КОМПАС COMPAS

On the quality assurance of PDG assessed physics data in computer readable forms

Vladimir Ezhela

Particle Physics Data Center at IHEP

Protvino, Russia

PDG meetings 2008

Motivation

Physics community needs the computer readable RPP data files of metrological quality to be consistently included into:

- » **MC-generators**
- » **Physics analyses systems**
- » **Software for models testing**

Historically it turns out that traditional chain to assure the quality of the published scientific data:

Authors → Journal peer reviewers → Editorial boards

and evolved publishing standards does not enough to represent multidimensional correlated data with the metrological quality

This presentation is to show that even the more powerful chain:

Authors → Journal peer reviewers → Journal editors →
→ RPP article finders → RPP encoders → RPP overseers →
→ Verifiers(Authors) → RPP peer reviewers → RPP editors →
→ Journal peer reviewers → Journal editors

used by PDG collaboration does not enough to represent RPP data with metrological quality needed for different applications

Problems Overview

1. RPP reference data adopted from other data centers
2. RPP reference data extracted from original papers
3. RPP evaluated data from reviews and mini-reviews
4. RPP evaluated data from fits and averages

1. RPP reference data adopted from other data centers

The NIST Reference on
Constants, Units, and Uncertainty

CODATA Internationally recommended values of the Fundamental Physical Constants

Reviews of Modern Physics

Over-rounding and improper incertanty propagation for derived quantities {*me*, *e*, *1/ α(0)*, *h*}

CODATA: 1986 (1987)	Symbol	Unit	Value(Uncertainty)xScale	Correlations		
Elementary charge	*e*	C	1.602 177 33(49) x 10^(-19)	*e*	*h*	*me*
Planck constant	*h*	J s	6.626 075 5(40) x 10^(-34)	0.997		
Electron mass	*me*	kg	9.109 389 7(54) x 10^(-31)	0.975	0.989	
1/(Fine struct. const.)	1/α(0)		137.035 989 5(61)	−0.226	−0.154	−0.005
CODATA: 1998 (2000)						
Elementary charge	*e*	C	1.602 176 462(63) x 10^(-19)	*e*	*h*	*me*
Planck constant	*h*	J s	6.626 068 76(52) x 10^(-34)	0.999		
Electron mass	*me*	kg	9.109 381 88(72) x 10^(-31)	0.990	0.996	
1/(Fine struct. const.)	1/ α(0)		137.035 999 76(50)	−0.049	−0.002	0.092
CODATA: 2002 (2005)						
Elementary charge	*e*	C	1.602 176 53(14) x 10^(-19)	*e*	*h*	*me*
Planck constant	*h*	J s	6.626 0693(11) x 10^(-34)	1.000		
Electron mass	*me*	kg	9.109 3826(16) x 10^(-31)	0.998	0.999	
1/(Fine struct. const.)	1/ α(0)		137.035 999 11(46)	−0.029	−0.010	0.029
CODATA: 2006 (2008)						
Elementary charge	*e*	C	1.602 176 487(40) x 10^(-19)	*e*	*h*	*me*
Planck constant	*h*	J s	6.626 068 96(33) x 10^(-34)	0.9999		
Electron mass	*me*	kg	9.109 382 15(45) x 10^(-31)	0.9992	0.9996	
1/(Fine struct. const.)	1/ α(0)		137.035 999 679(94)	−0.0142	−0.0005	0.0269

Correlator eigenvalues of the selected constants in CI units

1986: { 2.99891, 1.00084, 0.000420779, **-0.000172106**}

1998: { 2.99029, 1.01003, **-0.000441572**, 0.00012358 }

2002: { 2.99802, 1.00173, 0.000434393, **-0.000183906** }

2006: { 2.99942, 1.00006, 0.000719993, **-0.000202165**}

Correlation matrix(e, h, m_e, 1/α(0)) of uncertainties in "Energy" units

CODATA : 2006(8)	Symbol	[units]	Value (uncertainty) scale	Correlations		
				e	h	m_e
Elementary charge	e	[C]	1.602 176 487(40)10^{-19}			
Planck constant	h	[eVs]	4.135 667 33(10)10^{-15}	0.9996		
Electron mass	m_e	[MeV]	0.510 998 910(13)	0.9966	0.9985	
1/α(0)	$\alpha(0)^{-1}$		137.035 989 5 (61)	-0.0142	0.0132	0.0679

Eigenvalues → { 2.99721, 1.00275, 0.0000341718, 1.40788 10^{-6} }

Origine: Linear Uncertainties Propagation & Over-rounding

Unfortunately there is no possibility to assess the quality of the derived FPC-2006 because of absence of the corresponding LSA-2006 data files.

http://physics.nist.gov/cuu/LSAData/index.html

Moreover, the recently published data on the correlation matrix of the inputs to evaluate Rydberg constant by LSA method has two negative eigenvalues **(see Rev. Mod. Phys. 80 (2008) 633, TABLE XXIX.)** .

Most probably this is due to the over-rounding of the matrix elements when preparing data for traditional publication.

Having the LSA-2006 files and standardized formulae for the derived FPC (that could be obtained from the NIST-FCDC), COMPAS group can produce the fpcLive of metrological quality to be inserted into pdgLive

2. RPP reference data extracted from original papers

European Physical Journal C46 (2006) 1
Experiment CERN-LEP-DELPHI

A measurement of the tau hadronic branching ratios

Table10. Measured branching ratios in percent.The uncertainties are statistical followed by systematic

Decay mode		BranchingRatio(%)
$\tau^- \to h^-$	$\geq 0 K^0 \nu_\tau$	12.780 ± 0.120 ± 0.103
$\tau^- \to h^- \pi^o$	$\geq 0 K^0 \nu_\tau$	26.291 ± 0.201 ± 0.130
$\tau^- \to h^- 2\pi^o$	$\geq 0 K^0 \nu_\tau$	9.524 ± 0.320 ± 0.274
$\tau^- \to h^- \geq 1\pi^o$	$\geq 0 K^0 \nu_\tau$	37.218 ± 0.155 ± 0.116
$\tau^- \to h^- \geq 2\pi^o$	$\geq 0 K^0 \nu_\tau$	10.927 ± 0.173 ± 0.116
$\tau^- \to h^- \geq 3\pi^o$	$\geq 0 K^0 \nu_\tau$	1.403 ± 0.214 ± 0.224
$\tau^- \to 3h^\pm$	$\geq 0 K^0 \nu_\tau$	9.340 ± 0.090 ± 0.079
$\tau^- \to 3h^\pm \pi^o$	$\geq 0 K^0 \nu_\tau$	4.545 ± 0.106 ± 0.103
$\tau^- \to 3h^\pm \geq 1\pi^o$	$\geq 0 K^0 \nu_\tau$	5.106 ± 0.083 ± 0.103
$\tau^- \to 3h^\pm \geq 2\pi^o$	$\geq 0 K^0 \nu_\tau$	0.561 ± 0.068 ± 0.095
$\tau^- \to 5h^\pm$	$\geq 0 K^0 \nu_\tau$	0.097 ± 0.015 ± 0.005
$\tau^- \to 5h^\pm \geq 1\pi^o$	$\geq 0 K^0 \nu_\tau$	0.016 ± 0.012 ± 0.006

Table 11.Correlation matrix of the combined statistical and systematic uncertainties.The last three rows show the correlation with the topological branching ratios presented in [16].

$h^-\nu_\tau$	1.00											
$h^-\pi^0\nu_\tau$	−0.34	1.00										
$h^- \geq 1\pi^0\nu_\tau$	−0.47	0.56										
$h^- 2\pi^0\nu_\tau$	0.06	−0.66	0.15	1.00								
$h^- \geq 2\pi^0\nu_\tau$	−0.03	−0.74	0.15	0.81	1.00							
$h^- \geq 3\pi^0\nu_\tau$	−0.06	0.38	0.11	−0.86	−0.36	1.00						
$3h^\pm\nu_\tau$	−0.07	−0.08	0.15	0.00	−0.03	−0.02	1.00					
$3h^\pm\pi^0\nu_\tau$	−0.02	−0.01	−0.05	−0.03	−0.02	0.03	−0.53	1.00				
$3h^\pm \geq 1\pi^0\nu_\tau$	−0.04	−0.04	−0.13	−0.04	−0.06	−0.02	−0.56	0.75	1.00			
$3h^\pm \geq 2\pi^0\nu_\tau$	−0.01	−0.01	−0.04	0.03	−0.02	−0.06	0.26	−0.78	−0.16	1.00		
$5h^\pm\nu_\tau$	−0.01	−0.01	0.01	0.00	0.00	0.00	−0.02	−0.03	−0.01	0.03	1.00	
$5h^\pm \geq 1\pi^0\nu_\tau$	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	−0.05	−0.05	−0.57	1.00
B1	0.09	0.10	0.26	0.04	0.11	0.03	−0.50	−0.25	−0.39	−0.06	−0.03	0.00
B2	−0.09	−0.10	−0.26	−0.04	−0.11	−0.03	0.50	0.25	0.39	0.06	0.03	0.00
B3	−0.02	0.00	0.00	0.00	0.00	0.00	−0.03	0.03	0.00	0.00	0.72	0.40

Table 11.Correlation matrix of the combined statistical and systematic uncertainties as it is reproduced in **pdgLive-2007(8).**

$h^-\nu_\tau$	1.00											
$h^-\pi^0\nu_\tau$	−0.34	1.00										
$h^- \geq 1\pi^0\nu_\tau$	−0.47	0.56										
$h^- 2\pi^0\nu_\tau$	0.06	−0.66	0.15	1.00								
$h^- \geq 2\pi^0\nu_\tau$	−0.03	−0.74	0.15	0.81	1.00							
$h^- \geq 3\pi^0\nu_\tau$	−0.06	0.38	0.11	−0.86	−0.36	1.00						
$3h^\pm\nu_\tau$	−0.07	−0.08	0.15	0.00	−0.03	−0.02	1.00					
$3h^\pm\pi^0\nu_\tau$	−0.02	−0.01	−0.05	−0.03	−0.02	0.03	−0.53	1.00				
$3h^\pm \geq 1\pi^0\nu_\tau$	−0.04	−0.04	−0.13	−0.04	−0.06	−0.02	−0.56	0.75	1.00			
$3h^\pm \geq 2\pi^0\nu_\tau$	−0.01	−0.01	−0.04	0.03	−0.02	−0.06	0.26	−0.78	−0.16	1.00		
$5h^\pm\nu_\tau$	−0.01	−0.01	0.01	0.00	0.00	0.00	−0.02	−0.03	−0.01	0.03	1.00	
$5h^\pm \geq 1\pi^0\nu_\tau$	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	−0.05	0.05	−0.57	1.00

Omitted in RPP

-0.05

In original text

This matrix is assigned in the RPP data block to the observables of Table 12 not to the observables of Table 10 as it is in the original paper.

*"Using the world averages [18] for the channels involving **K^0** and neglecting this contribution for channels with more than three charged pions or kaons, we can derive the branching ratios shown in Table 12. In this subtraction, the total error on the world average was added in quadrature to the systematic error of these measurements."*

Table 12. Measured branching ratios in percent after subtraction of the contributions of channels including K^0. The uncertainties are statistical followed by systematic

Decay mode	Branching Ratio (%)
$\tau^- \to h^- \nu_\tau$	11.571 ± 0.120 ± 0.114
$\tau^- \to h^- \pi^o \nu_\tau$	25.740 ± 0.201 ± 0.138
$\tau^- \to h^- 2\pi^o \nu_\tau$	9.498 ± 0.320 ± 0.275
$\tau^- \to h^- \geq 1\pi^o \nu_\tau$	36.641 ± 0.155 ± 0.127
$\tau^- \to h^- \geq 2\pi^o \nu_\tau$	10.901 ± 0.173 ± 0.118
$\tau^- \to h^- \geq 3\pi^o \nu_\tau$	1.403 ± 0.214 ± 0.224
$\tau^- \to 3h^\pm \nu_\tau$	9.317 ± 0.090 ± 0.082
$\tau^- \to 3h^\pm \pi^o \nu_\tau$	4.545 ± 0.106 ± 0.103
$\tau^- \to 3h^\pm \geq 1\pi^o \nu_\tau$	5.106 ± 0.083 ± 0.103
$\tau^- \to 3h^\pm \geq 2\pi^o \nu_\tau$	0.561 ± 0.068 ± 0.095
$\tau^- \to 5h^\pm \nu_\tau$	0.097 ± 0.015 ± 0.005
$\tau^- \to 5h^\pm \geq 1\pi^o \nu_\tau$	0.016 ± 0.012 ± 0.006

But it is impossible to do this evaluation reliably simply because there are no proper correlator of the corresponding “world averaged” tau branchings.

As a rule, PDG shows correlators in % for the pure informational purposes – to show highly correlated observables under study. → **The PDG correlators for branchings, are badly over-rounded.**

There is another problem with DELPHI correlators – both “correlation” matrices, original and presented in RPP, have two negative eigenvalues.

Such papers should be returned by referees to the senders for corrections.

Such “data” should not pass to the RPP repository without comments on the data corruption in spite of being published in journals with high impact factor.

Section summary

The module to test intrinsic consistency of the correlated input and output RPP data is urgently needed.

COMPAS group can workout the mockups of such module in *Mathematica*

3. RPP evaluated data from reviews and min-ireviews

In majority of the reviews and mini-reviews the evaluated particle physics parameters (the best current values) did not supported by the properly organized computer readable data files with input data and results of evaluations

4. RPP evaluated data from fits and averages

A proposal to improve presentation the results of constrained fits in computer readable forms (on a few simplest examples)

“**5.2.3. *Constrained fits:***

… In the Particle Listings, we give the complete correlation matrix; we also calculate the fitted value of each ratio, for comparison with the input data, and list it above the relevant input, along with a simple unconstrained average of the same input. ….”

Excerpt from page 17 of the Review

We see that there are no “complete correlation matrix” neither in the book nor on the pdgLive pages. We have over-rounded correlators instead, and can extract (a crazy job) non-rounded ones by using corresponding URLs from the CONSTRAINED FIT INFORMATION pages.

Moreover, it seems, that both correlation matrices have another problem. It turns out that if we have three random quantities x1, x2, x4 such that they obey the relation

$$x1 + x2 + x4 = 1,$$

then their covariance matrix is degenerate 3×3 matrix and its non-diagonal matrix elements completely determined by the diagonal ones $\sigma_{mn} = 2\,\rho_{mn}\cdot\sigma_m\,\sigma_n$, where

$$\rho_{mn} = (\,\sigma_k^2 - \sigma_m^2 - \sigma_n^2\,)\,/\,(\,2\,\sigma_m\,\sigma_n\,)\ ,\ (k\neq m\neq n) = (1,2,4)$$

are the correlations. Inserting corresponding σ_m data from pdgLive we obtain: →

CONSTRAINED FIT INFORMATION π^0 DECAY MODES

An overall fit to 2 branching ratios uses 4 measurements and one constraint to determine 3 parameters.
The overall fit has a χ^2 = 1.9 for 2 degrees of freedom.
The following *off-diagonal* array elements are the correlation coefficients $\langle \delta x_i \, \delta x_j \rangle / (\delta x_i \, \delta x_j)$, in percent, from the fit to x_i, including the branching fractions, $x_i = \Gamma_i / \Gamma_{total}$.
The fit constrains the x_i whose labels appear in this array to sum to one.

x_1	100		
x_2	–100	100	
x_4	–1	–0	100
	x_1	x_2	x_4

Rounded correlator

{2.00005, 1., -0.00005}

	x1	x2	x4
x1	1.00		
x2	-0.999958	1.00	
x4	-0.005585791	-0.003579367	1.00

"URL-rounded correlator"

{1.99996, 1.00004, -1.02849×10^{-10}}

	Rounded Correlator		
x1	1	-0.999956	-0.0046875
x2	-0.999956	1	-0.0046875
x4	-0.0046875	-0.0046875	1

Eigenvalues. Rounded correlator: {**1.99996, 1.00004, 5.46851×10^{-8}**}
Eigenvalues. Non-rounded correlator: {**1.99996, 1.00004, -1.21385×10^{-16}**}

We have no explanations of the obtained estimates of the correlator and propose the slightly modified procedure for the constrained fit

The Review of Particle Physics
C. Amsler *et al.*, Physics Letters **B667** (2008) 1

Ratio (R)	R-Value	R-Uncertainty	Formula (F)
$\Gamma(e^+ e^- \gamma)/\Gamma(2\gamma)$	0.0125	0.0004	x2/x1
$\Gamma(e^+ e^- \gamma)/\Gamma(2\gamma)$	0.01166	0.00047	x2/x1
$\Gamma(e^+ e^- \gamma)/\Gamma(2\gamma)$	0.0117	0.0015	x2/x1
$\Gamma(\gamma\ \mathrm{Atom}(e^+e^-)/\Gamma(2\gamma)$	1.84×10^{-9}	0.29×10^{-9}	x3/x1
$\Gamma(2e^+ 2e^-)/\Gamma(2\gamma)$	0.0000318	3.0×10^{-6}	x4/x1
$\Gamma(e^+ e^-)/\Gamma(\mathrm{total})$	6.46×10^{-8}	0.33×10^{-8}	x5
$\Gamma(\mathit{undetected})/\Gamma(\mathrm{total})$	0.0	6.0×10^{-4}	1-x1-x2-x3-x4-x5

7 measurements, 5 parameters

Proposal for "new" forms of constrained fits

$$\chi^2 = \Sigma\,(R - F)_i\, W_{ij}\, (R - F)_j + (10^8/36)\cdot \mathrm{UnitStep}[x1+x2+x3+x4+x5-1]\cdot(1-x1-x2-x3-x4-x5)^2$$

	Value		Error		Rounded correlator				
x1	0.98798		0.00066		1.00	-0.42	0.00	0.00	0.00
x2	0.01198		0.00029		-0.42	1.00	-0.00	-0.00	0.00
x3	1.82×10^{-9}	±	0.29×10^{-9}	;	0.00	-0.00	1.00	0.00	0.00
x4	31.4×10^{-6}		3.0×10^{-6}		0.00	-0.00	0.00	1.00	0.00
x5	6.46×10^{-8}		0.33×10^{-8}		0.00	0.00	0.00	0.00	1.00

Eigenvalues. Non rounded correlator: {1.41895,1.00000,0.99999,0.99993,0.58113}

Eigenvalues. Rounded correlator: {1.42, 1.00, 1.00, 1.00, 0.58}

Minimun(χ^2) = 1.94 for 7- 5 = 2 degrees of freedom

In addition we can obtain the estimate for the fraction of the sum of possible undetected decays xU = 1-x1-x2-x3-x4-x5. Our calculations give:

	Value		Error		Rounded correlator					
x1	0.98798		0.00066		1.00	-0.42	0.00	0.00	0.00	-0.90
x2	0.01198		0.00029		-0.42	1.00	-0.00	-0.00	0.00	-0.02
x3	1.82×10^{-9}	±	0.29×10^{-9}	;	0.00	-0.00	1.00	0.00	0.00	-0.00
x4	31.4×10^{-6}		3.0×10^{-6}		0.00	-0.00	0.00	1.00	0.00	-0.00
x5	6.46×10^{-8}		0.33×10^{-8}		0.00	0.00	0.00	0.00	1.00	0.00
xU	4.92×10^{-6}		600.0×10^{-6}		-0.90	-0.02	-0.00	-0.00	0.00	1.00

Eigenvalues. Non rounded correlator: {1.98, 1.02, 1.00, 1.00, 1.00, 7.08×10^{-17}}

Eigenvalues. Rounded correlator: {1.98563, 1.01533, 1.00, 1.00, 1.00, -0.00095}

Now we have complete information to formulate the result:

1. For the vector {x1,x2,x3,x4,x5} we have correct estimates for the adjusted values of components, their standard deviations and positive definite correlation matrix which may be uniformly rounded to be presented in integers % ;

2. For the extended vector {x1,x2,x3,x4,x5,xU} we have correct estimates for the adjusted values of components, their standard deviations and positive semi-definite correlation matrix expressed with 16 digits to the right of decimal point.

To express results in a more compact forms the directed rounding procedures should be designed and implemented to preserve the properties of the correlator.

Summary

We have problems with numerical expression and presentation of correlated multidimensional data in publications and in computer readable files.

These problems are common in the whole scientific community and originated in the absence of the widely accepted standard to express numerically the multidimensional correlated data.

As metrologists moves too slow, we propose PDG to workout the needed standard and implement it in PDG activity and in PDG publications: traditional and electronic. The physics community will follow PDG. Physics authors will produce data of high metrological quality.

COMPAS group will participate in this activity if it will be accepted by PDG collaboration.

We, PDG, will not stay alone! The movement to standardize the quality of e-data has started already

American Nuclear Society

Why Should Companies Support Standards Development?

Written by Suriya Ahmad for ***Nuclear Standards News***
(Vol. **33**, No. 6; **Nov-Dec, 2002**).

As professionals working in the nuclear energy industry, we are committed to the benefits that nuclear technology provides humankind.

The future of nuclear energy depends on maintaining a strong safety record, economics, and effective waste management.

So, how does the industry gather and maintain the information needed to meet these goals? It is done, in a large part, through the use of **voluntary consensus standards**.

Voluntary consensus standards represent the best knowledge of the field. They are written by groups of volunteers who are regarded as the technical experts in the nuclear energy industry.

THE NATIONAL ACADEMIES PRESS | THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

SHARING PUBLICATION-RELATED DATA AND MATERIALS: RESPONSIBILITIES OF AUTHORSHIP IN THE LIFE SCIENCES
Washington, D.C. **2003** www.nap.edu

DATA SCIENCE Journal CODATA

Special Issue "Thousand Words"

David R. Lide, Data quality - more important than ever in the Internet age
CODATA DSJ **6 (2007)** 154-155

Ray P. Norris, How to Make the Dream Come True: The Astronomers' Data Manifesto
CODATA DSJ **6 (2007)** S116-S124

Vladimir V. Ezhela, Multimeasurand ISO GUM Supplement is Urgent
CODATA DSJ **6 (2007)** S676-S689, **Errata, DSJ 7 (2008) E2-E2**

Shuichi Iwata, SCIENTIFIC "AGENDA" OF DATA SCIENCE
CODATA DSJ **7 (2008)** 54-56

Conclusion

- Scientific measured data to prove the discovery of a phenomenon and data needed to use the phenomenon in practice are the data of different quality.
- Current practice to select scientific papers for publication is not enough to assure the scientific data to be of metrological quality.
- Current practice of selecting measured data from publications to assess them as the reference data for scientific and industrial applications is too soft to prevent proliferation of incomplete or corrupted data.
- Necessity of the special standardized procedures and means to “sieve and seal” the measured scientific data to be qualified as data of metrological quality and recommended for publication is argued.
- It is time to think on the extended form of the scientific publication, namely: any paper, reporting measured (or evaluated) data, should be accompanied by data files where data are completely presented in computer readable form of sufficient numeric precision to preserve the results obtained.