УДК 539.2

V.I. Belousov, I.I. Degtyarev, V.V. Ezhela, O.V. Zenin, Yu.V. Kiyanov, V.N. Larin,
K.S. Lugovsky, S.B. Lugovsky, F.F. Tikhonin, N.P. Tkachenko

Institute for High Energy Physics, Protvino, Russia

NECESSITY AND FEASIBILITY OF INTER-UNIVERSITIES KNOWLEDGE GENERATION BASE FOR RESEARCH AND EDUCATION IN GENERAL AND FUNDAMENTAL PHYSICS

The necessity and feasibility to create an Inter-Universities Knowledge Generation Base (KGB) in the field of the fundamental physics phenomenology on the cooperative basement of existing open access information resources of scientific and industrial data centers is argued. It is deemed that such KGB, being used in the education and research at physics universities, will be helpful to make (step by step) the new measured "discovery data" to be "the discovery data of metrology quality". From the other side this will be helpful for Metrology to assimilate and standardize timely the new measuring methods and new data structures (emerging in Science) for possible applications in science, e-publishing, and industry.

***Keywords**: Quality of Scientific Data, Quality of e-Published Measured Data, Peer Reviewing Guide for e-Science, Knowledge Generation Base.*

Introduction

The quality of scientific data products permanently declared as the crucial point for all scientific inferences (See recent papers on the subject [1 – 6] and references therein). But the current practice of expressing and representing measured or evaluated data in publications and in the open access scientific and industrial data resources are far from being the practice with quality assurance of measured data (See refs. [7 – 21]). This brief survey of the data quality published in peer reviewed scientific journals of high impact factor clearly shows that we have a growing number of evidences that the current practice of knowledge generation, transfer, and preservation is obsolete and wasteful. This obsoleteness is one of the major factors that are slowing down the scientific and technological developments. Let us look at the crude "anatomy" of this obsoleteness:

1. Metrology of fundamental science is urgent

For the first sight the title of the section seems nonsense. Indeed, the mission of the science is to discover and decipher the primary entities of Nature and their relations. Scientists need as much freedom as currently possible in their searches and methods. The "uniformity of measurements" and "uniformity of the measured data quality" being imposed on scientific activity could restrict creativity of science. But the "primary entities and relations" should be standardized somehow to compose the "alphabet and grammar" of Nature to be used coherently in SEMI activities and in Society. Formalizing, maintenance, and tracing developments of this "alphabet and grammar" are the mutual mission of Science and Metrology.

Nevertheless, the critical comments to the measured data presentations in recent scientific publications [7 – 21] show that the situation is unsatisfactory. Presented critical comments can be grouped into generalized items as follows:

a) ignorance of existing metrology recommendations by scientists and publishers resulted in incomplete data representations;

b) absence of the in/out data quality assurance programs in traditional and electronic publishing processes;

c) data corruption caused by publication space constraints (over-rounding);

d) data corruption caused by too tight formats to store numbers (over-rounding);

e) lack in duality: human/computer usability of the reported data.

The items listed above show the apparent lack of metrological expertise of the reported measured scientific data in spite of the long existing general recommendations of CODATA Guide (1973) [22]:

"… It is a statement of minimum information that is needed to ensure that the reader can understand the quantitative data, can assess their precision and accuracy, and can recalculate the results when values for auxiliary data change.

The authors of a paper has the primary responsibility for providing the reader with the type of information outlined in this Guide.

The Guide also provides journal editors and referees with a set of consistent, considered criteria for judging the completeness and acceptability of papers in so far as the reporting of numerical quantities are considered. The recommendations reflect the experience of data evaluators: hence, adherence to them will permit evaluators to consolidate the author's results with existing data and facilitate their incorporation in critical compilations. …" (An excerpt from Introduction of [22])

The more elaborated recommendations are presented in the ISO GUM (1995) [8].

So, the fact of publication of measured data in refereed scientific journals does not assure the quality of the reported data, the representation of data in the integrated collections, gathered from publications and assessed by experts does not assure the quality of the integrated data.

We can attribute these sad evidences to the absence of proper metrological education in scientific community and from the other side to the lack of knowledge of the metrological problems of fundamental science in metrology community. Indeed, there are no consensus guides to express and report multidimensional jointly measured data that are urgently needed in SEMI activities**.** Existing international and national guides to express and report the measured or evaluated data are formulated in detail for one measured quantity only.

It should be noted, however, that there is the special Working Group of the Joint Committee for Guides in Metrology at the BIPM that takes care of the further developments of the GUM with working plan for the next few years as follows:

ISO/IEC FDGuide 98-1 Uncertainty of measurement -- Part 1: Introduction to the expression of uncertainty in measurement;

ISO/IEC NP Guide 98-2 Uncertainty of measurement - Part 2:Concepts and basic principles;

ISO/IEC NP Guide 98-3 Uncertainty of measurement - Part 3: Guide to the expression of uncertainty in measurement (GUM:1995);

ISO/IEC Guide 98-3/Suppl 1 Propagation of distributions using a Monte Carlo method;

ISO/IEC Guide 98-3/NP Suppl 2 Models with any number of output quantities ISO/IEC Guide 98-3/NP Suppl 3 Modelling;

ISO/IEC NP Guide 98-4 Uncertainty of measurement – Part 4: Role of measurement uncertainty in conformity assessment.

ISO/IEC NP Guide 98-5 Uncertainty of measurement – Part 5: Applications of the least-squares method.

It is hard to imagine when these documents will be available for the interested specialists. The GUM Suppl 1 will be issued, probably, this year (after 15 years of the first edition of the GUM). It is natural: the broader domain of applicability of the particular standard the slower it evolves to fit the changing environment (too many nuisance parameters to reach consensus).

There is an interesting (but prone of danger) initiative to make standards development faster was born recently in American Nuclear Society:

"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. "An excerpt from the article: "Why Should Companies Support Standards Development?", written by Suriya Ahmad for *Nuclear Standards News* (Vol. 33, No. 6, Nov-Dec, 2002).

The rest of this note is to argue that there is a way to consolidate the metrological efforts in SEMI activities by transferring the part of the activity of the national and international data centers to education researchers via creation the network of university hosted knowledge generation bases (KGB) at leading profile universities. To our understanding the fundamental atomic-, nuclear-, particle- and astrophysics are the most advanced fields to begin creation the new generations of scientists and engineers educated and trained in science and metrology and motivated for creation the best currently possible scientific data products.

2. KGB for research and education

The Knowledge Generation Base is deemed as the repository for the sequential versions of the Computerized Objects of Knowledge ─ COK(n) with clearly defined domain of applicability (DA).

The COK(n) should be composed of a few (more then seven) interoperable modules designed in an advanced computer algebra (CA) system:

ExpData(n) maintained DA-related collection of assessed experimental data (collected from literature or extracted from scientific data resources and passed through the data quality control subsystem).

ModeLib(n) ─ maintained DA-related library of theoretical model codes (collected from literature and coded or extracted from scientific code libraries) to calculate values of observable quantities to be compared with data from ExpData(n).

Evaluator(n) ─ module for the cross-assessments of data from ExpData(n) and relevant models from ModeLib(n) via adjustments of models to the experimental data for determination the best values of the model parameters and the region of applicability of the particular model with predefined confidence level.

Sketcher(n) ─ module to store the final results of the cross-assessments for all models with non-empty region of applicability.

Predictor(n) ─ module to calculate predictions for expected results of new measurements with estimates of precisions to resolve the competitive models of highest ranks and to prepare codes of highest rank models to refresh ModeLib.

Reporter(n) ─ module to prepare reports on physics results of cross-assessments to be published in peer reviewed physics journals with representation of numerical results with current level of metrological completeness and quality.

Motivator(n) ─ module to select new scientific publications that possibly contain new knowledge relevant to COK by counting for each new publication the number of it's references that are present in the target sample of references stored in the Motivator (the references to the papers, knowledge from which, were used to compose ExpData, or ModeLib, or Evaluator modules). Assimilation of selected by co-citation new publications in the COK(n) will give stimuli for creation the next version COK(n+1).

The proposed scheme is close to that of used (partly) in the well known in physics community data centers: FCDC, AMDC, NNDC, PDG but (unfortunately) without requirements to stick to the current level of the metrological completeness and quality in the spirit of CODATA Guide(1973) and ISO GUM(1995) (most probably, due to the lack of resources and manpower).

This scheme is suitable for introducing the Metrology and Industry approaches to the measured data quality into Science via Education and e-Publishing. Physics students participating in development and maintenance of the COKs will be well educated in the current metrological requirements to the measured data quality from the very beginning of their professional education and will bring corresponding knowledge to all physics areas where they will work in SEMI-activities. Indeed, if part of activities of the scientific and corporative physics data centers, being structured in COKs, will be transferred to physics universities (on the collaborative grounds), this will give them possibilities:

a) to introduce the technology of scientific data bases and knowledge generation bases into educational process on the level of the third year general physics courses (introductory parts of the atomic-, nuclear-, particle-, and astrophysics);

b) to take part in maintenance and development of the COKs, created in scientific or industrial data centers;

c) to create and maintain their own profile COKs, shared between the inter-universities KGB and data centers;

d) to collaborate with the metrology institutions and transfer university findings (obtained in due course of the COKs development and maintenance) that turned out to be useful for metrology development;

e) to start to create the new specialization in physics: physicist-systematist as the specialist for e-knowledge generation, knowledge management, knowledge preservation and transfer to future generations.

3. How to create a computerized object of knowledge

Approximate workflow of creating the COK and to organize its maintenance could be as follows:

1. Select subfield of interest in physics phenomenology to create **COK**.

2. Extract corresponding experimental data from literature or from existing data resources and create the **ExpData(n)**.

3. Collect from literature or from existing resources the relevant models and create the **ModeLib(n)** to calculate observable quantities and parameters evaluation.

4. Select or create the system to compare theoretical models with experimental data and create the Evaluator(n) for cross-assessments of data and models.

5. Create module **Sketcher(n)** to store results of cross-assessments of data and models.

6. Conduct qualifying comparisons the description of data from **ExpData(n)** by each model and select models with nonempty region of applicability at predefined confidence level.

7. Conduct ranking of selected models with respect to predefined data description quality indicators and form the **ModeLib(n)** as the library of models adjusted to **ExpData(n)**.

8. Select a few competitive models of highest ranks and form module **Predictor(n)**.

9. Form the file of basic references used to create the **COK (n)** including the references to the published results on the current version of the **COK** and create or update module **Motivator(n)**.

10. Prepare texts representing physics results obtained in due course of forming the **COK** and send them for publication in peer reviewed journals.

11. Prepare technical descriptions of the created **ExpData(n)** and **ModeLib(Nn** to fix the version of the **COK** in the repository and to be used in updating interfaces of interoperability with adjacent C**OKs**. Prepare on-line helps for custodians and users.

12. Version of the **COK** is considered as fully prepared if obtained physics and methodic results are published.

13. Activate module **Motivator(n)** to gather stimuli for the next generation **COK(n+1)**.

Literature List

1. NAP publication (2003) "Sharing publication-related data and materials: Responsibilities of authorship in the life sciences", Washington, D.C. [Электронный ресурс]. – Режим доступа к док: http://www.nap.edu/catalog.php? record_id=10613.

2. Nicole M. Radziwill. Foundation for Quality Management of Scientific Data Products / Nicole M. Radziwill // Quality Management Journal. – 2006. – 13(2). – P. 7.

3. David R. Lide. Data quality - more important than ever in the Internet age / David R. Lide // CODATA DSJ. – 2007. – 6. – P. 154-155.

4. Ray P. Norris. How to Make the Dream Come True: The Astronomer's Data Manifesto / Ray P. Norris // CODATA DSJ. – 2007. – 6. – P. 116-124.

5. Vladimir V. Ezhela. Multimeasurand ISO GUM supplement is Urgent / Vladimir V. Ezhela // CODATA DSJ. – 2007. – 6. – P. 676-789. [Errata: CODATA DSJ, 7, E2-E2].

6. Shuichi Iwata. Scientific "Agenda" of data science / Shuichi Iwata // CODATA DSJ. – 2008. – 7. – P. 54-56.

7. Measurement of the Tau Topological Branching Ratios at LEP / P.D. Acton, et al. [OPAL Collaboration] // Phys.Lett. – 1992. – B 288. – P. 373.

8. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML. "Guide to the Expression of Uncertainty in Measurement", ISO/IEC Guide 98:1995, ISBN 92-67-10188-9, Second Edition. – 1995.

9. By CLEO Collaboration (A. Anastassov et al.) (1997) "Experimental Test of Lepton Universality in Tau Decay", Phys. Rev. D55, 2559, [Erratum-ibid (1998), D58, 119904].

10. Mohr P.J. CODATA recommended values of the fundamental physical constants: 1998 / P.J. Mohr, B.N. Taylor // Rev. Mod. Phys. – 2000. – 72. – P. 351.

11. Measurement of the topological branching fractions of the tau lepton at LEP / P. Achard, et al. [L3 Collaboration] // Phys. Lett. – 2001. – B519. – P. 189.

12. A measurement of the tau topological branching ratios / P. Abreu, et al. [DELPHI Collaboration] // Eur. Phys. J. – 2001. – C20. – P. 617.

13. Audi G. The Ame2003 Atomic Mass Evaluation (II). Tables, Graphs and References / G. Audi, A.H. Wapstra, C. Thibaul, // Nucl. Phys. – 2003. – A729. – P. 337.

14. Mohr P.J. CODATA recommended values of the fundamental physical constants: 2002 / P.J. Mohr, B.N. Taylor // Rev. Mod. Phys. – 2005. – 77. – P. 1.

15. S. Schael, et al. [ALEPH Collaboration] Branching ratios and spectral functions of tau decays: Final ALEPH measurements and physics implications / S. Schael, et al. [ALEPH Collaboration] // Phys. Rept. – 2005. – 421. – P. 191.

17. Review of particle physics / W.M. Yao, et al. [Particle Data Group] // J. Phys. G. – 2006. – 33. – P. 1.

18. A measurement of the tau hadronic branching ratios / J. Abdallah, et al. [DELPHI Collaboration] // Eur. Phys. J. – 2006. – C 46. – P. 1.

19. R-matrix analysis of Cl Neutron Cross Sections up to 1.2-Mev / R.O. Sayer, K.H. Guber, L.C. Leal, N.M. Larson, and T. Rauscher // Phys. Rev. – 2006. – C73. – P. 044603.

20. Study of the e+ e- --> eta gamma process with SND detector at the VEPP-2M e+ e- collider / M.N. Achasov, et al. // Phys. Rev. – 2006. – D74. – P. 014016.

21. Mohr P.J. CODATA recommended values of the fundamental physical constants: 2006 / P.J. Mohr, B.N. Taylor, D.B. Newell // Rev. Mod. Phys. – 2008. – 80. – P. 633.

22. Guide for the Presentation in the Primary Literature of Numerical Data Derived from Experiments / D. Gavrin, T. Golashvili, H.V. Kehiaian, N. Kurti, E.F. Westrum Jr. [CODATA Task Group on Publication of Data in the Primary Literature)] // CODATA Bulletin. – 1973. – 9.

Поступила в редколлегию 15.07.2009

Рецензент: д-р техн. наук, проф. И.П. Захаров, Харьковский национальный университет радиоэлектроники, Харьков.

НЕОБХІДНІСТЬ І ДОЦІЛЬНІСТЬ ЗАГАЛЬНОЇ МЕЖУНІВЕРСИТЕТСЬКОЙ БАЗИ ЗНАНЬ ДЛЯ ДОСЛІДЖЕННЯ І ОСВІТИ В ЗАГАЛЬНІЙ І ФУНДАМЕНТАЛЬНІЙ ФІЗИЦІ

В.І. Бєлоусов, І.І. Дегтярьов, В.В. Ежела, О.В. Зенін, Ю.В. Кіянов, В.Н. Ларін,
К.С. Луговський, С.Б. Луговський, Ф.Ф. Тіхонін, Н.П. Ткаченко

Аргументується необхідність і доцільність створення Загальної межуніверситетської бази знань в області феноменології фундаментальної фізики, заснованої на базі існуючого відкритого доступу до інформаційних ресурсів наукових і промислових баз даних. Вважається, що така база, використовувана в освіті і дослідженнях в університетах фізики, буде корисною для створення (крок за кроком) нової виміряної «відкритої інформації», яка буде «відкритою інформацією метрологічної якості». З іншого боку це буде корисним для метрології засвоювати і стандартизувати нові методи виміру і нову структуру інформації (що з'являється в науці) для можливого вживання в науці, електронних виданнях і промисловості.

Ключові слова: *якість наукової інформації, якість електронних видань з виміряною інформацією, керівництво по перегляду видалених наукових даних, загальна основа знань.*

НЕОБХОДИМОСТЬ И ЦЕЛЕСООБРАЗНОСТЬ ОБЩЕЙ МЕЖУНИВЕРСИТЕТСКОЙ БАЗЫ ЗНАНИЙ ДЛЯ ИССЛЕДОВАНИЯ И ОБРАЗОВАНИЯ В ОБЩЕЙ И ФУНДАМЕНТАЛЬНОЙ ФИЗИКЕ

В.И. Белоусов, И.И. Дегтярев, В.В. Ежела, О.В. Зенин, Ю.В. Киянов, В.Н. Ларин,
К.С. Луговский, С.Б. Луговский, Ф.Ф. Тихонин, Н.П. Ткаченко

Аргументируется необходимость и целесообразность создания Общей межуниверситетской базы знаний в области феноменологии фундаментальной физики, основанной на базе существующего открытого доступа к информационным ресурсам научных и промышленных баз данных. Считается, что такая база, используемая в образовании и исследованиях в университетах физики, будет полезной для создания (шаг за шагом) новой измеренной «открытой информации», которая будет «открытой информацией метрологического качества». С другой стороны это будет полезным для метрологии усваивать и стандартизировать новые методы измерения и новую структуру информации (появляющуюся в науке) для возможного применения в науке, электронных изданиях и промышленности.

Ключевые слова*: качество научной информации, качество электронных изданий с измеренной информацией, руководство по просмотру удаленных научных данных, общая основа знаний.*