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

The main sources of the corrupted data are:

- **Over-rounding;**
- **Usage of improper uncertainty propagation laws;**
- **Absence of the in/out data quality assurance programs in traditional and electronic publishing processes.**

As a rule, published multivariate data are damaged by over-rounding !!!

What is the over-rounding of multidimensional data?

Let us transform the "Greek" random vector with its scatter region

$$\begin{pmatrix} \zeta \\ \eta \end{pmatrix} = \begin{pmatrix} (\sqrt{2})\cdot(1.500 & 0.100) \\ (\sqrt{2})\cdot(0.345 & 0.001) \end{pmatrix}, \qquad r(\zeta, \eta) = \begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{pmatrix}$$

by 45 degrees rotation

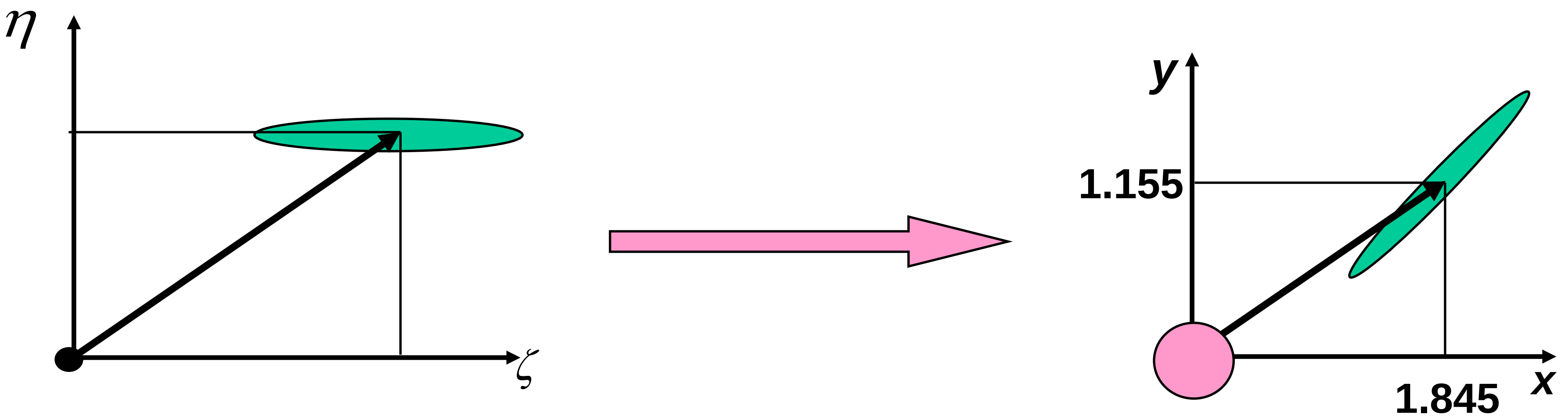
$$\begin{pmatrix} x = (\zeta+\eta)\,/\,(\sqrt{2}) \\ y = (\zeta-\eta)\,/\,(\sqrt{2}) \end{pmatrix}$$

to the "Latin" vector

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.845 & 0.100 \\ 1.155 & 0.100 \end{pmatrix}, \qquad r(x, y) = \begin{pmatrix} 1.00000 & 0.9998 \\ 0.9998 & 1.0000 \end{pmatrix}$$

Let us recall how data could be corrupted after this simplest data transformation

1. True calculations, qualitatively true picture



x = 1.845(100)

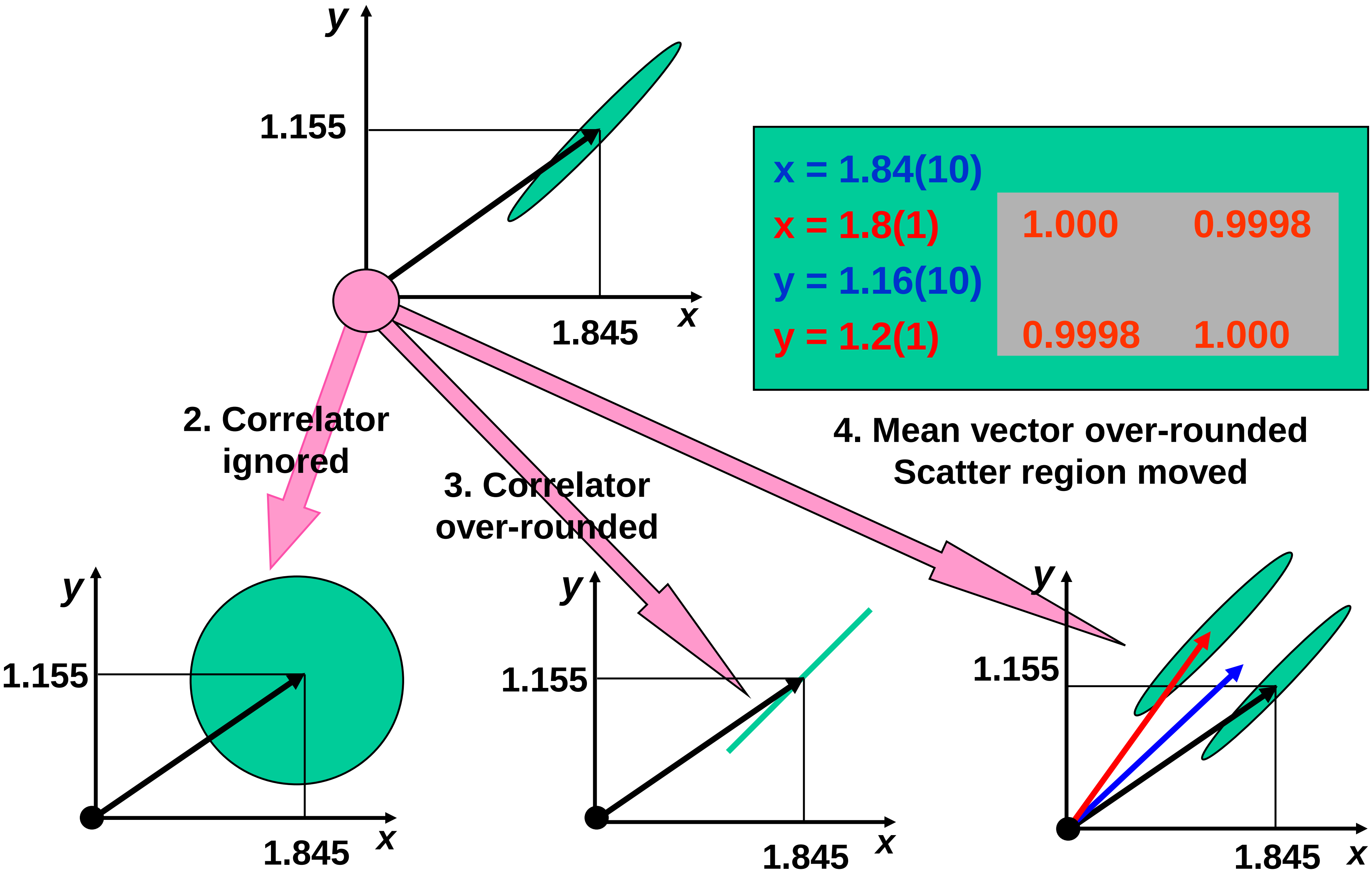
y = 1.155(100)

mean(uncertainty)

$$\begin{matrix} 1.000 & 0.9998 \\ 0.9998 & 1.000 \end{matrix}$$

correlator

All variants of correlated data corruption are copiously presented in scientific, educational, and technical resources



To escape corruption the safe rounding or directed rounding procedures are unavoidable!

Inputs from matrix theory for safe rounding:

Weil's theorem (see [29], [31]): Let $C = A + B$, where $A, B, C \in R^{n\times n}$ – symmetric matrices and $(\alpha_1 \leq \alpha_2 \cdots \leq \alpha_n)$, $(\beta_1 \leq \beta_2 \cdots \leq \beta_n)$, $(\gamma_1 \leq \gamma_2 \cdots \leq \gamma_n)$ their eigenvalues correspondingly.
Then $\forall i$ the following inequalities are valid

$$\alpha_i + \beta_{min} \leq \gamma_i \leq \alpha_i + \beta_{max} \ . \tag{9}$$

Gershgorin's theorem ([29], [30], [31]): Every eigenvalue α_i of the matrix A belongs to the interior of one of the circles

$$|A_{ii} - \alpha_i| \leq \sum_{j=1}^{n} |A_{i\neq j}| \ . \tag{10}$$

Schur's theorem ([31]): Let matrix $B \in R^{n\times n}$ is symmetric with values of the diagonal elements $b_1 \leq b_2 \leq \cdots \leq b_n$ (in any order) and eigenvalues $\beta_1 \leq \beta_2 \cdots \leq \beta_n$, then $\forall k \leq n$

$$\sum_{i=1}^{k} \beta_i \leq \sum_{i=1}^{k} b_i \ . \tag{11}$$

The equality take place only for $k = n$.

On the basis of Weil, Gershgorin, and Schur spectral theorems we propose the following **safe rounding thresholds** for:

Correlation coefficients

$$A \geq A_{\min}^{th} = \left\lceil \log_{10}\left(\frac{n-1}{2 \cdot \lambda_{\min}} \right) \right\rceil$$

Unitless uncertainties

$$P_U^{th} \geq \left\lceil \frac{1}{2} \log_{10}\left(\frac{n}{4 \cdot \lambda_{\min}} \right) \right\rceil$$

Unitless mean values

$$A_i \geq A_i^V = \left\lceil \frac{1}{2} \log_{10}\left(\frac{n}{4 \lambda_{\min} \left(U_i / [unit_i] \right)^2} \right) \right\rceil$$

where $\lambda_{\min}$ is the minimal eigenvalue of the correlator

Nonlinear Differential Uncertainties Propagation Law

C_i , $\langle \delta C_a , \delta C_b \rangle$ → $F_k(C_i)$, $\langle \delta F_m , \delta F_n \rangle$

I variables ***D*** functions

$$\left\langle \delta F_i , \delta F_j \right\rangle =$$

$$\sum_{k,l=1}^{T} \frac{1}{k!\,l!} \frac{\partial^k F_i}{\partial c_{\alpha_1} \ldots \partial c_{\alpha_k}} \left\langle \delta c_{\alpha_1} \ldots \delta c_{\alpha_k} , \delta c_{\beta_1} \ldots \delta c_{\beta_l} \right\rangle \frac{\partial^l F_j}{\partial c_{\beta_1} \ldots \partial c_{\beta_l}}$$

Correlator $\langle \delta F_m , \delta F_n \rangle$ is positive definite if $\langle \delta C_a , \delta C_b \rangle$ is positive definite and integers *I*, *D*, *T* obey inequality:

$$D \leq \frac{(I + T)!}{I! \times T!} - 1$$

In May 2005 the accurate data on basic FPC-2002 appeared for the first time. This gave us possibility for the further investigation of the derived FPC-2002 *{me, e, 1/α(0), h}* :

Linear Differential UPL (default machine precision)

2002: { 2.99825, 1.00175, 9.95751E-10, 9.23757E-17 }

Linear Differential UPL (SetPrecision[expr,30])

2002: { 2.99825, 1.00175, 9.95751E-10, -6.95096E-35 }

Non-Linear Differential UPL (second order Taylor polynomial) (SetPrecision[expr,100])

2002: { 2.99825, 1.00175, 9.95751E-10, 2.86119E-15 }

Comparison with CODATA recommended values of derived FPC-2002 {*me, e, 1/α(0), h*}

1. Insert values of the basic constants from LSA files into formulae

$$me = \frac{2R_\infty \cdot h}{c \cdot \alpha^2} = 9.109382551053865E\text{-}31$$

$$e = \sqrt{\frac{2 \cdot \alpha \cdot h}{\mu_0 \cdot c}} = 1.6021765328551825E\text{-}19$$

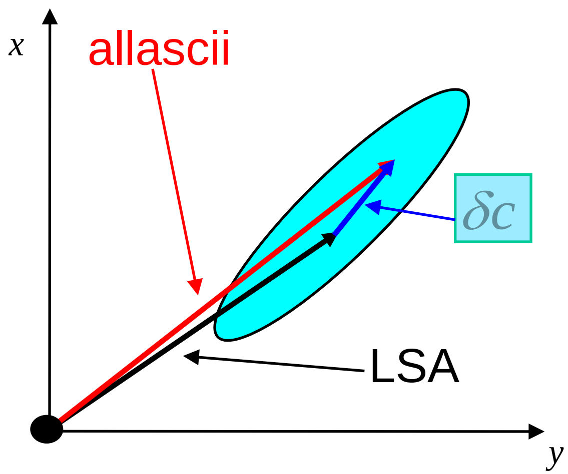
2. Biases were calculated supposing the multi-normal distribution for basic FPC. They are much less than corresponding standard deviations

	me	*e*	*1/a(0)*
bias	2.4943**E-66**	-2.6186**E-58**	1.7918**E-36**
sigma	1.5575**E-37**	1.791**8E-36**	**5.0E-7**

Where is the end of the rounded vector of the basic FPC-2002?

The end of the rounded vector should belong to the non-rounded scatter region.

To characterize the deviation we use the quadratic form χ^2



$$\chi^2 = \sum_{i,j}^{n} \delta c_i \cdot [\mathrm{cov}]^{-1}_{ij} \cdot \delta c_i$$

δc = c(allascii) –c(LSA)

$$\mu_i = \delta c_i / \sigma_i$$

Rounded vector belongs to non-rounded scatter region if:

$$\chi^2 = \sum_{i,j}^{n} \mu_i \cdot [cor]^{-1}_{ij} \cdot \mu_j < 1$$

We have 22 constants for which NIST give both allascii (rounded) and LSA "non-rounded" data for this test:

$$\chi^2 \approx 0.06$$!!!

But where is the end of the rounded vector for derived FPC-2002 ?

FPC	Our calculations with DifUPL(2,4,2)	Allascii (NIST-2002)	IMPROBABLE !!!
me	9.109382551053865E-31	9.1093826 E-31	
e	1.6021765328551828E-19	1.60217653 E-19	
1/α(0)	137.035999105576373	137.03599911	
h	6.626069310828000E-34 (LSA)	6.6260693 E-34	

$$\chi^2 = 2.18E+10$$

Thus, we see that the values of the derived vector components {*me,e, 1/α(0)*} presented on the NIST site in allascii.txt file are improbable!!!

The vector is out of the scatter region for the 10^{10} standard deviations

due to improper uncertainty propagation and 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

Some examples of “current doubtful practice” to express and publish measured data and how correlated data are reflected in RPP database

- **Physics Letters B288 (1992) 373,**

CERN-LEP-OPAL Experiment

- **Guide to the Expression of Uncertainty in Measurement**

(ISO GUM,1995)

- **Physical Review D55 (1997) 2259; D58 (1998) 119904E,**

CESR-CLEO Experiment

- **Reviews of Modern Physics 72 (2000) 351,**

CODATA recommended values of the FPC: 1998

- **Physics Letters B519 (2001) 191,**

CERN-LEP-L3 Experiment

- **European Physical Journal C20 (2001) 617,**

CERN-LEP-DELPHI Experiment

- **Nuclear Physics A729 (2003) 337,**

The AME2003 atomic mass evaluation (II)

- **Reviews of Modern Physics 77 (2005) 1,**

CODATA recommended values of the FPC: 2002

- **Physics Reports 421 (2005) 191,**

CERN-LEP-ALEPH Experiment

- **Physical Review D73 (2006) 012005,**

SLAC-PEP2-BABAR Experiment

- **Journal of Physics G33 (2006) 1**

Review of Particle Physics

- **European Physical Journal C46 (2006) 1,**

CERN-LEP-DELPHI Experiment

- **Physical Review C73 (2006) 044603,**

R-matrix analysis of Cl neutron cross sections up to 1.2 MeV

- **Physical Review D74 (2006) 014016,**

BINP-VEPP-2M-SND Experiment

- **Reviews of Modern Physics 80 (2008) 633,**

CODATA recommended values of the FPC: 2006

Physics Letters B288 (1992) 373

Experiment CERN-LEP-OPAL

Measurement of the τ topological branching ratios at LEP

"5. Summary and discussion

The inclusive branching ratios of the τ lepton to one, three and five charged particle final states are measured to be

B_1 = 84.48 +- 0.27 (stat) +- 0.23 (sys) %,
B_3 = 15.26 +- 0.26 +- 0.22 % and
B_5 = 0.26 +- 0.06 +- 0.05 % respectively.

These measurements have been obtained from a fit where **$B_1 + B_3 + B_5$ is constrained to equal one.**

The correlations between the fitted branching ratios are given by the matrix

$$\rho = \begin{pmatrix} 1. & -0.97 & -0.15 \\ -0.97 & 1. & -0.07 \\ -0.15 & -0.07 & 1. \end{pmatrix}."$$

Eigenvalues of this matrix are

{ **1.9677, 1.0118, 0.0205** }

Rounding Threshould = 2

If $\rho = \rho$(stat), it should be degenerate, but it is positive definite!

It is possible to restore the "true" statistical correlator from data on statistical errors, if they were obtained by the constrained fit :

$(B_1 + B_3 + B_5 = 1)$.

Indeed, in this case

$$\rho^{(stat)}_{mn} = (\sigma_k^2 - \sigma_m^2 - \sigma_n^2) / (2\, \sigma_m \sigma_n) , \; (k \neq m \neq n) = (1,3,5).$$

Inserting data on the statistical errors we will obtain a "true" correlator

$$\rho^{(stat)} = \begin{pmatrix} 1 & -0.975071 & -0.274691 \\ -0.975071 & 1 & 0.054487 \\ -0.274691 & 0.054487 & 1 \end{pmatrix}$$

with eigenvalues 2.02838, 0.97617, 3.46132E-17, where the minimal eigenvalue should be treated as zero (it is close to "default precision" which is 16 "significant digits").

Thus, the obtained matrix is degenerate and differs strongly out of the OPAL matrix.

From the systematic errors budget, taken from Table 7 of the paper

	Observables		
Sources	0.14	0.13	0.35
	0.12	0.12	0.12
	0.10	0.10	0.027
	0.10	0.10	0.00

we can calculate the covariance matrix of systematic uncertainties

Covariance matrix		
0.054	0.0526	0.00904
0.0526	0.0513	0.00869
0.00904	0.00869	0.002098

Adding it to the "true" statistical covariance matrix we will obtain the covariance matrix for the combined stat. and syst. errors

Total covariance matrix		
0.1269	-0.01585	0.00459
-0.01585	0.1189	0.00954
0.00459	0.00954	0.005698

Total correlation matrix		
1.	-0.129035	0.170695
-0.129035	1.	0.366519
0.00459	0.366519	1.

Eigenvalues of the total correlator are as follows
{ 1.36933, 1.09429, 0.536376 }

Now, it seems, we have complete presentation of OPAL result:

- **estimates of mean values,**
- **estimates of statistical and systematic covariances with true properties;**
- **estimates of the total covariances and correlations with quoting the data quality parameters (precision of calculations and rounding thresholds).**

Physical Review D55 (1997) 2559

Experiment CESR-CLEO

Erratum: Experimental tests of lepton universality in τ decay.

Phys. Rev. D58 (1998) 119904

TABLE XII. Correlation coefficients between branching fraction measurements.

C_τ	B_e	B_μ	B_h	B_μ/B_e	B_h/B_e
B_e	1.00	0.50	0.48	-0.42	-0.39
B_μ		1.00	0.50	0.58	0.08
B_h			1.00	0.07	0.63
B_μ/B_e				1.00	0.45
B_h/B_e					1.00

Unreliable !!!

Eigenvalues of this matrix are as follows:

$$(2.1735,\ 1.7819,\ 1.0550,\ -0.0075,\ -0.0028)$$

So, the Erratum to the Erratum is needed

European Physical Journal C20 (2001) 617
Experiment CERN-LEP-DELPHI

A Measurement of the τ Topological Branching Ratios

$$\begin{bmatrix} B_1 \\ B_3 \\ B_5 \end{bmatrix} = \left(\begin{bmatrix} 0.85316 \pm 0.000929_{stat} \pm 0.000492_{syst} \\ 0.14569 \pm 0.000929_{stat} \pm 0.000477_{syst} \\ 0.00115 \pm 0.000126_{stat} \pm 0.000059_{syst} \end{bmatrix}, \begin{bmatrix} 1.00 & -0.98 & -0.08 \\ -0.98 & 1.00 & -0.08 \\ -0.08 & -0.08 & 1.00 \end{bmatrix} \right)$$

Unreliable !!!

Published correlator is incorrect and over-rounded.

Our calculations, based on data presented in the paper give the "correct" safely rounded correlator:

$$\begin{bmatrix} 1. & -0.9924 & -0.0848 \\ -0.9924 & 1. & -0.0335 \\ -0.0848 & -0.0335 & 1. \end{bmatrix}$$

It seems that an Erratum to the paper is needed, because of the over-rounding and improper uncertainty propagation

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(%)
$\tau^- \to h^-$	$\geq 0K^0\ \nu_\tau$	12.780 ± 0.120 ± 0.103
$\tau^- \to h^-\ \pi^o$	$\geq 0K^0\ \nu_\tau$	26.291 ± 0.201 ± 0.130
$\tau^- \to h^-\ 2\pi^o$	$\geq 0K^0\ \nu_\tau$	9.524 ± 0.320 ± 0.274
$\tau^- \to h^- \geq 1\pi^o$	$\geq 0K^0\ \nu_\tau$	37.218 ± 0.155 ± 0.116
$\tau^- \to h^- \geq 2\pi^o$	$\geq 0K^0\ \nu_\tau$	10.927 ± 0.173 ± 0.116
$\tau^- \to h^- \geq 3\pi^o$	$\geq 0K^0\ \nu_\tau$	1.403 ± 0.214 ± 0.224
$\tau^- \to 3h^\pm$	$\geq 0K^0\ \nu_\tau$	9.340 ± 0.090 ± 0.079
$\tau^- \to 3h^\pm\ \pi^o$	$\geq 0K^0\ \nu_\tau$	4.545 ± 0.106 ± 0.103
$\tau^- \to 3h^\pm \geq 1\pi^o$	$\geq 0K^0\ \nu_\tau$	5.106 ± 0.083 ± 0.103
$\tau^- \to 3h^\pm \geq 2\pi^o$	$\geq 0K^0\ \nu_\tau$	0.561 ± 0.068 ± 0.095
$\tau^- \to 5h^\pm$	$\geq 0K^0\ \nu_\tau$	0.097 ± 0.015 ± 0.005
$\tau^- \to 5h^\pm \geq 1\pi^o$	$\geq 0K^0\ \nu_\tau$	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 ***K⁰*** *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
$\tau^- \to h^- \pi^0 \nu_\tau$	25.740 ± 0.201 ± 0.138
$\tau^- \to h^- 2\pi^0 \nu_\tau$	9.498 ± 0.320 ± 0.275
$\tau^- \to h^- \geq 1\pi^0 \nu_\tau$	36.641 ± 0.155 ± 0.127
$\tau^- \to h^- \geq 2\pi^0 \nu_\tau$	10.901 ± 0.173 ± 0.118
$\tau^- \to h^- \geq 3\pi^0 \nu_\tau$	1.403 ± 0.214 ± 0.224
$\tau^- \to 3h^\pm \nu_\tau$	9.317 ± 0.090 ± 0.082
$\tau^- \to 3h^\pm \pi^0 \nu_\tau$	4.545 ± 0.106 ± 0.103
$\tau^- \to 3h^\pm \geq 1\pi^0 \nu_\tau$	5.106 ± 0.083 ± 0.103
$\tau^- \to 3h^\pm \geq 2\pi^0 \nu_\tau$	0.561 ± 0.068 ± 0.095
$\tau^- \to 5h^\pm \nu_\tau$	0.097 ± 0.015 ± 0.005
$\tau^- \to 5h^\pm \geq 1\pi^0 \nu_\tau$	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.

Physics Letters B519 (2001) 191

Experiment CERN-LEP-L3

Measurement of the topological branching fractions of the τ lepton at LEP

"After combination of the systematic uncertainties the results for the branching fractions of the τ lepton decays into one, three and five charged particle final states are:

$$B(\tau \to (\text{1-prong})) = 85.274 \pm 0.105 \pm 0.073\%,$$
$$B(\tau \to (\text{3-prong})) = 14.556 \pm 0.105 \pm 0.076\%,$$
$$B(\tau \to (\text{5-prong})) = 0.170 \pm 0.022 \pm 0.026\%,$$

where the first uncertainty is statistical and the second is systematic."

Unfrortunately, there are no comments on the properties of the stat., or syst., or combined uncertainty matrices in the section describing the final results.

But in **pdgLive(2008)** we have some indication that there might be further comments from "L3-verifier"

From the footnotes to the measurements in the corresponding data blocks it is possible to form the correlation matrix

Source	$B(1\text{-prong})$	$B(3\text{-prong})$	$B(5\text{-prong})$
$B(1\text{-prong})$	**1.0**	**−0.978**	**−0.082**
$B(3\text{-prong})$		**1.0**	**−0.19**
$B(5\text{-prong})$			**1.0**

that is named as "correlations between measurements" there, and can be interpreted as the correlations of the total uncertainties. It does not coinside with the statistical correlator presented in the paper (Table 4).

"In the fit the constraint **$B(1) + B(3) + B(5) = 1$** is applied and the sum of N^i_{exp} is constrained to the number of observed τ decays. The following results are obtained:….."

"Table 4

The correlation coefficients obtained from a fit of the topological branching fractions

Source	$B(1\text{-prong})$	$B(3\text{-prong})$	$B(5\text{-prong})$
$B(1\text{-prong})$	**1.0**	**−0.978**	**−0.082**
$B(3\text{-prong})$		**1.0**	**−0.127**
$B(5\text{-prong})$			**1.0**

"

Our calculations, analogous to OPAL case based on the statistical errors from constrained fit presented in the paper give the "correct" safely rounded statistical correlator:

$$\rho^{(stat)} = \begin{pmatrix} 1 & -0.97805 & -0.104762 \\ -0.97805 & 1 & -0.104762 \\ -0.104762 & -0.104762 & 1 \end{pmatrix}$$

With eigenvalues: 1.97805, 1.02195, -1.01856E-17

From the systematic errors budget,
taken from Table 5 of the paper

	Observables		
	0.048	0.052	0.024
	0.01	0.01	0.001
	0.01	0.01	0.001
Sources	0.011	0.011	0.001
	0.035	0.035	0.003
	0.012	0.012	0.001
	0.017	0.017	0.004
	0.032	0.032	0.007

Systematic covariance matrix		
0.005307	0.005499	0.001592
0.005499	0.005707	0.001688
0.001592	0.001688	0.000654

With systematic errors as in the paper
0.0728492, 0.0755447, 0.0255734

Adding it to the "true" statistical covariance matrix we will obtain the covariance matrix for the combined stat. and syst. errors

Total covariance matrix		
0.016332	-0.005284	0.001350
-0.005284	0.016732	0.001446
0.001350	0.001446	0.001138

Total correlation matrix		
1.	-0.319646	0.313143
-0.319646	1.	0.331378
0.00459	0.331378	1.

Eigenvalues of the total correlator are as follows

{ 1.33208, 1.31076, 0.357163}

Now we have complete presentation of L3 result:

- **estimates of mean values;**
- **statistical and systematic covariances with true properties;**
- **estimates of the total covariances and correlations with quoting the data quality parameters (precision of calculations and rounding thresholds).**

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)

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 $\chi^2 = 1.9$ for 2 degrees of freedom.
The following *off-diagonal* array elements are the correlation coefficients $<\delta x_i \delta x_j> / (\delta x_i \cdot \delta x_j)$, 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.

$$\begin{array}{cccc} x_1 & 100 & & \\ x_2 & \underline{-100} & 100 & \\ x_4 & \underline{-1} & \underline{-0} & 100 \\ & x_1 & x_2 & x_4 \end{array}$$

Eigenvalues of the 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

Eigenvalues of the "URL-rounded correlator" {1.99996, 1.00004, -1.02849×10^{-10}}

“**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 RPP-2008

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

$$\mathbf{x1 + x2 + x4 = 1,}$$

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

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

are the correlations. Inserting corresponding $\boldsymbol{\sigma_m}$ data from pdgLive we obtain:

	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 why the obtained estimates of the correlator differs from that of presented in the RPP and propose slightly modified procedure for the constrained fit

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
Γ(*undetected*)/Γ(total)	0.0	6.0×10^{-4}	1-x1-x2-x3-x4-x5

7 measurements, 5 parameters

KOMПАСOMPAS

Proposal for “new” forms of constrained fits

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

	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}

Minimum(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

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.

Metrology in Fundamental Science is Urgent

Publication of data in refereed scientific journals does not assure the quality of the reported data:

- Absence of the in/out data quality assurance programs in traditional and electronic publishing processes;
- Incomplete presentation of measured data;
- Data corruption caused by publication space constraints (over-rounding);
- Lack in duality: human/computer usability;
- Multivariate data presented in publications often are corrupted

Presentation of data collected from publications and assessed by experts does not assure the quality of the integrated and assessed data:

- Absence of the in/out data quality assurance programs in loading/extracting data into/from databases;
- Incomplete presentation of measured data;
- Data corruption caused by too tight formats to store numbers (over-rounding);
- Lack in duality: human/computer usability;

Existing International and National Guides to express and report measured data are formulated for one measured quantity only.