple protocols with grid tools worldwide to disk pools.
- Straightforward procedure to add new (MC) samples.
- In the end nowadays all the data from JADE can fit to a modern USB stick.



JADE data in MPP: Bits statistics

Data are PAW ntuples, ASCII files and FPACK compressed data.

	MPCDF
Files:	5.4k
Volume:	600GB (645 × 10⁹ bytes)
Work area:	yes
Access:	Worldwide
Protocols:	Multiple, see list
Auth:	Grid certificate

Available at:

- <gsidcap://grid-srm.rzg.mpg.de:22128/pnfs/rzg.mpg.de/data/zeus/jade>
- <grid-gftp2.rzg.mpg.de>
- <davs://grid-dav.rzg.mpg.de:2880//zeus/jade>
- ...



MPP model for software preservation

Explicit effort put to make software it work in the next 10-15 years.
Previous efforts and high quality of code made it possible.

Key ideas:

- Rely on industry, not HEP-only standards.
- Enable integration and compatibility with new physics software, e.g. data bases and Monte Carlo generators.
- **i.e. make software analysis-ready**

JADE software in MPP

- Full chain for reconstruction of raw data to PAW/ROOT ntuples was resurrected.
- Main software for the analysis of early preserved and reconstructed data is vanilla PAW or ROOT(via h2root).
- Additional software includes:
 - event display;
 - Monte-Carlo generation packages;

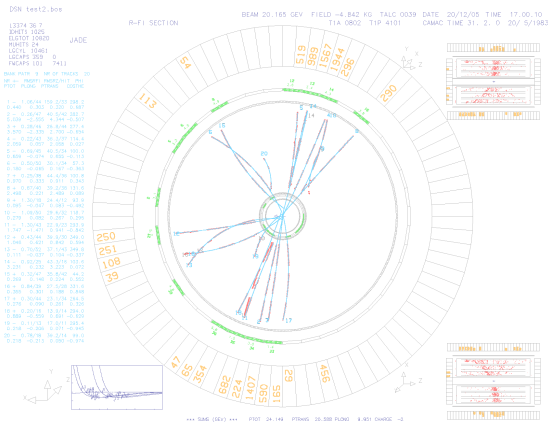
(Original)Software is available at <https://wwwjade.mpp.mpg.de>

JADE software porting quest in detail

- Original sources in FORTRAN IV (1974), FORTRAN 77, Sheltran, Mortran, Assembler ported in 2003 to AIX4.3@RS6000, the old routines replaced with CERNLIB libraries and compiled with IBM FORTRAN.

Huge work!

- In 2017 everything is problematic: AIX@ppc is Big Endian architecture, no CERNLIB in modern systems, IBM FORTRAN is not for free, 32 bit systems are dying out etc.



JADE software porting quest: step one

One can start from:

- PowerPC64 multi-arch CentOS7 Big Endian on QEMU2.7;
- CERNLIB compilation for PowerPC **Yes, for many flavours!**;
- Trial version of IBM xlf15 for Linux;
- Manual or semiautomatic fixes of differences between FORTRAN dialects (e.g. HOLLERITH, INTEGER vs. CHARECTER).

FYI: Some codes date back to 1974.

JADE software porting quest: step two

Once the software passes simplest tests:

- Switch to gfortran;
- Start to use cmake (best thing for FORTRAN!);
- Switch to Little Endian Linux and control I/O Endianess with `GFORTRAN_CONVERT_UNIT` ;
- Replace as many as possible CERNLIB routines with ROOT or dummy routines;
- The remaining routines just copy from CERNLIB (i.e. create "picocernlib");
- **Compile codes in native 64-bit mode.**

JADE software porting quest: step two

Once the code is compiled on a standard computer, one has to fix bugs in functionality or implement missing features.

Here the problems begin.

- Part of the code was lost long time ago: muon reconstruction is not available.
- Event display was ported to HIGZ/CERNLIB in 2003. To the version of CERNLIB that is hard to find. Most graphics primitives re-implemented in ROOT, but display is still not fully operational.
- The data is compressed with FPACK utility, but the way it was done is not known, same as the way to unpack it. A collaboration with J.Olsson might be helpful.

JADE software porting quest: step three

On the positive side:

- The key to productive usage of JADE data in the past was an option to generate MC with much newer generators. The option was implemented with integration of Pythia/Herwig/Ariadne codes into JADE software.
- Now it is simpler: Monte Carlo generators have standard output formats and only a converter from a standard format to JADE-specific input has to be developed.
- Done with HepMC3, very nice library.

A certain environment is needed for the analysis.

- Virtual machines(VM) are a very attractive **long-term** solution;
- The way other experiments (LEP/LHC) are going;
- The solution has very generic requirements, will survive for a long time.
- Unlimited number of installations → potentially usable on clouds;
- Usage not restricted to any laboratory or virtualisation software. Can run anywhere.
- Usage of ISO installation image assures vendor independence.

VM is available at: <https://www.jade.mpp.mpg.de>

JADE software environment/VM

VM includes:

- JADE software: ROOT, MC simulation, event display, file catalogue, setup scripts etc. See a detailed list in backup slides.
- Modern MC generators, FastJet, CERNLIB, PAW and other popular and “not really” packages.
- Anything you will want to install. . .
- Agree access and download it.
- VirtualBox images are provided as well.

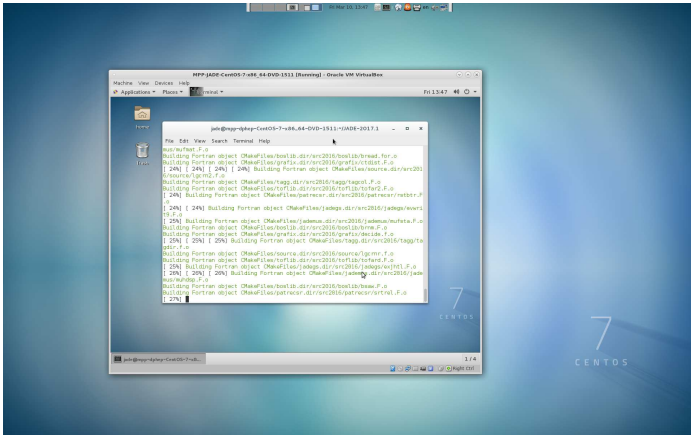
JADE software environment/VM

```
CentOS Linux (3.10.0-327.el7.x86_64) 7 (Core)
CentOS Linux (0-rescue-509eac0305b3419c88634ea8800a62c2) 7 (Core)

Use the ↑ and ↓ keys to change the selection.
Press 'e' to edit the selected item, or 'c' for a command prompt.
```

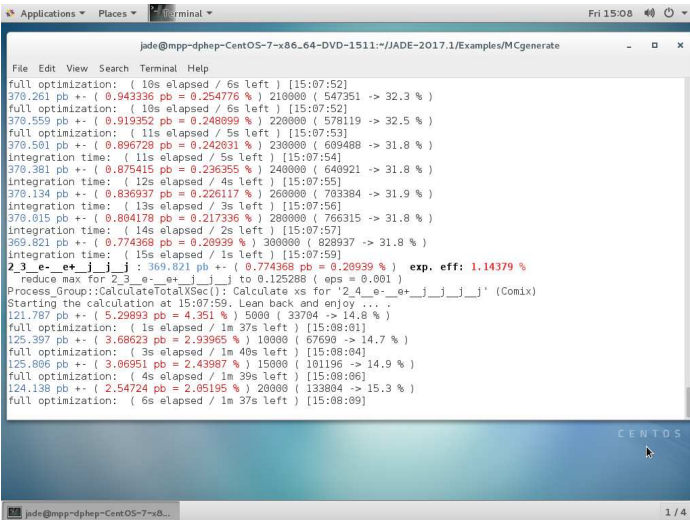
Generic start of VM.

JADE software environment/VM test



JADE software compilation in VM: takes about 4 minutes.

JADE software environment/VM test

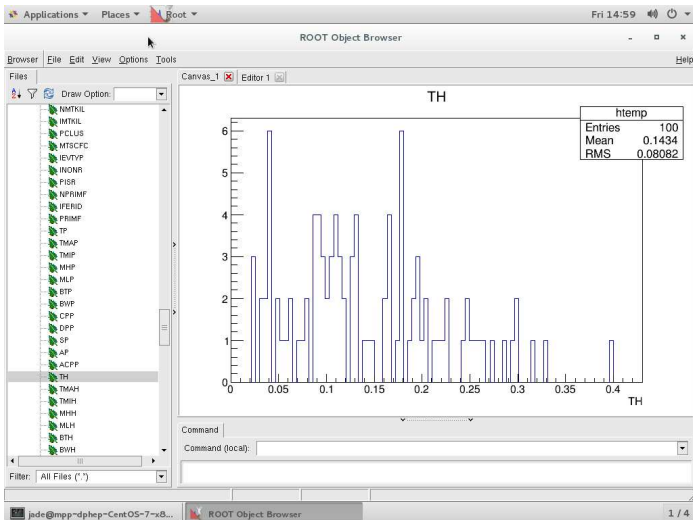


The screenshot shows a terminal window titled "jade@mpp-dpheap-CentOS-7-x86_64-DVD-1511:~/JADE-2017.1/Examples/MCgenerate". The terminal displays the output of a Monte Carlo simulation. The output includes several lines of performance metrics, such as "full optimization: (10s elapsed / 6s left)" and "integration time: (11s elapsed / 5s left)". The simulation results are summarized in a table-like format, showing the number of particles (pb) and the percentage of particles that are "pb" (likely "pb" for "particles per bin" or similar). The final line of the simulation output is "2_3_e-_e+_j_j_j : 369.821 pb +- (0.774368 pb = 0.20939 %) exp. eff: 1.14379 %". The terminal window also shows the system's desktop environment, including the "Applications" and "Places" menus, and the system clock showing "Fri 15:08".

```
Applications ▾ Places ▾ Terminal ▾ Fri 15:08
jade@mpp-dpheap-CentOS-7-x86_64-DVD-1511:~/JADE-2017.1/Examples/MCgenerate
File Edit View Search Terminal Help
full optimization: ( 10s elapsed / 6s left ) [15:07:52]
370.261 pb +- ( 0.943336 pb = 0.254776 % ) 210000 ( 547351 -> 32.3 % )
full optimization: ( 10s elapsed / 6s left ) [15:07:52]
370.559 pb +- ( 0.919352 pb = 0.248099 % ) 220000 ( 578119 -> 32.5 % )
full optimization: ( 11s elapsed / 5s left ) [15:07:53]
370.501 pb +- ( 0.896728 pb = 0.242031 % ) 230000 ( 609488 -> 31.8 % )
integration time: ( 11s elapsed / 5s left ) [15:07:54]
370.381 pb +- ( 0.875415 pb = 0.236355 % ) 240000 ( 640921 -> 31.8 % )
integration time: ( 12s elapsed / 4s left ) [15:07:55]
370.134 pb +- ( 0.836937 pb = 0.226117 % ) 260000 ( 703384 -> 31.9 % )
integration time: ( 13s elapsed / 3s left ) [15:07:56]
370.015 pb +- ( 0.804178 pb = 0.217336 % ) 280000 ( 766315 -> 31.8 % )
integration time: ( 14s elapsed / 2s left ) [15:07:57]
369.821 pb +- ( 0.774368 pb = 0.20939 % ) 300000 ( 828937 -> 31.8 % )
integration time: ( 15s elapsed / 1s left ) [15:07:59]
2_3_e-_e+_j_j_j : 369.821 pb +- ( 0.774368 pb = 0.20939 % ) exp. eff: 1.14379 %
reduce max for 2_3_e-_e+_j_j_j to 0.125288 ( eps = 0.001 )
Process Group::CalculateTotalXSec(): Calculate xs for '2_4_e-_e+_j_j_j_j' (Comix)
Starting the calculation at 15:07:59. Lean back and enjoy ...
121.787 pb +- ( 5.29893 pb = 4.351 % ) 5000 ( 33704 -> 14.8 % )
full optimization: ( 1s elapsed / 1m 37s left ) [15:08:01]
125.397 pb +- ( 3.68623 pb = 2.93965 % ) 10000 ( 67690 -> 14.7 % )
full optimization: ( 3s elapsed / 1m 40s left ) [15:08:04]
125.806 pb +- ( 3.06951 pb = 2.43987 % ) 15000 ( 101196 -> 14.9 % )
full optimization: ( 4s elapsed / 1m 39s left ) [15:08:06]
124.138 pb +- ( 2.54724 pb = 2.05195 % ) 20000 ( 133804 -> 15.3 % )
full optimization: ( 6s elapsed / 1m 37s left ) [15:08:09]
CENTOS
jade@mpp-dpheap-CentOS-7-x86... 1 / 4
```

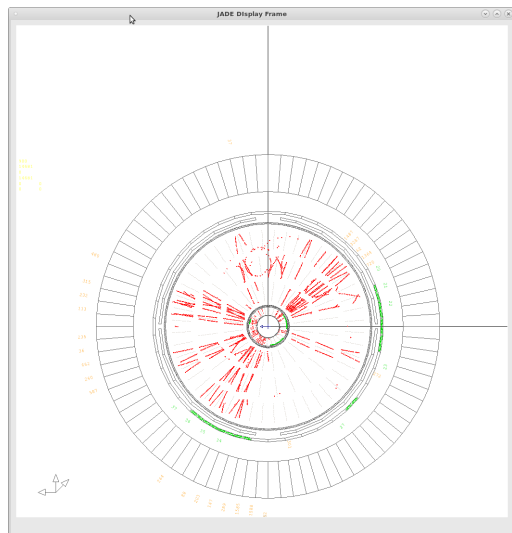
MC for JADE can be generated in VM or elsewhere. In this case SHERPA is running in VM.

JADE software environment/VM test



In the end of reconstruction a ROOT file with simple tree is delivered. Thrust distribution from the generated events is shown.

JADE software environment/VM test



Test of the reconstructed MC files with event display. **Not all functionality is restored. To be fixed..**

MPP Data Preservation summary

- Huge work was done since 1995!
- Many high quality, important physics results delivered.
- Data is accessible in MPCDF and software is **analysis-ready**.

Potential improvements are obvious: enable data reading, fixes to event display, improving documentation.

- Hadronisation effects.
- QCD with b -quarks and modern theory predictions, e.g. tests of future fully differential NNLO predictions for $e^+e^- \rightarrow \text{hadrons}$ with heavy quarks.
- ...

- Generic technologies for hardware and OS is an advantage.
- High code quality, clarity and stability is a key to success. Note: some routines are from 1974, i.e. 43 y.o. Few modern experiments can be sure that more than tiny fraction of their code will behave like that.
- Design of software is important.
- Good build system makes porting much easier
- Standardisation of I/O data formats is important.
- Documentation on simple things is important, i.e. meaning of arguments for utilities.

- Compressing the data might look like a good solution, but it is bad. Saving 1Tb of disk space today mean lose of days to read the data back tomorrow. **Do not compress your data!**
- Keeping reference data/MC/results is important.
- Standardisation of I/O data formats is important.

What other experiments can learn from JADE DP: Some anecdotes

- one "calibration" file, with luminosities for each run and fill, was stored on a private account and therefore lost when DESY archive was cleaned up;
- Jan Olsson, when cleaning up his office in 1997, found an old ASCII-printout of the JADE luminosity file. Unfortunately, it was printed on green recycling paper - not suitable for scanning and OCR-ing. A secretary at Aachen re-typed it within 4 weeks. A checksum routine found (and recovered) only 4 typos.
- an old version of the original BOSlib 1979 version was found, on request at the Tokyo computer centre.
- Peter Bock, when cleaning out an old lab at the Physics Institute at Heidelberg University, found a few 9-track tapes containing original JADE MC files which were very valuable for validating results of first re-analyses in 1997

What other experiments can learn from JADE DP: Some anecdotes

- First port attempt in 2016 was done with AIX 4.3 machine used as router in MPP, but turns out VirtualBox is more useful.
- Many problems in software porting were solved in the industry: byte ordering problem in the modern gfortran, availability of Linux on PowerPCs, good FORTRAN build system (cmake) is available now, xlf compiler is available on Linux now.
- Googling the arxiv used for the data, FPACK is a hard task.

List of custom/extra RPM packages on JADE VM

Some subpackages omitted.

blackhat-0.9.9-1.el7.centos.x86_64.rpm

blas-devel-3.4.2-5.el7.x86_64.rpm

cernlib-2006-36.el7.centos.i686.rpm

cernlib-devel-2006-36.el7.centos.i686.rpm

cernlib-packlib-gfortran-2006-36.el7.centos.i686.rpm

cernlib-static-2006-36.el7.centos.i686.rpm

cernlib-utils-2006-36.el7.centos.i686.rpm

epel-release-7-5.noarch.rpm

fastjet-3.1.2-1.el7.centos.x86_64.rpm

form-4.1-1.el7.centos.x86_64.rpm

geant321-2006-36.el7.centos.i686.rpm

gosam-2.0.3-1.el7.centos.x86_64.rpm

gosam-contrib-2.0-1.el7.centos.x86_64.rpm

Herwig-7.0.2-2.el7.centos.x86_64.rpm

lapack-devel-3.4.2-5.el7.x86_64.rpm

LHAPDF-6.1.6-6.el7.centos.x86_64.rpm

osg-ca-certs-1.55-1.el7.centos.noarch.rpm

osg-release-3.3-5.el7.centos.noarch.rpm

patchy-gfortran-2006-36.el7.centos.i686.rpm

paw-gfortran-2006-36.el7.centos.i686.rpm

PHOTOS-3.61-1.el7.centos.x86_64.rpm

pythia8-8.2.15-102.el7.centos.x86_64.rpm

qd-2.3.15-100.el7.centos.x86_64.rpm

qgraf-3.1.4-1.el7.centos.x86_64.rpm

root-5.34.36-1.el7.centos.x86_64.rpm

SHERPA-MC-2.2.0-3.el7.centos.x86_64.rpm

TAUOLA-1.1.5-1.el7.centos.x86_64.rpm

ThePEG-2.0.2-1.el7.centos.x86_64.rpm

vincia-1.2.02-1.el7.centos.x86_64.rpm

xbae-4.60.4-12.el7.centos.i686.rpm

All JADE software is packed as one rpm that is installed as is in /opt