

PROPERTIES OF NANOSTRUCTURES: DATA ACQUISITION, CATEGORIZATION, AND EVALUATION

AV Eletskaa^{1, 2}, AO Erkimbaev², GA Kobzev², MS Trachtengerts², and VY Zitserman^{2}*

¹National Research Centre “Kurchatov Institute”, Moscow, Russian Federation

Email: eletskii@mail.ru

^{2*}Joint Institute for High Temperatures, Russian Academy of Sciences, Izhorskaya 13, Bldg. 2, Moscow 125412, Russian Federation

Email: vz1941@yahoo.com

ABSTRACT

This article is devoted to general problems of development of reference data on properties of nanosized objects. It has been shown that the peculiar features of physical characteristics of nanostructures influence the behavior of an expert engaged in building the relevant computer database of property data. The building procedure includes comprehensive data systematization on the basis of classification of nanostructures and detailed identification of a nano-inherent object within the selected class. The key features of data on nanosized objects are discussed, including variation of property nomenclature, dimensional effects, and a high level of data uncertainty. The approaches to data systematization proposed in the article are considered in terms of ISO recommendations. Along with systematization, we propose a procedure for data certification taking into account a quantitative statement of uncertainty as well as quality indicators. The latter indications address the completeness of the description of both an object and a measurement method as well as the reproducibility of results. As an example, property data of carbon nanoforms (nanotubes, graphene, etc.) are analyzed.

Keywords: Nanostructure, Nanomaterial, Nano-object, Nanostructured material, Numeric properties data, Dimensional effect, Logical structure, Semistructured data, Data certification, Data evaluation

1 INTRODUCTION AND SETTING UP THE PROBLEM

This paper focuses on building numerical databases on the properties of nanoscale objects. The main focus is a system of *nanodata* (defined as data on nanosized materials) in general with essential and specific features, such as the existing body of data, logical structure, format, representation in a database (DB), etc. The variety of synthesized nanoforms and types of objects with unique properties defined by dimensional factors at the nanoscale makes it impossible to use usual database approaches when applied to numeric nanodata.

Here we mainly summarize the general concepts and procedures. This problem was already encountered while building a database for carbon nanoforms - fullerenes, graphenes, nanocapsules, nanotubes, nanodiamonds, etc. (Erkimbaev, Zitserman, & Kobzev, 2010). Other structures of carbon, such as nanofibers, nanoconic tips, and nanocapsules, are objects with the same class of problems (Hu, Shenderova, & Brenner, 2007). All these nanosized materials were discovered during the last decades, but until now, there has been no internationally adopted nomenclature or general specification of the data. The status in this area dramatically differs from that with common materials.

On the other hand, the large body of existing relevant data makes it promising today to choose nanocarbon systems to develop a suitable approach for evaluating and managing existing property data for the multitude of nanoscale objects. In addition we already have done a significant amount of analytical work on the subject, including reviews by one of the authors (Eletskaa, 1997; Eletskaa, 2002; Eletskaa, 2004; Eletskaa, 2007; Eletskaa, 2009).

2 KEY FEATURES OF NANOSTRUCTURES AND NANOMATERIALS

2.1 Main peculiarities of numeric data

According to the ISO recommendation (ISO/IEC, 2010), the generic term for all discrete nanoscale objects is **nanomaterial**, which is subdivided into two subclasses: nano-objects and nanostructured materials. The former includes objects with any external dimension in the nanoscale (approx. 1 nm to 100 nm); the latter includes materials having internal structure or surface structure in the nanoscale. Even a cursory examination of properties of nano-objects shows three main features which must be taken into consideration in any effort to compile and disseminate fully evaluated nanomaterials property data:

- *A large variety of existing object types cannot be confined to fixed property nomenclature.* Different kinds of nano-objects have their own lists of important features that should be incorporated in a database. They demand the development of a flexible logic structure capable of supporting such data.
- *Nano-objects lie in an intermediate position between single molecules and bulk substance.* For this reason it is necessary to ascribe the nomenclature of macro properties to nanoscale objects. Examples are found in mechanical properties and thermal conductivity of carbon nanotubes (CNT) and graphene, phase transitions in clusters (Berry & Smirnov 2009), and variations in the phase diagram for diamond-graphite as one passes from bulk to nanoscale objects (Yang & Li, 2008).
- *Properties of nano-objects show significant dependence on production methods (processing history, fabrication treatments, etc.).* Sometimes it is impossible to reproduce results of measurements in the same or other laboratories. The uncertainties may be due to methods and conditions of synthesis as well as uncontrollable factors, such as defects in structures, impurity on surface, etc.

One can also select other peculiarities that distinguish nano-materials from common substances. This problem has been considered by Rumble, Freiman, and Teague (2012) in a report on a recent ICSU - CODATA Workshop. The authors noted additional characteristic properties unique to nanomaterials, in particular: surface to volume ratio (surface areas up to 1000 m²/gm); different bulk and surface electronic structures; quantum size effects; large influence of small amount of impurities, etc. These and other similar characteristics are used as identifiers in a detailed description of a specific nanomaterial (See Section 3). The activity of ISO as well as the European Committee for Standardization (**CEN**) in the field of nanotechnologies, including preparation of standards for classification, terminology, measurement, characterization, etc., are discussed in detail in the report by Jean-Marc Aublant (2012).

2.2 Dimensional effects

The fundamental and universal reason for deviations of numerical nanodata from related bulk properties lie in the dependence of properties (structural, thermodynamic, electronic, transport, etc.) on the characteristic size of a nano-object. What is more, a distinction needs to be drawn between irregular dimensional dependence (with specific maxima in some cases) and regular (monotonous) dependence inherent in the bulk objects. The first is the irregular dependence of a property on the number of particles that shows extremes at so-called "magic numbers", corresponding to the maximum of the cluster stability. Irregularities due to size of nano-objects are observed also in mass spectra, ionization potentials, and other properties. The effect of size (dimensional effect) is the main reason for the data uncertainties. For example, thermal and electric conductivities of CNT depend significantly on tube length. This is caused by a change of the transport mechanism (from ballistic to diffusive) at some specific CNT length (Eletskii, 2009). The characteristic length at which the change occurs depends on concentration and type of defects that connect directly with processing history and conditions of specimen preparation.

As a result, specifying only CNT length is not sufficient to provide complete characterization. At least, length should be accompanied by data on synthesis method and processing history. A similar case is seen also for graphenes. Its transport properties are essentially dependent on the lengthwise and crosswise extents of a specimen and also on edge structure (chirality). If reliable information on size, structure, chirality, and defects of an object is deficient or lacking in any manner, the data on properties have rather essential uncertainty. Thus, results of measurements by Brown, Hao, Gallop, and Macfarlane (2005) show that thermal and electric conductivities drop off by a factor of 2-3 if magnified for individual single-walled CNT with increasing length of less than 1%.

Properties of multilayered CNTs and graphenes depend appreciably on the number of layers, which also can be considered as manifestations of a dimensional effect (Eletskii, Iskandarova, Knizhnik, & Krasikov, 2011). Thus, according to recent measurements by Ghosh, Bao, Nika, Subrina, Pokatilov, Lau, and Balandin (2010), the thermal conductivity of multilayered graphene decreases in inverse proportion to the number of layers n and reaches crystal graphite value at $n > 4$. In the case of CNTs, the effect is opposite — thermal and electrical conductivities of a specimen increase with increasing numbers of layers (Li, Lu, Li, Bai, & Gu, 2005)

In addition to dimensional effects, it is necessary to pay attention to the uncertainty of cross-section sizes. For example, measurements of thermal and electrical conductivities, elasticity modules, etc. may be fulfilled if cross-section data of an object are available. If the thickness (width) of an object is one or several layers of atoms, the choice of this parameter (thickness or width) becomes arbitrary and brings forth additional problems.

This problem may be demonstrated by measurements of the thermal conductivity of graphene, defined as the relation between heat flux through a sample and its temperature gradient (Nika, Pokatilov, Askerov, & Balandin, 2009). Obviously, the exact value of the graphene layer thickness is required for calculation of the temperature gradient. It is commonly accepted that the distance between the nearest layers in the crystal graphite, which equal 0.34 nanometers, can be used. Sometimes, however, the characteristic size of a carbon atom, smaller by a factor of 2 – 3, is also used. Thus, the arbitrary choice of the single-layer graphene thickness can result in more than 100% uncertainty in the estimation of thermal conductivity.

A similar problem to the previous one appears in the measurement of the Young's modulus of CNTs (Eletskii, 2007). This property is defined as the relation between the stretching force and increasing sample length. In turn, the specific stretching coefficient is calculated from a cross-section of the sample. There seems to be no escaping the conclusion that the uncertainty of the Young's modulus for a CNT can also reach 100% based on uncertainty in cross-section as discussed above.

The significant dependence of nanomaterial properties upon the size of structural units means that a new parameter, the *size of a unit* (crystalline particle, colloidal particle, etc.), should be taken into consideration. In many cases, subtle details, for example, the size distribution, the volume ratio $\Delta V/V$ of space between grains, and so forth (Suzdalev & Suzdalev, 2001), may appreciably affect the physical properties. Such supplementary data are also necessary for valid specification of a nanomaterial, along with a description of the material's origin and its processing history. For example, full details are ultimately necessary for carbon cloth-like materials made of single-wall CNTs, multilayered graphene paper, CNT yarn, etc.

It is necessary to bear in mind that both the geometrical and physical parameters of nanoscale units can show variations in values. Distribution of these parameters depends on the methods and conditions of production and noticeably affects the numeric property data (transport, mechanical, etc.). An example by Han and Ostrikov (2010) demonstrates the importance of detailed description. Nitric acid processing of single-walled CNT films changes the type of electrical conductivity from semi-conductor to metallic (Lobach, Buravov, Spizyna, Eletskii, Dementyev, & Maslakov, 2011). Such processing removes attached molecules or adsorbed radicals from the surface of the CNT and thus dramatically changes the electronic structure of the object.

The above example demonstrates once more that there are other factors that have influence on data uncertainty — in particular molecules or radicals adsorbed on the surface. The physical properties of such objects are determined by a relatively large contribution of the surface as compared to the bulk. Adsorption of radicals by the CNT or graphene surface is responsible for the variation in the electronic structure that has an immediate impact on the electrical properties. Thus, the electrical conductivity of a pure graphene sheet is 100-1000 times larger than that of graphene partially oxidized with 10 % of oxygen (Hernandez, Nicolosi, Lotya, Blighe, Sun, De, et al., 2008). This change is caused by the energy gap that occurs in graphene oxidation. The thermal conductivity of graphene also decreases as the number of adsorbed radicals increases. The adsorbed radicals act as scattering centers for phonons, hindering collisionless movement along the specimen. There are some processes that remove radicals by heat or chemical treatment.

To sum up, reliable data on type and amount of adsorbed radicals are necessary, in addition to geometry and object structure characteristics, for unambiguous characterization of a nano-object as well as for data evaluation. Hence, the measured properties of nano-objects have unremovable uncertainty that stems from their atomic structure. Nevertheless, the needs of engineering design or scientific research demand that property data have a certain certification of quality or an integrated estimation of uncertainty. This estimation should be based on accessible data of the size and structure of object, method of measurement, method of synthesis, etc. More details are considered in Section 6.

2.3 Data complexity

In addition to high data uncertainty, the description of a nano-object involves yet another peculiarity. The point is that properties make sense and are of value for users only when expanded descriptions of measurement methods, state of the specimen, and environmental and other conditions are available. Identification should include the whole set of quantitative and qualitative features concerning structure, size, morphology, synthesis method, etc. It is pertinent to note that the same multifactor description is inherent not only for objects of the nanoworld but also for other materials as well. Their properties are always defined by a wide complexity of factors (technology and structural features, environment, etc.). This feature distinctly differentiates a material from common (pure) substances (or solutions) with properties defined solely by chemical composition and/or structural formula. In pursuing the goal of assessing materials data quality, the special concept of materials metrology (Munro, 2003) was designed to develop materials databases. The principles and practices set forth in this work demand that measurement results be presented together with data on the measurement method and object characteristics and that the reliability of numerical values be defined by the scope of the available data. As a result of experience acquired during the development a database for superconductors and ceramics, Munro (2003) has suggested procedures for comprehensive assessment of the quality of material data of any kind. The importance of these procedures increases when considering objects of the nanoworld, as the number of additional factors involved by synthesis and/or measurement methods also increases. Some of these new factors have more effect on numeric property data and formulations than might be expected.

The necessity of a comprehensive description of a nanomaterial is stressed also by Rumble, Freiman, and Teague (2012). In addition to the above-mentioned factors, they propose to consider the chemical reactivity (i. e., ability to interact with different objects), capacity to form associations (bonding, attachment, aggregation), and physical properties. For properties, they suggest two alternatives: (1) the description of a material includes its properties; or (2) a material is described without reference to its properties. In this article, we have accepted an approach to the systematization of data on properties of nanostructures based on a strong separation of two aspects: detailed description of a nanomaterial (identification) and description of properties (see next section and Figure 1). Separating the property data from the general description of a nanomaterial is justified because numerical property data include a large set of additional information: specification of the experiment, methods of processing or estimation, uncertainty analysis, etc.

3 THE GENERAL APPROACH TO COLLECTING AND PROCESSING OF NANO-OBJECT DATA

With the use of data analysis principles (Moniz, 1993; Munro, 2003; Newton, 1993) that have already been applied in material science and with the specific experience of data nanostructures evaluation (Eletskii, 1997; Eletskii, 2002; Eletskii, 2004; Eletskii, 2007; Eletskii, 2009), the approximate schematic in Figure 1 may be proposed. In this scheme, the design of a data collection goes in two directions: the characterization of objects and the specification of properties.

The first path (left side of the diagram) is based on a classification schema that allows the data to be organized or classified into categories in accordance with an object's topology, size, etc. An expert in database design needs to use an accepted scheme and must classify the object into one of its classes. Then the problem of its detailed identification arises. For example, a specific CNT can be identified by diameter, number of layers, chirality indexes, and, as discussed in Section 2, synthesis conditions. Generally, the set of identifying features is specific to each category and covers a rather large set of characteristics, such as the monomer formula, the number of monomers in clusters or nanostructures, morphological features, thermal prehistory, external factors, etc. Identification procedures are discussed in detail in Section 5.

The second path (right side of diagram) shown in Figure 1 describes the work steps applicable for a detailed elaboration of the property data. The initial step involves a choice of state parameters and property nomenclature. Besides temperature and pressure, data on structure and dispersion (size distribution) as parameters may be considered. It is interesting that the dispersion of data as an additional state parameter was used long before the emergence of present nanotechnologies. Therefore, the dilemma in characterization appears: whether to qualify

the dispersion of an object (for instance, cluster) as a state parameter or to take it as an identification parameter. The appropriate decisions must be made in the data analysis of every object.

The nomenclature of properties depends essentially on the object class (headings of classification) as well as on the specific purpose or functions to be served by the data. It is essential that the nomenclature for clusters, nanotubes, and similar structures is extended to include both molecular and macroscopic properties. This feature shows the intermediate nature of nanostructures, between a molecule and a bulk substance. For example, the whole set of mechanical characteristics peculiar to engineering materials may also be attributed to nanotubes (Eletskii, 2007). Another example is adsorption properties that are mostly used to characterize porous materials (Eletskii, 2004).

When the nomenclature is established, the main work can begin – extraction of relevant numerical values and formulas from publications and other sources. Specific features of an object and data presentation in the original source dictate the type and format of the data to be accepted for data input. It is always preferable to store data on nano-objects as raw data, which are present mainly in three basic forms (tabular, formulas, graphic), because they have an exceptional variety of property value representations.

Regardless of the adopted representation for numeric data, the data are accompanied by some metadata in any information system (Erkimbaev, Zitserman, Kobzev, & Fokin, 2008; National Information Standards Organization, n.d.). This term (metadata) denotes data about the data, i.e., structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. When applied to a database of properties, metadata include names, designations, units of measure, measurement method (or that of evaluation), presentation of property data, and more importantly, possible uncertainty. Data presentation, as a concept, is wide and has different meanings that cover the main details of property definition and type of data used as well as a set of measurements for the multidimensional data with lists of their types and values.

Some elaboration may be critical to understanding or using property data because property definition is connected frequently with context: a measurement method, model, scope of use, etc. Metadata permit linkage of the property definition with the context; for instance, data must be accompanied by measurement methods, model, application, and so on. The typical example is "hardness of material", a property defined by method of measurement (*Knoop hardness*, *Vickers hardness*, *Rockwell hardness*, etc.). Differences in definitions of thermodynamic properties relate to the accepted reference state, methods of numerical data representation (direct, difference from values at a reference state, ratio of the value to that at reference state, etc.), and temperature scale. The metadata are also important for uncertainties, as their evaluation and representation have diverse kinds: absolute and relative values, the same value for an entire data set or presented for every experimental point, confidence interval at a given level of confidence, etc. The last step finishes the data evaluation (Section 6), i.e., it integrates numeric property data and metadata presented in separate files.

4 CATEGORIZATION OF NANO-OBJECTS

As a rule, the first step involves identification of an object (Figure 1), which sets it apart from the classification heading according to an adopted set of identifying details. We have here an analogy to identification in chemistry when a substance from a group (elements, oxides, hydrides, etc.) is defined by its chemical formula. At the same time, the classification of nanosize objects is much more complicated because the chemical composition or even existence of the structural form (as in organic chemistry) is not sufficient for object identification. For example, fullerenes and CNTs may be considered to be both clusters and large molecules. Fullerenes as molecules do not raise any doubt, but because of appreciable diversity in the size and structures of CNTs, ambiguity in their classification seems to be a cause of unavoidable difficulty. Problems arise also when we consider the family of graphenes. The classical definition of graphene corresponds to a single-layered hexagonal graphite structure. However, many authors consider structures consisting of two or even several closed layers to be graphene as well. Therefore, it is necessary to define and adopt the number of graphene layers when graphene converts to graphite. This question is typical for nanostructures with dimensional effects.

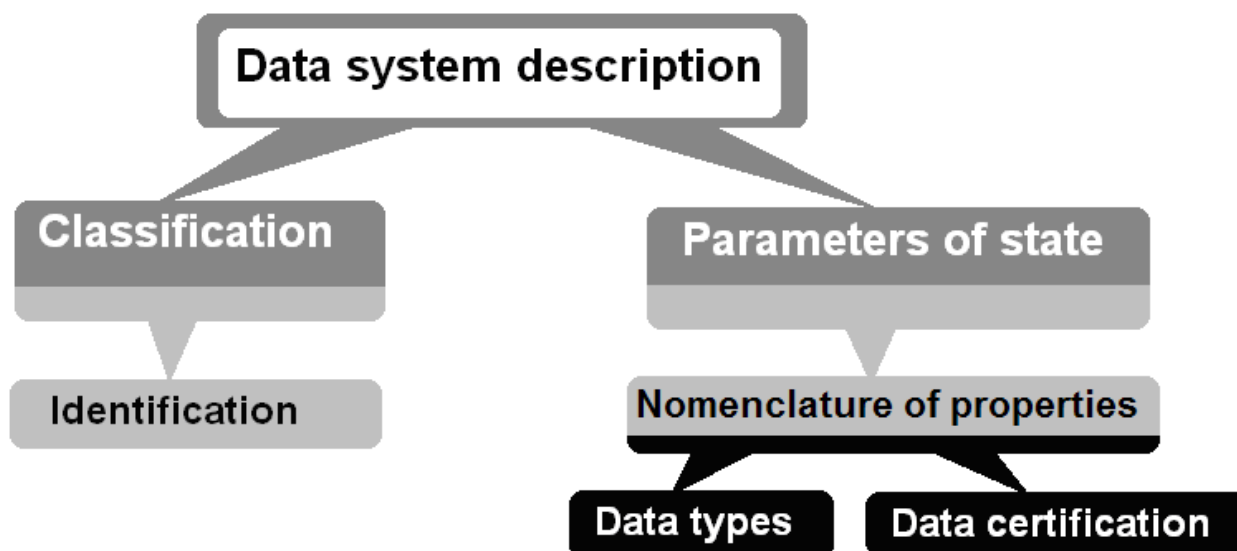


Figure 1. Schematic description of data system designing

A simple scheme was suggested by Suzdalev and Suzdalev (2001), who divided the whole nanoworld into two types — separated individual nanoclusters and nanocluster systems (materials). Moreover, they introduced six cluster types, based only on methods of synthesis: molecular ligands, gas ligandless, colloidal, solid-state, matrix, and film. Thus, all kinds of fullerenes and CNTs come under the heading of ligandless gas phase clusters.

Pokropivny and Skorokhod (2008) distinguished four types of objects using dimensions as criteria. This number can have four values from 0 to 3. Value $K=0$ means a cluster with length no more than 100 nm in every dimension. In contrast, the value $K=3$ is applied to common macroscopic substances or materials. The prefix *nano* in this case shows only the size of the elements that make up the material. The intermediate values $K=1, 2$ are applied to 1-dimensional and 2-dimensional structures, which have macroscopic size along one or two dimensions, respectively, for example, nanowires and nanofilms.

In the case where these criteria are used to certify structural elements that form the defined object, the dimension, however, can accept only 3 values ($L=0, 1, 2$). Then the class of objects that can be assembled by elements of same type may be described by the *nanofformula* $\mathbf{K D L}$. All the clusters of the C_N type that are chemical forms of carbon atoms are referred to the sole $\mathbf{0D0}$ class because dimension values ($K=0, L=0$) refer equally to cluster and monomer. On the other hand, nanotubes or graphene assembled by those elements are defined by formulas $\mathbf{1D0}$ and $\mathbf{2D0}$. If an object is assembled by elements of several types, the formula assumes the form $\mathbf{KD}\{L,M,N \dots\}$ and $\mathbf{K} \geq \max\{L,M,N \dots\}$. The number of classes defined with this classification is essentially limited. For example, when only three structural elements are used, the number of classes is no more than 36 (Pokropivny & Skorokhod, 2008).

The most exhaustive approach for a classification scheme has been proposed by ISO Technical Committee 229 on Nanotechnology. The authors of report ISO/TR 229 (2010) propose a hierarchy system called a **nano-tree**, “upon whose basis wide ranges of nanomaterials can be categorized, including nano-objects, nanostructures and nanocomposites of various dimensionality of different physical, chemical, magnetic and biological properties”. The higher levels define an object by four criteria: dimensionality; external and inner structure; chemical origin; and properties and behavior. Notably, as distinct from the schema of Pokropivny and Skorokhod (2008), the number of nanoscale sizes is accepted as dimensionality. According to this, nanotubes or nanowires are considered as 2D objects, while nanofilms (e. g., graphene) are considered as 1D ones. In this definition, clusters fall into the 3D category. The second level (second branch of the nano-tree) distributes objects over three blocks (branches): one-component, multi-component, and nanostructured, each of which has a set of types. For example, an object hierarchy is established: 1D→one-component→nanofilms. The next two hierarchy levels distribute objects over their chemical origin (metal, ceramics, polymers, etc.) and properties (physical, chemical, mechanical, etc.). The schema as a whole looks rather cumbersome while it does not cover the overall class of nanomaterials and inevitably requires revision as new materials appear. For this reason, the schema of Pokropivny and Skorokhod (2008) has been used as the basis for classification in the present work. This schema

is quite easily adaptable and is dependent only on the dimensionality of the nanostructures themselves and their components. All other identifiers determining an object's structure, chemical origin, composition, etc., are considered separately. Their totality distinguishes the object under the heading of the classification scheme.

5 OBJECT IDENTIFICATION

The identification procedure sets an object unequivocally apart from a class of similar objects (CNTs, graphenes, clusters, etc.) that fall under the same class heading. For instance, it is necessary to use identifiers of an object, such as chemical composition, size, and structure to set it apart from other members of the class defined by a topological nanoformula. Thus, the topological formula **0D0** for cluster A_N should be followed by the chemical formula of a monomer A , monomer number N , and by the symbol of the point group (D_{3h} , T_d , O_h , etc.). Precisely such tables of atomic and molecular clusters data were used for building the Cambridge Cluster Database (Cambridge Cluster Database, n. d.). Identification by number N appears to be impractical when it comes closer to 10^3 – 10^4 . In this case the more convenient characteristic is the linear size expressed in nanometers accompanied by the crystal type and features of morphology.

The representation of modeling results for several carbon cluster families (Yu, Chaudhuri, Leahy, Wu, & Jayanthi, 2009) may be an illustration for the above. The proper identification is achieved there by specifying only cluster diameter and structure type: bucky-diamonds, icosahedral clusters, fullerenes, and fullerene-like structures (carbon cages and carbon onions). The analysis of nanodiamond detonation synthesis (Benedek, Milani, & Ralchenko, 2001; Danilenko, 2006; Shenderova & Gruen 2006) needs also to consider nanocluster types. Particles of nanographite, nanodiamond, and nanodiamond covered with a graphite layer, appear in the reaction zone. Each of these is qualified by diameter while the nanodiamond is also characterized by layer thickness.

The other broad class of nanostructures is nanotubes. These objects are included in a class defined by the topological formula **1D0**. Specific objects may be set apart from others in that class by monomer chemical formula (e.g. C, BN, and BeO), chirality indexes, diameter, and number of walls. In addition, exact identification needs additional data on structural defects, surface state, and other factors induced by material synthesis. **These examples show that the identification rules cannot be set *a priori*, i.e., specific peculiarities of every class must be taken into account.**

Identification of nanomaterials as macroscopic objects is more difficult than the cases above because it is necessary to elucidate a **source** material (matrix) description without taking into account nanosized inclusions and treat these inclusions separately as well. There is a general recommendation for identifying materials by the ASTM Committee E49 on Materials Databases (now defunct) with seven distinct categories (See Table 1).

Table 1. The ASTM scheme for material identification (Moniz, 1993)

Number	Indicators
1	Primary identifiers (e.g., material class)
2	Characterization of the material (e.g., in terms of its chemical composition)
3	Widely recognized specification codes (e.g., M-52 steel)
4	Source of the material and its processing history
5	Geometrical details for the specimens used in the measurements
6	Any fabrication treatments (e.g., heat treatments)
7	Service history experienced by the specimen used in the measurements

It is obvious that identifiers for separate nanostructures can be applied also for nanosized inclusions. Some new parameters, such as volume or weight fraction of inclusions, dispersion, volume fraction of intergrain area, and so forth (Suzdalev & Suzdalev, 2001), are necessary as the case might require. As a whole, nanostructure identification should meet two requirements: (1) use of an extensive set of identifiers, defining size, chemical composition, structure, and other factors; and (2) possibility to change the defining set while going from one nanostructure class to another one.

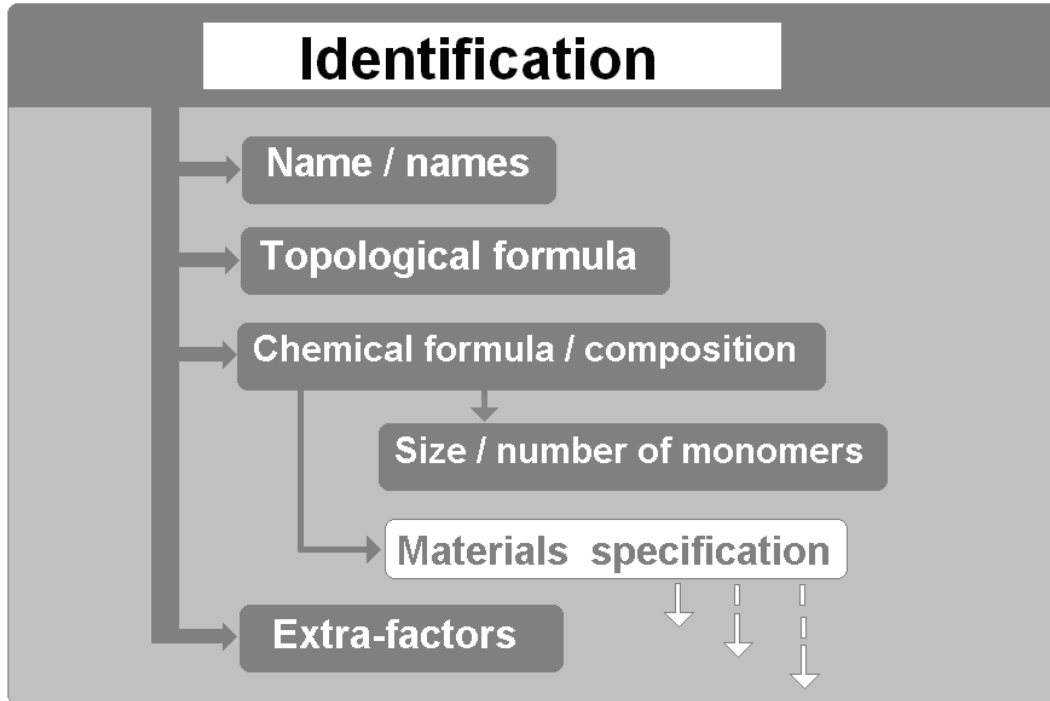


Figure 2. Schematic of the nanoscale objects characterization

Figures 2 and 3 present a summary "frame", i.e., the exemplary listing of data blocks that are needed for identification of objects. The scheme named *Identification* displays the need to consider supplementary identifiers for macroscopic objects (e.g., crystal symmetry data). The last of the blocks (*extra-factors*, expanded in Figure 3) refines data on structure, influence factors, synthesis, etc. for both an object and a particular specimen. Logical structure and type of data in every block vary with nanostructure type. This is the foundation to implement the context-dependent **semistructured** data concept as introduced by Abiteboul, Buneman, and Suciu (2000) and Erkimbaev et al. (2008).

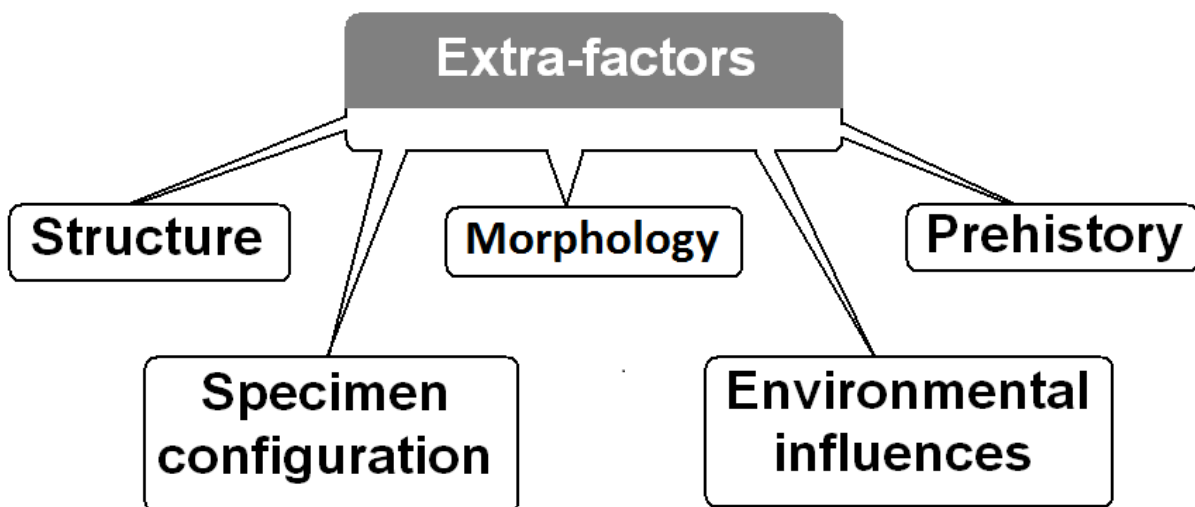


Figure 3. Data block "extra-factors"

6 DATA CERTIFICATION

6.1 Data certification procedures

Data certification (DC) is a set of procedures that fulfills multi-aspect evaluation of data presented and results in estimation of total uncertainty, i.e., estimated error value and/or some data quality indicator. In some exceptional cases, it may be sufficient to make a decision that the data are acceptable for an application in accordance with some criteria.

According to the simplified scheme in Figure 4, the first DC step includes three procedures for evaluation of reliability, completeness, and consistency of the data. The first procedure should show whether identification (specification) of the object is completely presented in the data set. This part of certification is important as materials properties are of no value without detailed characteristics of the material. The identification may be called complete when the values of all identifying details are known and specified and the blocks, schematically outlined in Figure 2, are filled out.

The second DC procedure (Figure 4) should provide the answer to the same question, relating now to the measurement (or prediction) method, i.e., whether the description of the method is sufficient for evaluation of results. Development of nanotechnologies is the result of widely used, high-precision physical methods that allow determination of structural characteristics and chemical composition of a sample: electronic (ionic) microscopy; Raman-spectroscopy; electronic spectroscopy (Auger-spectroscopy, X-ray photoelectron spectroscopy, electron energy loss spectroscopy, and many others). The foremost goal in describing the measurement method is to present sufficient information for comparability and estimation of uncertainties. Sufficient details on the applied method allow the estimation of reliability, taking into account that all techniques have limited ranges of practical use.

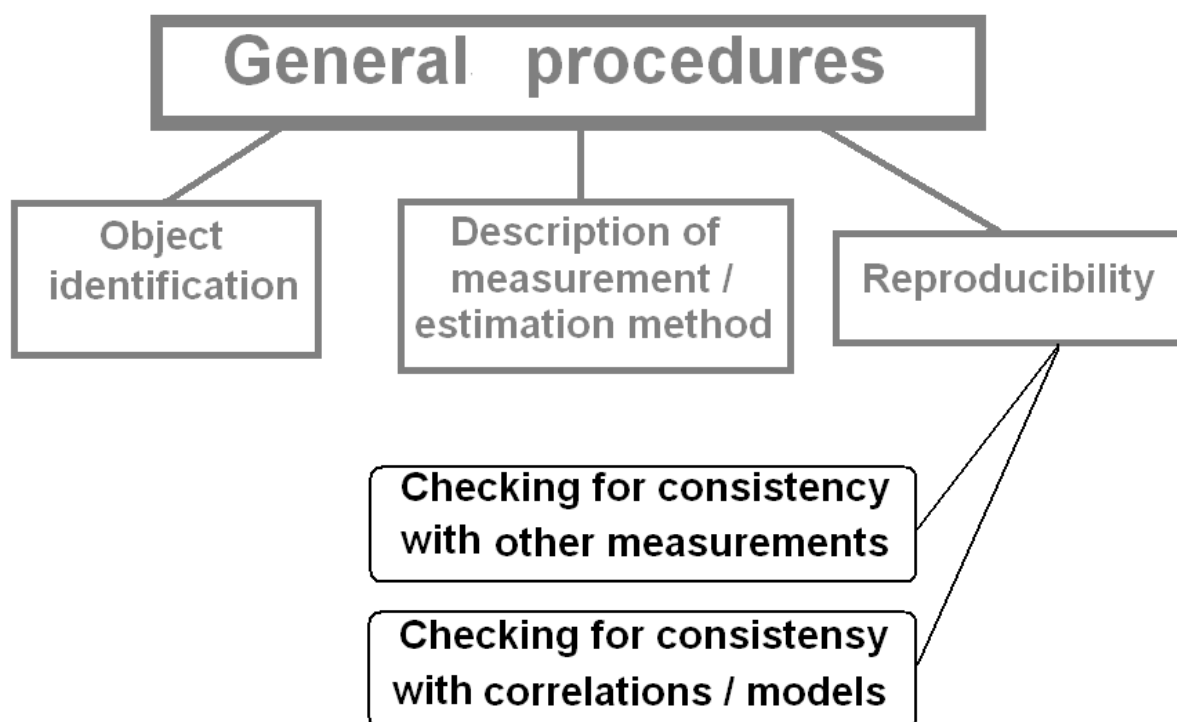


Figure 4. Schematic of the data certification

Correlations and theoretical methods (presumably using quantum chemistry techniques) must be characterized in the same manner with respect to assessment of reliability and completeness. As a rule, the universally accepted name of a method and its version, the actual program used, model parameters, and so on, must be presented. For example, the investigation of carbon clusters energetics (Yu et al., 2009) was accompanied by

detailed descriptions of such points: the modeling is based on a self-consistent and environment-dependent (SCED) Hamiltonian, implemented in the framework of a linear combination of atomic orbitals (LCAOs). As the method is a semiempirical one, the description includes the set of parameters used in the work, the so called optimized carbon parameters. The essential point in the description and estimation of reliability of a theoretical method is the comparison of results with available experimental data and/or alternative calculations.

The third procedure, shown in Figure 4, enables a preliminary conclusion on reproducibility of measurements. Reproducibility is the mutual agreement among independent measurements, conducted and reported by the same or different laboratory. A higher level of confidence can be reached when validation of the value estimated by several predictive methods (quantum chemistry or semiempirical) is possible.

When coupled with object and measurement method specifications, reproducibility allows us to provide ultimate assessment of the data quality. Qualitative (intuitive) characteristics may be discriminated by means of level quality indicators (high, middle, low) that are convenient in database building. The corresponding metadata that define the quality of data, include three indicators in this case: formally defined reproducibility of data as well as completeness of object and method specifications.

6.2 Components of nano-object data uncertainty

In addition to quality assessment, numerical estimation of the data uncertainty must be introduced. Different versions of estimation procedures for uncertainty are necessary because of the variety of nanostructure types, forms, measurement methods, even when all objects are in the same nanocarbon family. **The first variant** is applicable when the data type depends only slightly on features peculiar to nanosize objects. Thus, all publications on thermodynamic properties of fullerenes and fullerites represent results of the same form that are common in the thermochemistry of traditional substances. Calorimetric methods together with standard estimations of molecular constants allow usage of common estimations of uncertainty based on usual statistical methods (Diky & Kabo, 2000). The metadata must account for representation of uncertainty: standard deviation, level of confidence for the interval, and a combined uncertainty that includes extension of uncertainties from variables to the property. It is a common practice, when applied to nanostructures, to represent uncertainty as a root-mean-square deviation expressed in absolute or relative values.

The second variant may be used when "ineradicable" uncertainty stems from the dimensional effect or from a specific synthesis method. The expert evaluator must take into consideration possibilities of the method in combination with reproducibility of obtained results. As a result, the expert can offer an estimation of uncertainty in the form of a value interval but without probability interpretation, i.e., without a distribution law within the interval.

The third variant can be applied when theoretical methods of calculation, especially methods of quantum chemistry, are used. Notice that nanostructures data have a specific feature — published data, calculated by theoretical methods. These data are becoming more numerous, along with improvement of data quality (accuracy and reliability). As appropriate estimations of uncertainty, some indicators may be considered: for example, (1) systematic error inherent in each method and usually presented in publication and (2) qualitative assessment of the reproducibility (measure of agreement) received from comparison with similar calculations or available experiments. The expert estimation presents the results in the form of a possible value interval as well as those pointed out in previous variants.

6.3 Data quality categories

Quality indicators, combined with the numerical estimation of uncertainty, are a foundation for data to be distinguished by the categories, as already has been defined for common materials (Munro, 2003). Table 2 shows how numeric data may be assigned to one of the eight proposed categories defining their reliability. Experimental as well as theoretical (predicted, calculated, etc.) data may be assigned to three quality categories (1-3 for experimental, 4-6 for theoretical data). For instance, experimental data may be assigned to category 1 if the statistical error is known (that is the common situation in the research of macroscopic objects) and each indicator that defines reproducibility and completeness of data has a high quality level. In more difficult cases (e.g., the occurrence of dimensional effects), the same level of reliability is assigned when uncertainty is shown as an interval of values and the same level of quality indicators is present.

A decrease in quality and/or absence of a numerical estimation of uncertainty transfers the experimental data to category 2 or 3. Theoretical data may be qualified in the same way — by the availability of systematic error that is inherent in the model and quality indicators (categories 4-6).

Table 2. The categories proposed for property data of nano-objects (based on Munro, 2003)

№	Data category	Assigning data to categories
1	Experimental, high level of reliability	Availability of numerical estimations of uncertainties at a high level of quality indicators
2	Experimental, middle level of reliability	Availability of numerical estimations of uncertainties at a middle level for at least one of the quality indicators
3	Experimental, low level of reliability	Availability of numerical estimations of uncertainties at a low level for at least one of the quality indicators Absence of numerical estimations of uncertainty
4	Theoretical (estimated), high level of reliability	Availability of systematic error typical of a given method at a high level of quality indicators
5	Theoretical (estimated), middle level of reliability	Availability of systematic error typical of a given method at a middle level for at least one of the quality indicators
6	Theoretical (estimated), the low level of reliability	Availability of systematic error typical of a given method at a low level for at least one of the quality indicators Unavailability of systematic error data
7	Commercial	Selected property data, included in the documentation given by the manufacturer or the supplier of a product
8	Typical	Typical values accepted from the literature, without indication on reliability level Provide order of magnitude estimations or functional trends Data should be qualified as unevaluated.

In actual practice, additional categories may be needed: commercial and typical. Data of the first kind are provided by manufacturers to inform a customer about the principal characteristics of products (for example, fullerenes or CNTs). Typical data are derived from preliminary reports with no indication of reliability or uncertainty. Nevertheless, these data may be used for order of magnitude estimates or functional dependencies. In the absence of alternatives, commercial or typical data may be loaded into a database with the indication that any assessment of reliability is not available.

7 CONCLUSION

The present work suggests general principles for data collection on properties of nano-objects. These principles may be used as guidelines for database building. The paper includes a description of a logical data structure adapted to specific features of nanostructures and methods of data certification and categorization.

When creating a logical data structure, an expert defines the necessary set of identifiers and nomenclature of the properties as well as specifies types and data formats for each attribute. The anticipated requirements for data structure, stemming from specific properties of nanostructures, can be met by the widely used semistructured data (SSD) model (Abiteboul et al., 2000; Erkimbaev et al., 2008). The model has been introduced to accommodate all forms and kinds of data that come from multiple sources with differences in notation, meaning, etc. In the SSD model the information that is typically associated with a schema is contained within the data themselves, which is sometimes called "self-describing". In such a database there is no clear separation between the data and the schema, and the degree to which it is structured depends on the application. The benefit is that these data do not have a rigidly and predefined schema.

Some different approaches to consider when building databases of that type were proposed, based, for example, on XML tools (Graves, 2002). We have applied technology that used a free distributed object-relation database system with open code PostgreSQL (Stones & Matthew, 2005), which combines the traditional relational data

model with maintenance of a “fuzzy” data structure. The relative stability of the data “frame” described above (Sections 4 and 5) serves as an additional argument in support of the technology inherent in conventional tools. The potential of PostgreSQL has appeared to be quite sufficient for databases on nanocarbon properties (Erkimbaev et al., 2010), despite the exclusive variety of structures and materials of this class.

Along with the variety of data structures, there is a marked feature of numeric data for nano-objects — their level of uncertainty is quite high. Some of the factors responsible for the basically unavoidable high uncertainty are due to the nanoscale nature of the object. Uncertainty also arises when taking into consideration both types of data — experimental and theoretical and, in particular, results of quantum chemical calculations in absence of a reliable approach to the assessment of confidence.

So, it is possible to adequately describe confidence in data only by involving the complete body of available information on uncertainty and numeric assessments as well as quality indicators. In Section 6, we developed a procedure that may be used for certification of data by introducing quality indicators and assigning data to categories of reliability (Table 2). In a broad sense, reliability is defined by the available numeric value of uncertainty, the completeness of the data in both object and research methods, and the reproducibility of the measurements/estimations. The metadata include **quality indicators** (object, method, reproducibility) with estimations according to a three-level scale, the category of reliability, and the indexes defining the uncertainty (random or systematical, absolute or relative, etc.).

The conceptual scheme described in this paper may be adjusted to arbitrary nanoscale objects by changing (or expanding) the classification scheme, identification standards, properties nomenclature, and if necessary, the certification procedure. From a practical standpoint it is very important that the approach proposed here be fully suited to making such adjustments even if an expert may lack sufficient knowledge of computer technology.

8 ACKNOWLEDGEMENTS

The authors thank the anonymous referee who steered our attention to international activity within the area of exploration under consideration in the present article, including data categorization on properties, standards in the field of identification, and classification of nanostructures, etc.

9 REFERENCES

- Abiteboul, S., Buneman, P., & Suciu, S. (2000) *Data on the Web: From Relations to Semistructured Data and XML*, Burlington, Massachusetts: Morgan Kaufman Publishers.
- Aublant, J.-M. (2012) *International and European Standardization of Nanomaterials*. ISCU-CODATA workshop, Paris, 23-24/02/2012. Retrieved from the World Wide Web, October 20, 2012: <http://www.codata.org/Nanomaterials/Index-agenda-Nanomaterial.html>
- Benedek, G., Milani, P., & Ralchenko, V.G., (Eds.) (2001) *Nanostructured carbon for advanced applications, NATO Science Series. II. Mathematics, Physics and Chemistry – Vol. 24*. Berlin, Heidelberg, Dordrecht, New York: Springer.
- Berry, R.S. & Smirnov, B.M. (2009) Phase transitions in various kinds of clusters. *Physics Uspehi* 52(9), 137–164.
- Brown, E., Hao, L., Gallop, J.C., & Macfarlane, J.C. (2005) Ballistic thermal and electrical conductance measurements on individual multiwall carbon nanotubes. *Appl. Phys. Lett.* 87, Ar# 023107
- Cambridge Cluster Database of Global Minima (n.d.). Wales group homepage. Retrieved from the World Wide Web, October 20, 2012: <http://www-wales.ch.cam.ac.uk/CCD.html>
- Danilenko, V.V. (2006) Features of carbon condensation in a detonation wave and the conditions for optimal synthesis of nanodiamonds. *Journal of Superhard Materials*, 28 (5), 2006, 7-22.
- Diky, V.V. & Kabo G.J. (2000) Thermodynamic properties of C₆₀ and C₇₀ fullerenes. *Russian Chemical Reviews*, 69(2), 95-104.

- Eletskii, A.V. (1997) Carbon nanotubes. *Physics Uspehi* 40(9), 899–924
- Eletskii, A.V. (2002) Carbon nanotubes and their emission properties. *Physics Uspehi* 45(4), 369–402.
- Eletskii, A.V. (2004) Sorption properties of carbon nanostructures. *Physics Uspehi* 47(11), 1119–1154.
- Eletskii, A.V. (2007) Mechanical properties of carbon nanostructures and related materials. *Physics Uspehi* 50(3), 225–261.
- Eletskii, A.V. (2009) Transport properties of carbon nanotubes. *Physics Uspehi* 52(3), 209–224.
- Eletskii, A.V., Iskandarova, I.M., Knizhnik, A.A., & Krasikov, D.N. (2011) Graphene: fabrication methods and thermophysical properties. *Physics Uspechi* 181(3), 233- 268.
- Erkimbaev, A.O. Zitserman, V. Yu, Kobzev, G.A., & Fokin, L.R. (2008) The logical structure of physicochemical data: Problems of numerical data standardization and exchange. *Russian Journal of Physical Chemistry A* 82(1), 15-25.
- Erkimbaev, A.O., Zitserman V. Yu, & Kobzev, G.A. (2010) Systematization of Data on the Physical and Chemical Properties and Application of Carbon Nanostructures. *High Temperature*, 48 (6), 830–836.
- Ghosh, S., Bao, W., Nika, D.L., Subrina, S., Pokatilov, E.P., Lau, C.N., & Balandin, A.A. (2010) Dimensional crossover of thermal transport in few-layer graphene. *Nature Materials* 9(7), 555-558
- Graves, M. (2002) *Designing XML Databases*. New Jersey: Prentice Hall.
- Han, Z. J. & Ostrikov, K. (2010) Controlled electronic transport in single-walled carbon nanotube networks : Selecting electron hopping and chemical doping mechanisms. *Appl. Phys. Lett.* 96, Ar#233115.
- Hernandez, Y., Nicolosi, V., Lotya, M., Blighe, F.M., Sun, Z., De, S., McGovern, I.T., Holland, B., Byrne, M., Gun'ko, Y.K., Boland, J.J., Niraj, P., Duesberg, G., Krishnamurthy, S., Goodhue, R., Hutchison, J., Scardaci, V., Ferrari, A.C., & Coleman, J.N. (2008) High-yield production of graphene by liquid-phase exfoliation of graphite. *Nature Nanotechnology* 3(9) 563-568
- Hu, Y., Shenderova, O.A., & Brenner, D.W. (2007) Carbon Nanostructures: Morphologies and Properties. *Journal of Computational and Theoretical Nanoscience* 4 (2), 199-221.
- ISO/IEC TS 80004:1: 2010, Vocabulary – Core Terms. Retrieved from the World Wide Web, October 20, 2012: www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=55967
- ISO/TR 11360:2010, *Nanotechnologies - Methodology for the classification and categorization of nanomaterials*. ISO technical committee ISO/TC 229, Nanotechnologies. Retrieved from the World Wide Web, October 20, 2012: www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=55967
- Li, H.J., Lu, W.G., Li, J.J., Bai, X.D., & Gu, C.Z. (2005) Multichannel Ballistic Transport in Multiwall Carbon Nanotubes. *Phys. Rev. Lett.* 95, Ar#086601
- Lobach, A.S. Buravov, L.I., Spizyna, N.G., Eletskii, A.V., Demytyev, A.P., Maslakov K.I. (2011) Temperature-dependent resistance of Single-Walled Carbon Nanotube Films. *Khimiya Vys. Energii* 45(4), 1–7.
- Moniz, B. (1993) Nomenclature and current standards for identification of engineering materials. In Newton, C.H. (Ed.) *Manual on the Building of Materials Databases*. ASTM Manual Series: MNL 19. Philadelphia: American Society for Testing and Materials.
- Munro, R.G. (2003) *Data Evaluation, Theory and Practice for Materials Properties*. Special Publication 960-11. Washington, DC: Materials Science and Engineering Laboratory, NIST.

National Information Standards Organization (n.d.) Understanding Metadata. Retrieved from NISO Website, October 20, 2012: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

Newton, C.H. (Ed.) (1993) *Manual on the Building of Materials Databases. ASTM Manual Series: MNL 19*. Philadelphia: American Society for Testing and Materials.

Nika, D.L., Pokatilov, E.P., Askerov, A.S., & Balandin, A.A. (2009) Phonon thermal conduction in graphene: Role of Umklapp and edge roughness scattering. *Phys. Rev. B* 79, Ar#155413

Pokropivny, V.V. & Skorokhod, V.V. (2008) New dimensionality classifications of nanostructures. *Physica E*. 40 (7), 2521-2525.

Rumble, J., Freiman, S., & Teague C. (2012) *The Description of Materials*. ISCU-CODATA workshop, Paris, 23-24/02/2012. Retrieved from the World Wide Web, October 20, 2012: <http://www.codata.org/Nanomaterials/Index-agenda-Nanomaterial.html>

Shenderova, O.A. & Gruen, D.M. (Eds.) (2006) *Ultrananocrystalline Diamond. Synthesis, Properties, and Applications*, Norwich - New York: William Andrew Publ.

Stones, R. & Matthew, N. (2005) *Beginning Databases with PostgreSQL. From Novice to Professional. Second Edition*. New York: Apress, Inc.

Suzdalev, I.P. & Suzdalev, P.I. (2001) Nanoclusters and nanocluster systems. Assembling, interactions and properties. *Russian Chemical Reviews* 70(3), 177–210.

Yang, C.C. & Li, S. (2008) Size-Dependent Temperature–Pressure Phase Diagram of Carbon. *J. Phys. Chem. C* 112 (5), 1423-1426.

Yu, M. Chaudhuri, I., Leahy, C., Wu, S.Y., & Jayanthi, C.S. (2009) Energetics, relative stabilities, and size-dependent properties of nanosized carbon clusters of different families: Fullerenes, bucky-diamond, icosahedral, and bulk-truncated structures. *J. Chem. Phys.* 130, Ar#184708.

(Article history: Received 2 June 2011, Accepted 28 November 2012, Available online 11 December 2012)