ISSN 0147-6882, Scientific and Technical Information Processing, 2015, Vol. 42, No. 4, pp. 211–222. © Allerton Press, Inc., 2015.
Original Russian Text © A.O. Erkimbaev, V.Yu. Zitserman, G.A. Kobzev, M.S. Trakhtengerts, 2015, published in Nauchno-Technicheskaya Informatsiya, Seriya 1, 2015, No. 10, pp. 8–20.

A Universal Metadata System for the Characterization of Nanomaterials

A. O. Erkimbaev, V. Yu. Zitserman, G. A. Kobzev, and M. S. Trakhtengerts

Joint Institute for High Temperatures, Russian Academy of Sciences, Moscow, Russia

e-mail: adilbek@ihed.ras.ru, vz1941@mail.ru, gkbz@mail.ru, trachtengerts@yahoo.com

Received May 26, 2015

Abstract—This paper presents a detailed analysis of a newly developed system for the identification of nanomaterials and nanostructures of an arbitrary nature. The system has been created as a result of the joint work of the Committee on Data for Science and Technology (CODATA) and the International Council for Science (ICSI). It is shown that at its core the developed system is an extended set of metadata that reflect the conceptual framework of a scientific discipline. Their features include a wide coverage of the sum of the properties and characteristics, as well as the possibility of adding new headings, i.e., support for the evolution of the conceptual scheme that meets the requirements of the subject field. Limitations in the possibilities of the representation of certain categories of nanomaterials, especially of macroscopic objects, are found. It is shown that it is possible to overcome them by the active use of existing ontologies in the field of materials science.

Keywords: metadata, nanostructures, nanomaterials, reference data

DOI: 10.3103/S0147688215040024

INTRODUCTION

Metadata (MD), or in other words, "data about data," are an indispensable attribute of any information storage, from traditional libraries to computer databases, Web-portals, and social networks. A more precise definition that was given in [1] treats metadata as "the key to understanding the content of an information system that provides identification, classification, positioning, integration, and interpretation of information." In the normal practice of libraries or archives metadata describe the resource by providing its address in the repository or online, information on the authors and publishers, search indexes, etc. The role and significance of metadata increase sharply during the transition to scientific databases (DB). In particular, in the areas where the collection, processing, and exchange of data constitute the main part of the workflow, for example, in bio- and geoinformatics, climate research, and astronomical observations. In this case, the metadata structure, which is not limited to the bibliographic description, includes a set of attributes that are specific to the respective discipline that allows DB users to conduct highly targeted searches and allows the creators to effectively exchange data with thematically similar resources. There is a vast depository (www.dcc.ac.uk/), where metadata standards for a variety of disciplines, from astronomy to medicine and agriculture, are presented. As a rule, these standards define the information domain model that combines general information (the name and participants of the study, its timeframe, access conditions, formats and addresses of data sets, etc.) with metadata that provide details on the subject and method of study. Typical attributes that reveal the scientific content of data include the information about their origin (experiment, observation, and modeling), quality and uncertainty, measurement units, and nomenclature of objects, their properties, and states.

Many of these concepts can be determined by the reference to the corresponding position in the qualifiers or controlled vocabularies on the network. Thus, in chemistry the substance name can be determined by referring to a unique identifier, viz., a CAS Registry number that eliminates the ambiguity of synonyms and language differences. The reference to the IUPAC Compendium of Chemical Terminology [2] included in the metadata makes it possible to accurately identify almost all chemical terms and concepts that are associated with the structure and properties of matter, reaction mechanisms, etc. As an example, a reference to the URL http://goldbook.iupac.org/R05396.html uniquely defines the term *ring-opening polymerization*. A clear selection of terms by the reference to the appropriate dictionary eliminates the arbitrariness and ambiguity in interpretation while providing the use of software agents in the search for data and interpretation of concepts [3].

In [4], the authors analyzed the rich opportunities that are provided by metadata accompanied by a cer-

tain category of data on the properties of substances and materials. For the thermophysical properties of substances the corresponding system of metadata was developed in the form of ThermoML, which is a subject-oriented version of the XML language [5]. The proposed multi-level hierarchy of metadata made it possible to formulate the basis of the conceptual model, which includes the entire range of objects, their properties, phase states, methods of measurement and purification of samples, types of uncertainty, etc. The system can describe the data for more than 120 thermophysical properties from tables or analytical presentation forms and different data statuses (experimental, computational, and reference). Pure substances and mixtures of constant composition, as well as chemical and phase reactions, can be considered as objects. A detailed scheme of data presentation was developed that is now available online [6]. A similar scheme of the XML language is proposed for the representation and exchange of data in the field of materials science (MatML language [7]), the structure of which is filled with references to production conditions, technologies, and test methods.

Naturally, the extension of metadata tools to nanotechnology is planned given the unprecedented scale of this discipline and the increased information flow in the form of publications, databases, Web-portals, etc. associated with it. However, a number of problems arise when trying to systematize these data. These problems are caused by the subject specificity and its place in the common system of scientific and engineering knowledge. The analysis that has been carried out in a number of studies [8–10] has revealed an entire range of features that are related to nanomaterials:

(1) the breadth and fuzzy boundaries of the domain with the integration of many fields of physics, chemistry, biomedicine, etc.

(2) the rapid evolution of the discipline followed by the emergence of new materials, devices, as well as the introduction of new concepts and features.

(3) the multifactorial description of the nature of objects (nanomaterials, devices, etc.) with the involvement of a large set of parameters that determine structure, composition, morphology, and other features.

(4) the relationship of the nomenclature of properties and nanomaterial class and the combination of molecular and macroscopic characteristics in the nomenclature.

As a result, the classification of data on nanomaterials is much superior to the knowledge that has been generated in the traditional areas, such as in materials science and chemistry in terms of the complexity of classification. Repeated attempts to develop nanotechnology classifiers [9] have depended on a narrow segment and a limited objective (e.g., the rubrication of the Federal portal (www.nanoportal.ru)), and usually quickly lagged behind the current state of the domain. In order to resolve the problems that significantly hinder its development, international efforts have been undertaken by the Committee on Data for Science and Technology (CODATA) and the International Council for Science (ICSI), which created a special working group aimed at the development of general recommendations that are suitable for a wide range of experts. The objectives and principles of this work were described in [10, 11]. As a result, in 2015 a document entitled the Uniform Description System (UDS) was developed [12]. It is intended to provide a universal approach to the description of both nanomaterials themselves and their properties, production technologies, certification, technical documentation, etc. The UDS is designed to incorporate uniqueness in the definition of nanomaterials during their characterization, i.e., to reliably select one of the objects from the set of objects close to it in terms of properties or to confirm the identity of objects with regard to a set of attributes: size and shape, structure and chemical nature, surface state, synthesis conditions, etc.

In fact, the proposed document forms the metadata system, although it does not have their formal characteristics, such as an XML schema or ontology. It has the form of a text document for a multilevel (and multidimensional) questionnaire, so that a set of answers to questions posed to the expert gave a complete picture of the in accordance the conceptual scheme that was taken as a basis. Two aspects distinguish this system from similar ones [13, 14]: the maximum coverage of the entire set of properties and features and the ability to add new headings to the system, i.e., the support of the evolution of the conceptual scheme, which gives hope that most of the data systematization problems will be solved.

In this paper, upon detailed examination of the metadata system features we will try to define three issues: (1) how full and adequate the proposed supply system is, (2) the areas in which the work on the detailization and expansion of the system should be conducted, and (3) how to bring the system to the formalization level that meets the used metadata standards. Note that the UDS authors themselves see their system as a possible basis for the formation of the scheme of a DB or ontology and avoid their actual construction at this stage.

THE *UDS* LOGICAL STRUCTURE AND FORMAL IDENTIFICATION OF NANOMATERIALS

Like most metadata systems, the proposed description is constructed as a hierarchy of relevant information units called *categories*. Four top-level categories (Fig. 1) should provide the maximum completeness of the data that is necessary for a unique and comprehensive representation of nanomaterials. The first of these (**General Identifiers**) gives a formal definition of the material, including the assignment of names to this material and their assignment to the selected headings

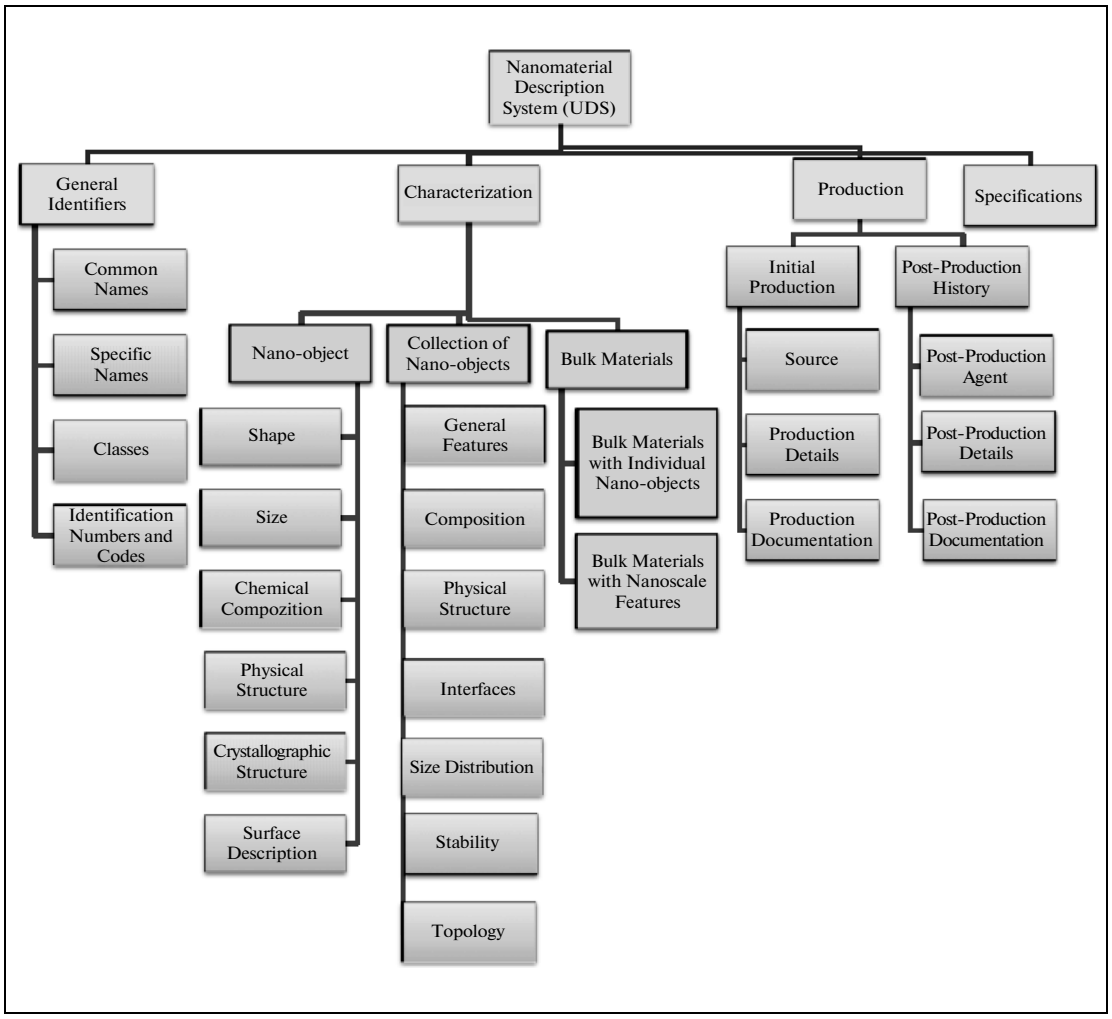

Fig. 1. The hierarchy of the main categories in the UDS [12].

of the classification scheme. This is the simplest identification level that does not affect the physical characteristics of an object and information about its production. The most complete identification is given by the second category (**Characterization**), in which all the items that are required for the discovery of information about the properties (the shape and size, chemical composition, internal and surface structure, etc.) are listed in detail. This set of data should provide an unambiguous selection of the described object from the set of other objects with a related structure. The last two categories (**Production** and **Specification**) cover all of the aspects that are related to the manufacture and supply of material to the market.

Each of these categories gives rise to the next level category (or subcategory). Thus, the first category **General Identifies** (see Fig. 1) is divided into four subcategories that correspond to different types of naming and classification of nanomaterials, while the category **Characterization** is divided into three subcategories according to its possible types. At the next level, categories that cover the physical characteristics of a material (shape and size, chemical composition, etc.) are introduced. A set of these characteristics (i.e., the characterization method) depends on the material type, which confirms the relationship between the nomenclature of properties and the nanomaterial class, as noted in the Introduction of this paper.

The lowest level of the hierarchy is represented by the so-called *descriptors,* i.e., the metadata that reveal the content of the result of measurements or calculations of corresponding characteristics with a descriptor that defines the data representation format as well: numeric (integer and float) or text (string). For example, in order to characterize the form of a nano-object (particle, cluster, nanotube, etc.), descriptors are necessary, viz., **the number of measurements on the nanoscale** or **the number of layers,** which are defined by integers, such as the general title of a **form** or **a geometric regularity** that assume a free text format. Obviously, descriptors that refer to the categories **Production** and **Specification** correspond not to physical characteris-

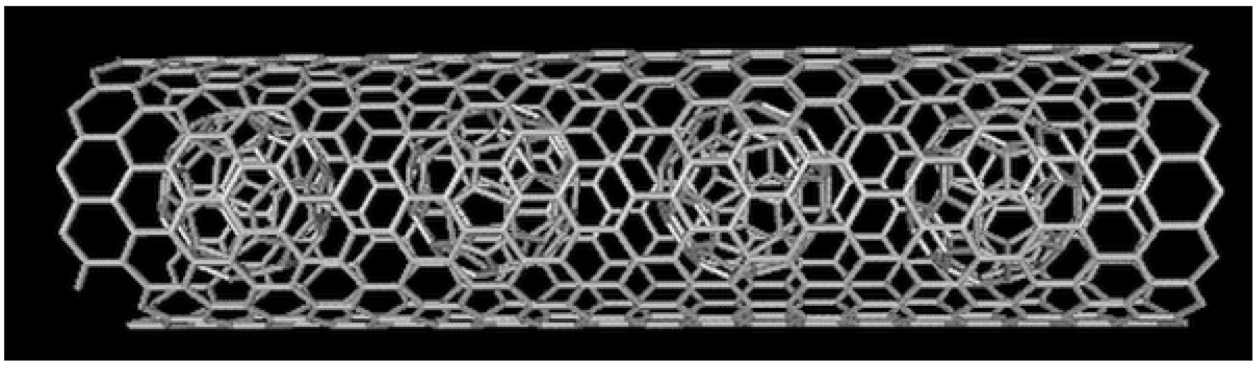

Fig. 2. The schematic view of a typical ensemble of nano-objects, the structure called a peapod.

tics but to technologies or documents. For example, the details of the technology are revealed by such descriptors as **the number of components** (belongs to the category **Source materials**), **General Description** (belongs to the category **Production Methods**) or **temperature** and **pressure** (belongs to the category **Conditions**). The first two of these descriptors require data in the text format and the next two require data in the numeric format with a floating point.

If the hierarchy of categories and subcategories structures the subject field by forming its conceptual scheme, at the level of descriptors it is possible to implement one of the most important metadata requirements that arise from the nature of nanomaterials, i.e., the need for a permanent adjustment related to the selection of a particular class and formulated tasks: basic research, the development of a commercial product, production technology, toxicology and environmental assessment, etc. Therefore, as noted in [12] in the general case the use of all descriptors is not required and their selection depends on the need for specific data. However, with the advent of new materials and methods for their production, as well as new features and events it can be necessary to expand the metadata system with new descriptors.

Typology of nanomaterials. As can be seen from Fig. 1, the metadata structure is associated with a specific agreement (the so-called nano-tree) that was adopted by the ISO in order to systematize the entire world of nanostructures [15]. In this agreement, the basic concept is a nanomaterial, which can be in one of two forms: nano-object and nanostructured material. The first form includes objects with at least one size corresponding to the nanoscale (up to 100 nm), and macroscopic objects whose internal or surface structures correspond to nanoscale refer to the second form. The authors of [12] added to nano-objects their ensembles/collections that are formed intentionally or accidentally. Thus, the ensemble corresponds to the same definition as the nano-object, i.e., it has from one to three sizes in the nanoscale. Examples of such objects are the so-called peapods that are formed by filling the inside of a carbon nanotube (CNT) with several fullerene molecules, C_{60} or C_{70} (Fig. 2).

Another change that was made in the nanomaterial systematics is associated with bulk materials; it is believed that they can be divided into two subclasses (see Fig. 1). The first subclass includes materials that contain identifiable nano-objects and the materials that exhibit a size effect because of the nano-sized internal or surface structure refer to the second one. Examples of materials of the first subclass are nanocomposites (e.g., a polymer with CNT inclusions) and of the second, nanostructured steels or ceramics.

The scheme in Fig. 1 captures only three basic types of nanomaterials but not their further division into more specific classes according to the shape, composition, etc., as was done in the ISO document [15]. This problem is solved by category **General Identifiers**.

Formal Identification of a Nanomaterial. The category **General Identifiers** includes the possible names of a nanomaterial and its assignment to a particular classification scheme. As an example, it is sufficient to assign one of the most common abbreviations (CNT) to popular objects (carbon nanotubes) and attribute them to the heading **1.1.2 Linear nanostructures** of the scientific heading list on the federal site Nanotechnology and Nanomaterials (www.portalnano.ru). In the UDS both tasks (naming and classification) are solved at two levels: arbitrary and in agreement with certain standards. It is proposed to use any English name that is encountered in the literature or their combination as an arbitrary name. For the cited example with carbon nanotubes these can be CNT and SWCNT.

As for the standardized name, in this case, as in chemistry, public classifiers should be used. In chemistry, in order to identify the substance, it is possible to refer to one of the most common classifications, such as the register of the Chemical Abstract Service, and the registers of basic reference books, viz., Beilstein for organic and Gmelin for inorganic substances. For nanomaterials the possibilities of such an identification are already limited because the discipline is quite

Table 1. The system of nanomaterials in the UDS, which was borrowed from the ISO document [15]

Nano-objects			Nanostructured materials
3D	2D	1D	
Nanoparticle	Nanofiber	Nanoplate	Nanopowder
Quantum dot	Nanotube		Nanocomposite
Onion	Nanorod		Solid nanofoam
	Nanowire		Nanoporous nanomaterial
			Nanofluid
			Nanoaerosol
			Nanostructured particle

new. In particular, these registers include only the nanostructures that can be determined by conventional chemical formulas, such as C_{60} or C_{70}.

A relatively wide range of nanomaterials is recorded in the chemical DB ChEBI [16], in which they are presented similar to conventional compounds. Each of the nanostructures is determined by a five-digit index, which can be used as a standardized name. For example, the index CHEBI:50796 determines nanotubes and CHEBI:50853 identifies quantum dots. A similar function can be assigned to the ontology NPO [13], which covers a larger list of nanomaterials, albeit with a bias towards medical and biological applications. Each object in this ontology is defined by the ID of the type of NPO_***; for example, NPO_606 determines CNT, and NPO_395 refers to a nanoparticle with a drug molecule embedded in its structure. The disadvantage of both systems is that they do not contain bulk nanomaterials (composites, powders, nanofluids, etc.), on which most of the technologies and devices are based.

The limited possibilities for the accurate naming of a nanomaterial compensate for its "embedding" in a classification scheme.[1] As is the case with names arbitrary schemas along with some classifications assigned by an authority may be usable. As a rule, an arbitrary classification is based on one of the physical characteristics of nanomaterials: topology, composition, manifestation of quantum effects, etc. Thus, the authors of [8, 9, 17] used a detailed classification based on topological features, when the number of measurements $K = 0–3$ at the macroscopic scale is taken as a basis. The value of $K = 0$ corresponds to the cluster and $K = 3$ corresponds to a macroscopic object in which the prefix "nano" refers only to the structural element. Intermediate values of $K = 1, 2$ correspond to one-dimensional and two-dimensional structures in which the macroscopic scale covers one or two dimensions (a nanowire or film). If the elements that make up the ensemble of nano-objects or bulk material are certified in the same way, it is possible to obtain a simple "nanoformula" of the type of **KD{L, M, N}**, where the indices L, M, and N refer to constituent elements of different types and the obvious inequality $K \geq \max\{L, M, N, \ldots\}$ holds. Thus, it is possible to classify not only nano-objects but also their ensembles, which is given by the scheme in Fig. 1.

With regard to carbon nanomaterials, it has repeatedly been proposed to classify them by the *hybridization* of the chemical bond, sp^{ν} index, where integer values $\nu = 1, 2, 3$ refer to the main allotropes of carbon (carbyne, graphite, and diamond) and intermediate ones that define mixed hybridization types cover many carbon nanoforms [18].

The authors of [12] recommend the so-called "nano-tree" that is presented in the ISO report [15] as the standard classifier. The nanomaterials are distributed by "tree branches" via the set of attributes, of which the topological attribute, the number of measurements at the nanoscale, which ranges from 1 to 3, is in first place.[2] In the case of this classification principle, clusters and nanoparticles have the dimension 3, while a nanowire or nanotube has the dimension 2 (two nanosizes in section); finally, the dimension of a film or graphene is 1 (one nanosize for the thickness). Each of the groups that are defined by these dimensions was divided into three subgroups: single-component, multicomponent, and nanostructured. Somewhat changed elements of this hierarchy are proposed in the UDS as a possible standard classifier. Changes only occur in that the mentioned three subgroups are brought to the top level, as shown in Fig. 1, and the dimensional sign is applied only to nano-objects (see Table 1, which shows typical nanomaterial groups).

Along with nano-tree that was recommended in the UDS as a recognized classifier (*assigned by an author-*

[1] This refers to the classifier of nanomaterials themselves, unlike the entire scheme in Fig. 1, which also includes their properties and production history.

[2] The separation of nanomaterials by the number of measurements at the nanoscale does not coincide with the common scale based on macroscopic measurements; in the latter case, the cluster dimension is 0 and the nanowire dimension is 1, etc. This classification was adopted in [8, 9, 17].

Table 2. The set of headings that is used in the PACS to select complex structured nanostructures

PACS codes	Complex nanostructures
81.07.Pr	Organic-inorganic hybrid nanostructures
81.07.Lk	Nanocontacts
78.67.Pt	Multilayers; superlattices; photonic structures; metamaterials
78.67.Sc	Nanoaggregates; nanocomposites
78.67.Ve	Nanomicelles
63.22.Np	Layered systems
62.23.St	Complex nanostructures, including patterned or assembled structures

ity in the terminology of [12]), other classifiers can be used, for example, an application for the PACS (Physics & Astronomy Classification Scheme) (www.aip.org/publishing/pacs/nano-supplement) described in detail in [9]. Its advantage is that it covers a large number of nanomaterials in accordance with different characteristics including complex structures such as ensembles or nanoproducts (Table 2).

Finally, the DB ChEBI [16] that was proposed for the naming of nanomaterials can be used to assign them to a specific heading. Each of the headings, depending on its place in the hierarchy, can act both as an nanomaterial identifier and as the class that covers the entire group of them, which is illustrated in Fig. 3, which shows fragments of the hierarchy of nanotubes: in the left side of the figure the code CHEBI:50796 defines the nanomaterial class that includes all of the nanomaterials recorded below; their codes can be regarded as standardized names. However, some of them, such as CHEBI:50594 for a carbon nanotube, can be considered as class identifiers and "subordinate" ones, as nanomaterial identifiers, for example CHEBI:50595 for single-walled carbon nanotubes.

Neither prepositional scheme can cover the entire set of nanomaterials, including newly synthesized ones. Therefore, the opportunities for its adjustment for the subject field and the needs of the expert community that are provided for in the UDS are very

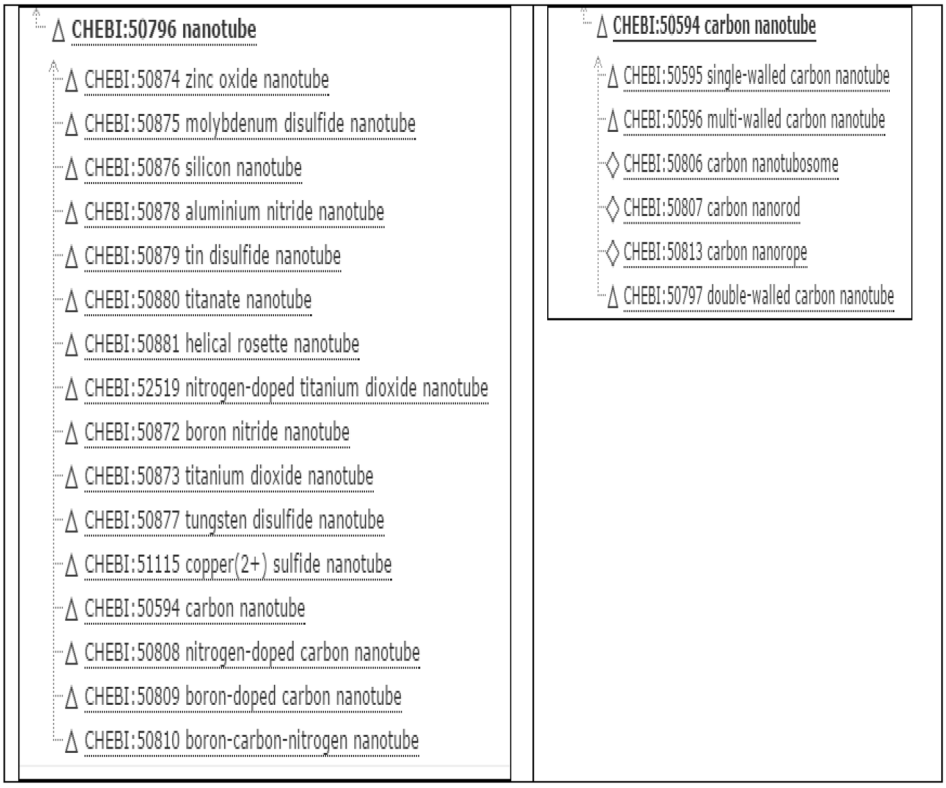

Fig. 3. Fragments of the hierarchy of the nanomaterial from the DB ChEBI [16].

Table 3. The categories and subcategories of the UDS that have been proposed to characterize individual nano-objects

Shape	Size	Chemical composition	Physical structure	Crystallographic structure	Surface description
Shape Type	Applicable Dimensions	Atomic Composition	Layered Nano-Object	Physical Structure Identification	General Surface Description
Shape Features	Derived Dimensions	Molecular Composition	Shell Structure	Unit Cell Information	Surface Treatment
	Internal Dimensions	Chemical Moieties	Physical Features	Basic Unit Cell Parameters	Surface Geometry
			Defects		Surface Electronic Properties
			Entrapment		Other Surface Properties
			Additions		

important. In [12] it was stated that a nanomaterial cannot be identified by a simple name (or a set of names), which distinguishes them from ordinary molecules. The solution to this problem, as previously emphasized in [8–11], lies in collecting the maximum amount of information on the properties and production technology, which is provided by the UDS in a systematized way.

IDENTIFICATION OF NANOMATERIALS BY THE COMBINATION OF PHYSICAL PROPERTIES

The main ideas of the multifactorial characterization of nanomaterial were expressed by the UDS authors in a report that was presented at the working group that was held in 2012 in Paris under the auspices of the ICSU and CODATA [11]. They proceeded from the factors that determine the key differences of nanomaterials from traditional materials (steel, alloys, ceramics, etc.): a large value of the surface/volume ratio, differences in the electronic structure of the surface and volume, quantum size effects; components that "hang" on the surface, chemical reactivities when comparing nanomaterials with the macroscopic analog, the synthesis of a nanoform that was previously unknown in the macrocosm (CNT, graphene, etc.), the strong influence of small impurities, the self-assembly of ordered nanostructures, and a heterogeneous structure at a small scale.

In addition to differences in the properties of nanomaterials, the authors of [12] took the diversity of the disciplines that need to give a detailed description of the nanomaterial and the problems that were solved into account: the design of the devices, prediction of properties, toxicological and environmental assessments, standards development, etc. The requirements to the list of determining factors were formulated on this basis. This list included only the factors that are necessary for identification. Many of the important nanomaterial properties (thermodynamic, electrical, optical, etc.) were not taken into account, thus reducing the entire metadata volume to a reasonable minimum. In accordance with the general metadata scheme (Fig. 1), this list is different for individual nano-objects, their ensembles (nanoproducts), and macroscopic materials.

The structure of metadata for individual nano-objects. Table 3 shows the final requirements for nanomaterials in the form of a hierarchy of categories that determine nanomaterials by geometry (size and shape), composition, physical characteristics, and data about the surface. Certain aspects of each of the characteristics have subcategories, for example, defining, on the one hand, the form type (cylinder, sphere, etc.) and, on the other hand, its features. The next (the lowest) level of the hierarchy is occupied by descriptors that define the data element that was proposed by an expert. The number of descriptors is sufficiently large in accordance with the general requirement for the multifactorial description of nanomaterials. Simultaneously, each of the UDS descriptors has an accurate *definition* that defines the content and format of the input data. For example, Table 4 shows a set of descriptors that are related to the shape and chemical composition of nanomaterials.

In addition to two major formats (text and numeric), the format called *enumeration*, where a data

Table 4. Descriptors that define the shape and chemical composition of nanomaterials

Shape		Chemical Composition		
Shape Type	Shape Features	Atomic composition	Molecular composition	Chemical Moieties
Number of dimensions on nanoscale	Type of feature	Number of different types of atom present	Number of different types of molecule present	
General shape	Regularity of feature	Atom present	Molecules present	
ISO 229 shape name	Number of feature	Composition percentage type	Molecular formula	
Specific shape	Symmetry of feature	Composition percentage	Molecular name	
Type of thickness of a nano-object with one dimension at the nanoscale		Type of composition	Structural formula	
Cross-sectional view for nano-object with two dimensions at the nanoscale		How measured	CAS Registry number	
Number of layers (for nano-object with two or three dimensions at the nanoscale)			IUPAC InChI	
Geometric regularity			Composition percentage type	
Shape symmetry			Composition percentage	
Symmetry components			Type of composition	
			How measured	

The list of descriptors in the subcategory **Chemical Moieties** (part or the functional group of the molecule) is identical to that given in the subcategory **Molecular Composition**.

element is obtained from a limited list or an external controlled vocabulary, is used in the UDS (although it is not specifically stipulated). Thus, the ***ISO 229 Shape name*** descriptor (Table 4) determines the form in accordance with a list of five possible forms, viz., a nanoparticle, nanorod, nanotube, nanoplate, and nanocone, that were adopted in the ISO standard [16]. Similarly, when determining the composition of a nano-object, names and designations of atoms are specified in accordance with the periodic table and the molecule is uniquely identify by a number from the CAS registry. Other descriptors that use this format (in Table 4 they are shaded darker) make it possible to determine the component percentage (by mass or number of particles), "composition type" (a separate measurement, a calculated value, or the result of averaging), and measurement method (the descriptor ***how measured***) from the list. This format appears to be very effective for creating the details of the category **Physical Structure**. As an example, in order to reveal the content of **Physical Features**, an a priori prepared list of them was proposed (hole, protuberance/appendage, end, cap, legs, etc.) that make it possible to reflect the nano-object morphology.

The *enumeration* format was widely used in ThermoML [5], a system of metadata for the representation of the thermal properties of materials that makes it possible to fix the property type, phase state, methods of measurement and cleaning, data status, type of uncertainty, etc. This format eliminates the arbitrariness in the recording of data and reveals rich opportunities in the machine processing in the search and justification of an inference. Note that the text format (string) also does not always presuppose a free description, but it can follow strict rules as, for example, in the recording of a molecular formula, name, and structural formula that follow the accepted standard of one of the public chemical DBs, for example, ChEBI [16] or ChemSpider (www.chemspider.com). The descriptor ***IUPAC InChI*** contains a special type of chemical data, the so-called linear notation, as a string of characters that encode information about the structure and topology of the molecule.

A complete list of all of the descriptors is presented in the text of the UDS guidelines at the IUPAC site [12]). By the extension or modification of the list of descriptors it is possible to *fine tune* metadata struc-

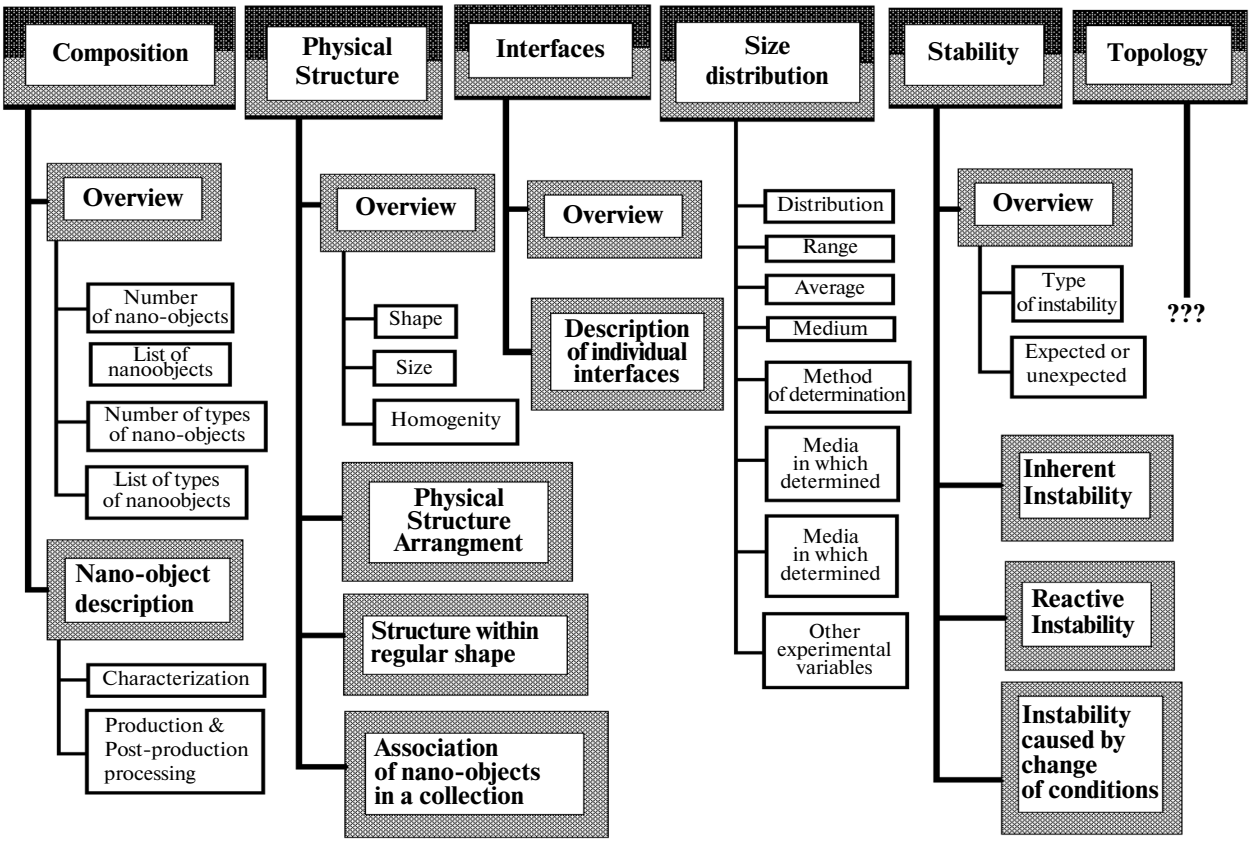

Fig. 4. The hierarchy of categories and descriptors that define an ensemble of nano-objects.

tures for requirements that are caused by the characteristics of objects or tasks.

Metadata structure for an ensemble of nano-objects. As can be seen from the diagram in Fig. 4, the metadata structure can be divided into three main blocks: the ensemble composition, description of each of its components, and description of the ensemble as a whole. The composition is set almost the same as the composition of the molecule, i.e., types of individual nano-objects and their number in the ensemble are specified. As an example, for the structure in Fig. 2 (a peapod), the description should designate two types of nano-objects (CNT and fullerene) in the quantity of one CNT and four fullerene molecules. The next block should describe each nano-object in accordance with the predetermined structure (Table 3) along with information on the production cycle. Finally, five categories (**Physical Structure, Interfaces, Size Distribution, Stability, and Topology**) represent different aspects of the ensemble as a whole. The first of these categories gives the general characteristics: the shape and size, relative position of objects, their association, etc. The category **Interfaces** includes a description of the surfaces that separate the individual nano-objects, their shape, dimension, regularity, etc. The majority of the descriptors provide only a qualitative characteristic. In contrast, descriptors of the category **Size Distribution** mainly involve numerical characteristics, while the measurement method and the medium can be an element of the classifier (*enumeration* format).

The category **Stability** determines the ensemble stability with respect to the spontaneous decay or external impact: thermal, chemical, etc. First, the subcategory **Overview** sets the instability type from a list and describes the decay expectancy. The following subcategories reveal these instability types in detail.

Finally, the last category (**Topology**) in the interpretation of the authors of [12] characterizes the coherence and continuity of the ensemble that manifest themselves in their influence on the properties. Molecular biology provides an example, where the object functionality depends on the *homeomorphism,* i.e., on the ability of an object to continuously transform into another one without gaps and binding. In general, ensemble topology is quite a subtle concept that can be revealed only on a qualitative level without hierarchical layering.

Bulk (Macroscopic) nanomaterials. A special category hierarchy was not proposed with regard to this nanomaterial class. For the two possible types of bulk nanomaterials that are marked in Fig. 1, the possibilities of the UDS are different. If the bulk nanomaterial includes identifiable nano-objects, an important role

Table 5. The final characteristics of a commercial product

Post-Production Process Result
Nanomaterial produced
Purity
Composition
Yield
Physical state
Date produced
Location
Post-producing organization
Batch number
Post-producing documentation

is played by the degree of freedom that these nano-objects have within the respective phase. In both the liquid phase and the solid phase, the free movement makes it possible to use only the regulations that are proposed for the nano-object itself or for its ensemble. If a nano-object (for example, a metal nanoparticle) is rigidly fixed in a solid matrix, then its description should be accompanied by the description of the matrix itself following the standards that are proposed for typical materials: metals, ceramics, polymers, etc. As an example, for the nanocomposite formed by the CNT inclusion in the polymer matrix it will be necessary to specify both materials. Obviously, there can be intermediate cases of a nano-object with limited freedom within the matrix, which will require the improvement of the UDS.

The second of the two mentioned types in Fig. 1 (*Bulk Materials with Nanoscale Features*) covers nanostructured materials (steel, ceramics, etc.) whose elements (crystallites) fall into the nanoscale, which significantly affects physical and performance properties. Materials of this type are already widely used in mechanical engineering, the power industry, nuclear technology, and other branches of the economy [19, 20]. Unfortunately, the current UDS version does not make it possible to identify bulk nanomaterials. In the last section of this paper, we will discuss the possibility of their description based on the available ontologies in the field of materials science.

METADATA FOR PRODUCTION HISTORY AND DELIVERY CONDITIONS

The last two categories of the top-level hierarchy in Fig. 1 (**Production** and **Specification**) relate to the production and certification of nanomaterials as a commercial product. A key feature of nanomaterials, i.e., a strong dependence of their properties on the manufacturing technology, forces one to reflect the inextricable connection between properties and production conditions the metadata system. The completeness of information in the UDS is provided by the division of the category **Production** into two, which reflect the initial stage and stages following the production (see Fig. 1). For the initial stage metadata cover the composition and amount of the raw materials, method of production and formulation, equipment, nature and conditions of the environment, and the data for the desired product. A similar metadata structure in the category **Post-Production History** represents the next stages of the lifecycle: cleaning, storage, transport, etc. The resulting group of descriptors is the most significant. It is a kind of "passport" for a product that is delivered to the consumer (Table 5).

Once a nanomaterial becomes a commercial product, there is a need for a collection of accompanying documents that are combined by the term "specification," which is an integral part of any production, from raw materials to high-tech equipment. For each of the branches of science and production requirements for the composition and content of the specification are regulated by national or international standards. For nanomaterials, the guiding document is the ISO standard [21], which determines the required amount of data on the properties, conditions of storage and delivery, control measurements, etc. As for the category **Specification** in the UDS, it forms the backbone of metadata that define formal information: the name, version, and number of specifications, the date of its release and the official status of the manufacturer.

It should be noted that the hierarchy of categories that describe the production history and refer to the specification is almost identical for many sectors, especially those that are related to the production of materials.[3] Therefore, for these purposes, it is possible to use a number of vocabularies and ontologies of the general technical profile, for example, the standard ISO 10303-235: **Engineering Properties for Product Design and Verification** [22]. This will make it possible to somewhat reduce the load on the UDS by transferring links to other ontologies and vocabularies.

THE IMPROVEMENT POTENTIAL OF THE *UDS* SYSTEM

Certainly, the extreme complexity of the system of tasks that are posed by the developers, i.e., the unambiguous and comprehensive identification of all of the possible types of nanomaterials, does not make it possible to consider the proposed version as the final one. In addition to the variety of objects, the description standardization is complicated by the variety of requirements that the researchers, product developers, consumers, creators of standards, developers of health and environmental standards, computer scientists,

[3] This refers to the closeness of information models for different industries but not to the content of specifications which differ for each sector and type of product.

etc. have to deal with. In the UDS document [12] it was repeatedly emphasized that the accumulation of knowledge can require new descriptors, i.e., new positions in the metadata structure. Moreover, according to the authors of [12] there is almost no scientifically based foundation for the standardization of nanostructured bulk materials (the second group *Bulk Materials* shown in Fig. 1) or some characteristics of ensembles such as topology (see Fig. 4). Therefore, the developed metadata system stimulates work on its improvement and adaptation to individual segments of nanotechnology. There are two possible directions of this activity: (1) the revision and expansion of the logical structure, including classifications, selection of formats, and coordination with controlled vocabularies and ontologies, and (2) the expansion of the subject by the introduction of additional descriptors.

The first direction is a necessary condition in the case of the transition to strictly formalized ontologies or DBs, which is considered in the UDS as one of the most important applications of the developed system.[4] Some of the required steps were discussed during the preparation of this paper. Among them is the best possible use of descriptors of the *enumeration* format instead of free text, which is a prerequisite for the machine implementation of search and inference. This type of data can be used in the determination of chemical composition, the type of physical and morphological features, methods of measurement, etc.

The second direction is the widespread use of links to external resources in the form of controlled vocabularies, for example, the above-mentioned DB ChEBI [16] or ontology NPO [13] for the naming and systematization of nano-objects. External resources can also be used as sources of terminology, for example, to characterize types of surfaces or elements of a crystal structure. The compendium of chemical terminology **Goldbook** [2] and the CIF international standard of crystallographic data (www.iucr.org/resources/cif) are possible resources. The active involvement of vocabularies and ontologies that are available on the Internet is now regarded as one of the most important conditions for the global data integration [3].

The most difficult task in the UDS improvement is the identification of bulk materials, for which the system developers did not construct a taxonomy of categories and descriptors. The above method of searching for external sources makes it possible to find a natural way for its solution, while expanding the logical structure of metadata and the subject field. Accordingly, two types of bulk nanomaterials will require various resources. The first nanomaterial type includes bulk materials that contain identifiable nanoscale objects. Typical representatives of such materials are nanocomposites, for example, polymers with nanoscale inclusions. The ontology for this nanomaterial class has already been developed [14]. It is based on the taxonomy of classes and instances including classes such as **MaterialType, Nano-object, Nanocomposite,** as well as classes that combine chemical characteristics, such as **ChemicalIdentity, ChainComposition, and Structure**. The class **MaterialType** includes five subclasses (**CarbonBased, Ceramic, Metallic, Polymer, and SiliconBased**), which combined with a fairly simple taxonomy of nano-objects (**NanoFiber, NanoFilm, NanoLayer, NanoPowder, and NanoSurface**) makes it possible to reflect the formation of different types of nanocomposites that are determined by the choice of the matrix and filler. When building an ontology [14], the same principle was used that is recommended for the UDS expansion: the class **Polymer** was borrowed (together with its subclasses) from the ontology **ChemAxiom** and the class **Metallic,** from the ontology **PeriodicTable**.

However, the group of nanomaterials that contain identifiable nano-objects is not limited to composite materials. Many materials that are constructed of nano-objects without separation of the matrix and the filler, as in the case of a composite, can be named. An example of such a material is fullerite, a molecular crystal that is formed from C_{60} or C_{70} molecules because of the van der Waals interaction [23]. Such a crystal can also exist in the form of one-dimensional or two-dimensional polymers. Multiple films, fibers, and samples of nanothread and nanopaper are made of directed carbon nanotubes [24]. A macroscopic ultralight aerogel with a density on the order of mg/cm^{-3} gives a beautiful example of such a nanomaterial [25]. This aerogel has the form of a three-dimensional carbon structure, which is built from graphene sheets that are connected by the edges of the CNT. All such materials are inherently related to the ensemble of nano-objects given that macroscopic objects will occur in this class. For this purpose, lists of classes and properties that are used to characterize them should be extended so that it would be possible to represent such typical macroscopic properties, e.g., density, Young's modulus, and specific heat.

With regard to the second group of nanomaterials (see Fig. 1), which does not contain identifiable nano-objects, the current UDS version does not provide any recommendations for their description. However, in fact these materials can be attributed to the classes of traditional structural materials, which opens up the interesting possibility of using ontologies that were developed in materials science for data exchange on physical and performance characteristics. Nanomaterials include structural materials that acquire nanoscale features because of special techniques, i.e., a significant dependence of their properties on the size of the structural unit [19, 20]. Therefore, it is necessary to introduce a new parameter in the ontology along

[4] A phrase from the section Use of the Uniform Description System [12]: "The UDS provides a backbone for building the database schemas and ontologies that are at the core of a nanoinformatics resource so that information from different resources can be compared and contrasted correctly."

with the usual parameters that define the state of the environment (temperature, pressure, etc.). Moreover, during the characterization of a nanomaterial more subtle characteristics, for example, the distribution of structural units by size, can be required. In order to characterize such materials, including their mechanical, thermal, and service properties, one of the ontologies that was developed in materials science can be used, for example [26, 27]. However, when using the ontology on structural materials, it is necessary to accommodate the need to expand the set of state parameters.

Thus, the obvious way to expand the metadata system is its maximum possible integration with existing ontologies, both those that reflect some nanomaterial classes and general content ontologies that are related to chemistry, materials science, and other disciplines. This same strategy, as noted above, can be used to reflect not only the physical characteristics of nanomaterials but also the product lifecycle.

REFERENCES

1. Bezdushnyi, A.N., Vezhnevets, A.A., Serebryakov, V.A., and Shkotin, A.V., Metadata: Definition and use. www.elbib.ru/index.phtml?page=elbib/rus/methodology/metadata
2. IUPAC Compendium of Chemical Terminology - the Gold Book. http://goldbook.iupac.org/
3. Erkimbaev, A.O., Zitserman, V.Yu., Kobzev, G.A., Serebrjakov, V.A., and Teymurazov, K.B., Publishing scientific data as linked open data, *Sci. Tech. Inf. Process.*, 2013, vol. 40, no. 4, pp. 253–263.
4. Erkimbaev, A.O., Zitserman, V.Yu., and Kobzev, G.A., The role of metadata in the creation and application of information resources on the properties of substances and materials, *Sci. Tech. Inf. Process.*, 2008, vol. 35, no. 6, pp. 247–255.
5. Frenkel, M., Global communications and expert systems in thermodynamics: Connecting property measurement and chemical process design, *Pure Appl. Chem.*, 2005, vol. 77, no. 8, pp. 1349–1367.
6. ThermoML Schema. http://media.iupac.org/namespaces/ThermoML/ThermoML.xsd
7. Kaufman, J.G. and Begley, E.F., A data interchange markup language, *Adv. Mater. Processes*, 2003, pp. 35–36.
8. Eletskii, A.V., Erkimbaev, A.O., Zitserman, V.Yu., Kobzev, G.A., and Trakhtengerts, M.S., Thermophysical properties of nanoobjects: Data classification and validity evaluation, *High Temp.*, 2012, vol. 50, no. 4, pp. 488–495.
9. Builova, N.M., Eletskii, A.V., Zitserman, V.Yu., and Kobzev, G.A., The systematization of sources and data on nanotechnologies, *Sci. Tech. Inf. Process.*, 2013, vol. 40, no. 4, pp. 212–223.
10. Rumble, J. and Freiman, S., Describing nanomaterials: Meeting the needs of diverse data communities, *Data Sci. J.*, 2012, vol. 11, p. 1.
11. Rumble, J., Freiman, S., and Teague, C., The description of materials, *ICSU-CODATA Workshop on Nanomaterials*, 2012.
12. Uniform Description System for Materials on the Nanoscale. Prepared by the CODATA-VAMAS Working Group on the Description of Nanomaterials. http://www.codata.org/nanomaterials
13. NPO: NanoParticle Ontology for Cancer Nanotechnology Research. http://www.nano-ontology.org
14. Karpova, I., Poryseva, E., Kazakov, G., and Kol'tsova, E., The development of ontology in nanocomposite materials, *Inf. Resur. Ross.*, 2012, no. 2, pp. 5–9.
15. *ISO/TR 11360:2010, Nanotechologies - Methodology for the Classification and Categorization of Nanomaterials*, 2010. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=55967
16. Degtyarenko, K., de Matos, P., Ennis, M., et al., ChEBI: A database and ontology for chemical entities of biological interest, *Nucleic Acids Res.*, 2008, vol. 36, Data base issue. – D344–D350.
17. Pokropivny, V.V. and Skorokhod, V.V., New dimensionality classifications of nanostructures, *Physica E.*, 2008, vol. 40, no. 7, pp. 2521–2525.
18. Heimann, R.B., Evsyukov, S.E., and Koga, Y., Carbon allotropes: A suggested classification scheme based on valence orbital hybridization, *Carbon*, 1997, vol. 35, pp. 1654–1658.
19. Andrievskii, R.A., *Osnovy nanostrukturnogo materialovedeniya: vozmozhnosti i problemy* (Fundamentals of Nanostructured Materials Science: Opportunities and Challenges), Moscow: Binom. Laboratoriya Znanii, 2012.
20. Gusev, A.I., *Nanomaterialy, nanostruktury, nanotekhnologii* (Nanomaterials, Nanostructures, and Nanotechnology), Moscow: Fizmatlit, 2007, ed. 2.
21. *ISO/TS 12805:2011, Nanotechnologies-Material Specifications-Guidance on specifying nano-objects*, 2011. www.iso.org/obp/ui/#iso:std:51766:en]
22. Swindells, N., The representation and exchange of material and other engineering properties, *Data Sci. J.,* 2009, vol. 8, pp. 190–200.
23. Markin, A.V., Thermodynamics of crystal polymeric nanostructures of Fullerene-C60, *Cand. Sci. (Chem.) Dissertation*, Nizhni Novgorod, 2004.
24. Eletskii, A.V., Carbon nanotubes, *UFN*, 1997, vol. 167, no. 9, pp. 945–972.
25. Laverov, I., Aerogel of graphene and carbon nanotubes lacks flaws of its predecessors. http://elementy.ru/news/432045
26. Cheung, K., Hunter, J., and Drennan, J., Matseek: An ontology-based federated search interface for materials scientists, *IEEE Intell. Syst.*, 2009, vol. 24, no. 1, pp. 47–56.
27. Ashino, T. and Fujita, M., Definition of a web ontology for design-oriented material selection, *Data Sci. J.*, 2006, vol. 5, pp. 52–63.

Translated by O. Pismenov