The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/0264-0473.htm

Using a digital library instead of a traditional database

ABCD-based infrastructure for nanotechnology

Michael Trachtengerts, Adilbek Erkimbaev,
Vladimir Zitserman and Georgii Kobzev
Thermophysical Data Department,
Joint Institution for High Temperatures of Russian Academy of Sciences,
Moscow, Russia

Received 17 August 2014
Revised 17 August 2014
Accepted 13 October 2014

Abstract

Purpose – The purpose of this paper is to reveal main advantages of digital libraries in comparison with technology of common database for data-oriented fields of modern science. As an example, the subject domain "nanomaterials and nanotechnologies" with new features due to evolution of concepts and objects is presented.

Design/methodology/approach – An analysis of the information system ABCD as a basis for science-oriented digital library was fulfilled. Also, a survey of peculiarities of data in fast developing fields of science was prepared.

Findings – The results of this paper showed that functional capacities of ABCD satisfy requirements for complex collections and archives of scientific documents. Based on the ABCD tools and this concept, the digital library for storage and systematization of data and documents on nanomaterials and nanotechnologies for the power engineering was constructed. The library combines opportunities of bibliographic, full text and factual information systems.

Originality/value – This paper gives the foundation for creation of a library that combines services of bibliographic, full text and factual (numerical) information systems. Some analyses of ABCD tools were made before elsewhere, but they did not point on data peculiarities of complexly organized domains: semi-structured data, multitude formats (text, image and tables), interconnection of content with external sources located on other servers or in the Web.

Keywords Digital libraries, Data analysis, Nanomaterials, Information strategy, ABCD system, Database creation

Paper type Research paper

Introduction

The problems arising in gathering, mining and sharing of scientific data with common – usually relational – databases (DBs) are widely known and discussed by IT experts (Erkimbaev *et al.*, 2008; Fox and Harris, 2013). Restricted schemes of data descriptions grounded on sets of names and attributes for each relation table are not sufficiently robust for the sophisticated domain-specific structures of scientific data. As a rule, relational DBs support limited sets of data types that raise difficulties

The Electronic Library
Vol. 33 No. 5, 2015
pp. 916-927
© Emerald Group Publishing Limited
0264-0473
DOI 10.1108/EL-08-2014-0136

This work was supported by the Russian Foundation for Basic Research (Grant No. 13-08-00404) and the target funding program "Informatization" of the Russian Academy of Sciences.

during the expansions of subject domains. At the same time, modern data-dominated science with its extreme variety in domain-specific types requires new approaches to organize complex and changing data structures. Thus, it is difficult to apply the conceptual model inherent to each of the scientific disciplines with the predefined schemes of a relational DB. To some degree, the same problems are seen also in object-oriented and object-relational DBs.

There are also other reasons to expect difficulties in digital data management. Among these are the diversity of non-structured documents which need to be included in the data infrastructure, including the results of physical, natural and computer experiments; three-dimensional (3D) objects; temporal series; and multimedia. Representing the enormous volume of data produced by some science sectors (e.g. earth and space science, biomedicine, material science and others) also imposes a problem – the "big data problem". While the absolute size of the data explosion varies among disciplines, the general trend in rapid growth in all disciplines from astronomy to biomedicine can be clearly seen. According to an estimation provided in Fox and Harris (2013), by the year 2020, the annual growth in volume of scientific data will be approximately 35 zettabytes (10^{21} bytes). Traditional DBs are not suitable for processing such large amounts of multi-dimensional data. They are comparatively low in productivity because of the large number of keys required to fulfill a necessary search, the fixing histories of transactions and their specific memory management.

The scientific community has suggested a set of alternative approaches to overcome the limitations of traditional DBs for use by data-intensive science. Today, contemporary scientific work environments and research data infrastructures are based on a wide range of different conceptions and technical architectures. Among them are domain-specific versions of XML, for example, CML, MatML, ThermoML and so forth (Murray-Rust and Rzepa, 2002a, 2002b) and various tools for data storage and exchange. To address this problem, increasing attention in recent years has been paid to the Semantic Web which helps to integrate data sets and services with the help of ontology-based techniques (Bizer, 2013; Erkimbaev *et al.*, 2013).

Many scientific disciplines and technologies are focused on semi-structured data (SSD). This applies when data do not fit into a rigidly predefined schema, as required in traditional DBs. Eletskii *et al.* (2012a, 2012b) and Stones and Matthew (2005) showed that simpler, though less universal, decisions based on the object-relation DBs system with open code PostgreSQL may be applied. The system combines a traditional relational data model with maintenance of "fuzzy" data structures, commonly used in nanotechnologies.

In this publication, the authors demonstrate that many of the problems pertaining to e-science infrastructure can be solved by moving away from the concepts of traditional DB and moving to digital libraries (DLs) instead. DLs offer systems of new quality for storage and integration of thematic resources. At the Santa Fe workshop held in 1997, when declaring Phase 2 of the Digital Libraries Initiative program (DLI-2), the symposium content emphasized that DLs would be more powerful when compared to traditional public libraries (Griffin *et al.*, 2005). As a result, there are a variety of resources for developing the necessary environments capable to realize various functions for special libraries. Extra functions of DLs include:

- storage and data processing for a wide set of data that differ both in content and in types of media (textual, image, audio, etc.);
- multi-dimensional categorization according to the conceptual scheme of a subject domain;
- support of dynamic (rapidly changing) structures of resources and conceptual schemas; and
- integration of collected documents with documents/data of the same domain presented in a global network.

As to science as a whole, and especially to some its special branches (chemistry, material science, life sciences), the possibility of the so-called *semantic enrichment* is best matched to modern practices for science data management. In this approach, terms or data presented in an electronic publication are linked to relevant information in Web portals and sites. They may give access to reference books, encyclopedias and other valuable sources on library servers or elsewhere. In this paper, the authors analyzed some features of a rather simple and freely available DL tool referred to as ABCD. ABCD is the acronym for the project that in Spanish has the following full name: *Automatisación de Bibliotécas y Centros de Documentación* (Trakhtengerts, 2009; De Smet, 2009). It is a multi-lingual Web application used to manage document collections in libraries and information centers, providing such functions as data collection, inventory, categorization, export/import in various formats, administration and more. The system can be used by small or large groups of scientists. In spite of the simplicity and compactness of this system, it is possible to use it to create an electronic library of documents and data for nanomaterials applied in power engineering that nowadays presents a complicated and rapidly developing domain. Publications such as Eletskii *et al.* (2012a, 2012b) and Rumble and Freiman (2012) showed many special difficulties in data and document categorization that stem from novelty and dynamism of arising problems. Techniques for standardization, conceptual framework, classification schemes and terminology are still in progress. This paper presents the new DL approach as having advantages over common information systems.

Peculiarities of data and documents in nanoscience

Nanotechnology and nanoscience as a whole is a rather complicated and interesting subject domain with respect to structures of data and documents (Eletskii *et al.*, 2012a, 2012b; Rumble and Freiman, 2012). First of all, identification of nano-objects should involve a set of quantitative and qualitative features concerning structure, size, morphology and synthesis method. These features distinctly differentiate a nano-object from common substances or solutions defined solely by chemical composition and/or structural formula. As this takes place, a number of parameters and their choice vary significantly, depending on nanostructure class and corresponding SSD model (Abiteboul *et al.*, 2000). Changing lists of descriptive signs is the main reason for the impossibility of managing SSD by relational structures. In addition, some problems concerning classification also arise. Classification schemes are prone to variations, similar to properties nomenclature. Regular updating of the classification scheme appears to be a distinctive feature of the SSD model.

According to ISO recommendations (ISO/TR 11360, 2012), the world of nanostructures may be divided into two mega classes: nanostructured materials and nano-objects. Both of these were classified by dimension as well. Macroscopic substances containing nano-sized units refer to the first category, while the second one covers all imaginable kinds of synthesized nanostructures with the mandatory presence of at least one nano-metric dimension. In relation to classification, it is necessary to consider, in addition to the dimensional factor, the chemical features, structure, surface condition and a number of other impacts. During the evolution of the subject domain, some new headings in new classes arise, for example, graphene-based materials. Such depth and evolution of the subject area needs regular updating of qualifiers and reservation for some positions for new materials, devices and technologies.

The significant feature of nanoscience data is the intimate linking with external sources. These data contain a wealth of contextual information about synthesis/production methods, methods of research for physical properties or application features and operational features of devices. In pursuing the goal of assessing materials data quality in a developing DB, a special concept of materials metrology (Munro, 2003) was designed. The principles and practices set forth in this work demand that the results of measurements are presented in line with the data methods applied. The reliability of numerical values should be defined with the necessary scope of the available data. The importance of these procedures highlight the steps toward nanoworld objects due to the rising number of additional factors in synthesis and/or measurements.

The second type of contextual information consists of handbooks, thesauruses and reviews available on servers or on the Web. Applying these links raises the information value of concrete data sets due to the evolution of general information about related classes of nanomaterials (nanotubes, graphenes, etc.). Generally speaking, many science fields have a special need to consider concrete data in conjunction with third-party documents.

An urgent issue for nanotechnologies is the enormous quantity of sources producing the flow of updating factual data. It seems possible to overcome these problems by a combination of bibliographic data with full text, tables of numerical data and graphics. Such combinations allow for the management of optimum collection volume with the use of external reference data. In this case, the collection of bibliographic data from external sources may be fulfilled in a semi-automatic mode. For example, the Web of Science from Thomson Reuters may provide a fast data stream for a collection (Trakhtengerts, 2011). Thus, peculiarities of subject domain data impose heavy demands on the functionalities of a DL, including widespread use of contextual information, integration of diverse resources and a variety of data types and formats (text, images, tables, mathematical expressions, etc.).

Methodology

ABCD was chosen as a basic platform for information management in nanotechnologies because the authors had a positive experience with the previous version of the same software family – the documentary product CDS/ISIS by UNESCO. ISIS is still used by many of data centers and other documentary collections, mostly in developing countries, such as Latin America, Africa and South-East Asia. Familiar with its easy operation, the authors previously used it as a traditional DB for thermophysical properties of gases, liquids, metals, organic compounds and related materials (Trakhtengerts and Nun, 1989). It is reasonable that when working with more complex problems in using of nanomaterials with indistinct systematization, the authors took steps toward the implementation of the new software tool. At the beginning of the 2000s, the information service BIREME which has its main center in University São Paulo, Brazil, joined UNESCO's activities in information technologies and the development of the next generation of ISIS family software – the ABCD project (Dhamdhere, 2011; Trakhtengerts, 2009).

The new participants of the UNESCO project kept the attractive features of the basic ISIS software family, namely, the possibility for unlimited sets of local DBs, clearness of structures in document records, variety of methods to retrieve information and powerful instruments to format search results for printing. Users could access the Internet via the ABCD Z-39 protocol. The software includes all major library functions, including acquisitions, bibliographic DB management, user management, loan management, serial control and end-user searching on local and external bibliographic resources and library portals. ABCD also provides solutions for document centers as well. The centers usually have different services, as they have special collections, such as the necessity to reveal content (by providing abstracts, using thesauri, etc.), and require more flexibility in bibliographic structures, maintaining factual data, including numerical tables, images and multimedia. In fact, the transition from the old ISIS technology to ABCD means the transition to a new software capable of generating and managing DLs for subject domains with complex structured data. Dhamdhere (2011) noted some features of ABCD that allow considering the system as a DL tool. Among these are:

- unrestricted use of DB structures, which means that each record can apply its special structure;
- full-text indexing, "inverted file" – capabilities to extract individual words from the fields of records in the DB and to index them for search purposes;
- some PHP tools; for example, a PHP library that offers a full HTML editor, which can be embedded into a cataloguing form to create full documents;
- extracting texts from a Word document for inclusion into a field of an ABCD record;
- capability of using the Word indexing technique for retrieval; and
- access with the same interface to electronic documents located on the library server or elsewhere.

Flexibility of the DB structure is of special value when complex organized subject domains (biomedicine, material science and related areas) are characterized by semi-structured or non-structured data. Additional aspects which distinguish DL tools from old ISIS products are support of qualifiers, bibliographic formats and combination bibliographic and factual information. It is important for a developer of an information system that ABCD is capable of organizing DLs with the integration of reference catalogues, full-text collections, factual data as numerical tables, still images and more. The authors note several steps forward with a comparison between old ISIS versions and some of the technologies applied in ABCD. The most important for the creation of DLs are:

- Now the size of one document can be 1 Mb and the whole resource can be up to 4 Gb. It is expected that these restrictions in forthcoming versions will be expanded.

- Text fields in a document can be presented in HTML format that significantly improves its view. The mode "search through full text" is still supported.
- Fields may contain links to sources on the Internet and they may be used from within search results. Thus, chains of related documents may be combined. The use of Unicode in following versions is also planned.
- The 30-character indexing keys have been changed to 60-character keys, which provide improved control over sorting procedures, including those based on thesauri tables or DBs of authors.

Web-based technology allows HTML-encoded text strings for full-text indexing, hyperlinks to help pages and other important functions. By using this central part, it is possible to automate a small library with mostly narrow users. However, the most powerful functions of ABCD are tools which permit creating new DBs and modifying DB structures. Because ISIS DBs do not require sophisticated normalized relational structures and still can cope with elements in "many-to-many relations" (such as authors and their publications), ABCD can relatively easily cope with any locally created DB.

The structure and functionality of created DLs are defined by systems of fields, each of which is assigned with a certain data type and data format ("input type"). There are ten types in the total list with the main ones being *Field*, *Subfields* and *Group*. *Field* is the basic (most used) unit within a record. *Group* is a set of repeated fields of the same kind. *Subfields* are inner elements of a field. For example, *Group* for the field "Author" may include subfields "Last Name" and "First Name". It is convenient to use input type *Table* for the representation of a structure (see below for more detail about "input type"). Besides these three, there are special types of data (*Line* and *Heading*) for grouping and placement of fields on a worksheet for data entry: *Date (MARC 005)* – the special date field with tag 005 that is used in MARC, *Operator* and *Date* – records about the person responsible for data input and editing.

The second attribute of the field ("input type") provides a variety of record forms and gives the essential expansion of opportunities wanted. Along with standard text (*Text/Textarea*), the format under the name *HTML area* is also provided. This allows font management to enter graphic elements, hyperlinks and so forth. Additionally, two formats are aimed exclusively toward communication with external resources – *External HTML* and *Upload file*. Input type referred to as *HTML area* allows entering complicated blocks of text information with tables, drawings, hyperlinks and so forth. As this takes place, there are no a priori requirements to the logical data structure. This is the main condition imposed on semi-structured data (Abiteboul *et al.*, 2000). Aforesaid formats for linking with external resources permit the realization of the full potential of DLs: bibliographic records in conjunction with full-text sources, engaging external content and miscellaneous media presentations (text, image, video, audio, etc.).

The two other formats – *Select simple* and *Select multiple* – realize qualifier functions with permission to allocate one or several elements from the preliminary made list. They fix the names of defining signs, avoiding ambiguity and arbitrariness when free terms are assigned during the first step of data input and/or search. Typical spheres of application of these formats are classification of diseases and drugs in medical systems, substances and materials in physical and chemical data systematization and nanomaterials used in nanotechnologies. In doing so, a "semistructured (dynamic) classifier" (Builova *et al.*, 2013) was offered to support the evolving structure of data.

This is a subject catalog with a relatively stable framework and a limited number of hierarchy levels providing the possibility for the experts, when necessary, to add the appropriate levels refining individual headings. Dynamic versions of subject catalogs can be created this way, which allows categorization of individual segments of nanotechnologies with the required degree of depth. The ABCD tools allow for the easy realization of the specified requirements to be met in complexly organized and dynamically developing subject domains.

ABCD supports more formats useful in the development of information resources. For example, input type *Table* allows inserting of a compound data block into a field like *Group* with *Subfields*. There are some formats for a standard interface (*Checkbox* or *Radio*) to select an option, for example, to order documents in a certain way. The general list of input types totals 15 and includes types for office information (date, password, hidden text invisible to the user, etc.). These features of ABCD present considerable opportunities for developing a DL. Also, such opportunities can be realized with the use of very limited computer resources.

Results and discussion

Defined by size, nanotechnology encompasses nearly all branches of modern industry from space equipment up to health care. Such a range, together with the regular emergence of new devices and technologies, presents a considerable challenge to the IT designer (Erkimbaev *et al.*, 2008; Eletskii *et al.*, 2012a, 2012b). The wide variety of subjects stemming from rapid progress results in considerable difficulties in the management of information resources.

First of all, an accurate criterion is necessary for the allocation of a certain segment of subject domain: class of nanomaterials, technologies and devices and types of sources. All systems of general purpose, similar to the well-known Web of Science or Chemical Abstract Service, are focused on the last criterion and represent a flow of scientific publications, mainly based on the general content of summaries, keywords or subject headings.

Profound analysis of a structure is possible only for a narrow-domain segment. An example of this kind of resource is the DB on carbon nanomaterials supported by Ioffe Physico-Technical Institute (www.ioffe.ru/db_vul/).

The creation of a DL "Nanomaterials for Power Engineering" described in this work is aimed at providing adequate information for this dynamically developing area. Its function is to systematize the data covering both fields – nanomaterials (structures, physical properties and synthesis methods) and their technical applications (power technologies, the equipment and so forth). The possibilities of ABCD are appropriate to combine a rather detailed system of headings with data blocks of arbitrary size and structure.

The logical data structure is organized in a system of 25 fields, as shown in Table I. They store all necessary elements of the document, defining source (bibliographic information), type of a nanomaterial and power technology or equipment. The fields include text, tabular and graphic arrays, with in-depth information on material properties, synthesis method and domain of its applicability.

The first block of data (Fields 1-3 and 14-20) refers to the source description: author, title and similar. An important search option is the opportunity to set up the record and its type. Field 2 specifies the record type (bibl; full text, data), allowing the separation of

No.	Title	Searching field	Input type/format
1	Record index	Yes	Text/Textarea
2	Record type	Yes	Select simple
3	Document type	Yes	Select simple
4	Energy sector	Yes	Select multiple
5	Energy function	Yes	Select multiple
6	Object	Yes	Text/Textarea
7	Nanomaterial [free title]	Yes	Text/Textarea
8	Nanomaterial by rubricator	Yes	Select multiple
9	Chemical [free title]	Yes	Text/Textarea
10	Chemical by rubricator	Yes	Select multiple
11	*Synthesis*		*HTML area*
12	*Properties*		*HTML area*
13	*Application*		*HTML area*
14	Authors	Yes	Text/Textarea
15	Title rus	Yes	Text/Textarea
16	Title orig	Yes	Text/Textarea
17	Source		Text/Textarea
18	Year	Yes	Data
19	Language	Yes	Select simple
20	Affilation	Yes	Text/Textarea
21	*Full text*		*Upload file*
22	*WEB source*		*External HTML*
23	*More information*		*HTML area*
24	*Abstract*		*HTML area*
25	Comments		Hidden

Source: www.nsf.gov/pubs/1998/nsf9863/nsf9863.htm (accessed 5 August 2014)

Table I.
Fields of the DL "Nanomaterials for power engineering"

library segments relating to bibliographic, full text and factual information. Field 3 extracts document type from the preliminary prepared list (article, book, network document, etc.).

The second block of data includes necessary information about a nanomaterial. The block includes two formal characteristics: material type (Field 8) and chemical identity (Field 10). Data entry is carried out by the format *Select multiple*, that is the selection of more than one element from a list of predefined options. Fields 7 and 9 are directed to use any additional terms (free titles), making clear the structure and composition of the nanomaterial. Four fields (7-10) provide nanomaterial identification. At the same time, records may contain non-structured information on physical and performance properties of a material – Fields 11 and 12 filled by text, tables, drawings, hyperlinks and similar. Storage of complicated data is provided through the utilization of the format *HTML area*.

The third block (Fields 4-6 and 13) provides all necessary information on application. Fields 4 and 5, which use qualifiers, show related branches of power engineering and the function of corresponding devices or technologies. Free names may be used for the specification of these concepts in Field 6 (object) in the same way as was applied in the second block. More detailed description based on non-structured data (or SSD) can be included in Field 13 (*Application*) when using the format *HTML area*.

A key element in the proposed categorization is the classification scheme of nanomaterials focused on geometry, physical parameters, chemical nature, structure and so forth (Table II). Authors have analyzed a majority of rubricators applied in recent years (Builova *et al.*, 2013) and proposed an advanced schema based on the principle which was earlier formulated in Eletskii *et al.* (2012a, 2012b).

Table II. Nanomaterials classification

1.0	*Nanostructures*
1.1	0D nanoclusters
1.2	0D nanocryatals
1.3	OD fullerenes, endofullerenes, related materials
1.4	0D quantum dots
1.5	1D nanowires, nanorods, nanofibers
1.6	1D nanotubes
1.7	2D nanostructured and nanocomposite films
1.8	2D nanoporous surfaces
1.9	2D nanomembranes
1.10	2D graphene and graphene based materials
1.11	3D nanostructured materials
1.12	3D nanocomposite materials
1.13	3D nanoporous materials
1.14	3D nanopowders
1.15	3D nanofluids
2.0	*MISCL*
2.1	Ordered assemblies (multiple-layer and multiband nanostructures)
2.7	Nanoscale heterojunctions
3.0	*Functional nanomaterials*
3.1	Supported catalysts
3.2	Intercalated compounds and solid electrolytes
3.3	Sensor nanocomposites
3.4	Hydrogen absorbed materials
3.5	Nanostructured metals with special mechanical properties
3.6	Nanostructured ceramic and composite materials and coats
3.7	Nanostructured polymers, fibers and its composites
3.8	Layered magnetic materials and superlattices
3.9	Piezoelectrics
3.10	Superconducting materials
3.11	Thermoelectrics
3.12	Luminescent materials
3.13	Bionanomaterials
4.0	*Structural materials*
4.1	Technical-grade iron and carbon steels
4.2	Alloy steels
4.3	Non-ferrous metals and its alloys
4.4	Refractory materials and its alloys
4.5	Non-metallic materials

Source: https://sites.google.com/site/abcdlibraryautomationsoftware/ (accessed 5 August 2014)

The main feature of the system is a subject priority with a "stable framework" (Table II). Its variations concern only some elements. The head level of hierarchy is formed by four branches – nanostructures, MISCL (the mixed types), functional nanomaterials and structural nanomaterials. The first branch covers four classes of structures which are defined topologically (0D, 1D, 2D and 3D). The chosen number of the most typical representatives in each class is shown in the table. The headings named "functional nanomaterials" and "structural nanomaterials" allow for specifying a nanomaterial class due to its physical characteristics and applications. The scheme does not mean that all classes are comprehensively covered because some sophisticated structures sometimes appear. Thus, adoption of a stable classification is impossible. The graphene-nanotube structures (Du *et al.*, 2011) are examples of a case when classification is unambiguously difficult.

Capabilities of a DL may be considerably higher when some fields are reserved for hyperlinks to external resources. According to Table I, three fields (21-23) provide such features. The first of them includes a hyperlink on the full text of the document (article, the report, etc.), placed on the local server. The same document but localized on a remote server is called from Field 22. Wide lists of comments and hyperlinks to reference data or help files located on a server or on the Web can be inserted into Field 23, which uses the *HTML area* format.

In view of the quick progress of nanoscience, the library content has a higher value when reference data on material properties, synthesis and application are available. As an example of linked Web resources, a glossary of nanotechnology and related terms (http://thesaurus.rusnano.com), the similar dictionary (http://understandingNano.com) or the encyclopedia (www.nanowerk.com/n_encyclopaedia.php) may be used. The vast collection of help files which includes numerous reviews on properties of the most widespread nanomaterials that were published in recent years is available to users.

To sum up, ABCD tools are effective for creating and supporting information infrastructure corresponding to the requirements of a versatile and dynamically developing area, such as nanotechnology. The tools successfully combine the possibilities of detailed data categorization and storage of semi-structured information in this domain – properties, synthesis and applications of nanomaterials and use of external resources via hyperlinks.

References

Abiteboul, S., Buneman, P. and Suciu, S. (2000), *Data on the Web: From Relations to Semistructured Data and XML*, Morgan Kaufman Publishers, Burlington, MA.

Bizer, C. (2013), "Interlinking scientific data on a global scale", *Data Science Journal*, Vol. 12, pp. GRDI6-GRDI12.

Builova, N.M., Yeletskii, A.V., Zitserman, V.Y. and Kobzev, G.A. (2013), "The systematization of sources and data on nanotechnologies", *Scientific and Technical Information Processing*, Vol. 40 No. 4, pp. 212-223.

De Smet, E. (2009), "ABCD: a new FOSS library automation solution based on ISIS", *Information Development*, Vol. 25 No. 1, pp. 61-67.

Dhamdhere, S.N. (2011), "ABCD, an open source software for modern libraries", *Chinese Librarianship: An International Electronic Journal*, Vol. 32, available at: www.iclc.us/cliej/cl32dhamdhere.pdf (accessed 5 August 2014).

Du, F., Yu, D., Dai, L., Ganguli, S., Varshney, V. and Roy, A.K. (2011), "Preparation of tunable 3D pillared carbon nanotube-graphene networks for high-performance capacitance", *Chemistry of Materials*, Vol. 23 No. 21, pp. 4810-4816.

Eletskii, A.V., Erkimbaev, A.O., Kobzev, G.A., Trachtengerts, M.S. and Zitserman, V.Y. (2012a), "Properties of nanostructures: data acquisition, categorization, and evaluation", *Data Science Journal*, Vol. 11, pp. 126-139.

Eletskii, A.V., Erkimbaev, A.O., Zitserman, V.Y., Kobzev, G.A. and Trakhtengerts, M.S. (2012b), "Thermophysical properties of nanoobjects: data classification and validity evaluation", *High Temperature*, Vol. 50 No. 4, pp. 488-495.

Erkimbaev, A.O., Zitserman, V.Y., Kobzev, G.A. and Fokin, L.R. (2008), "The logical structure of physicochemical data: problems of numerical data standardization and exchange", *Russian Journal of Physical Chemistry A*, Vol. 82 No. 1, pp. 15-25.

Erkimbaev, A.O., Zitserman, V.Y., Kobzev, G.A., Serebrjakov, V.A. and Teymurazov, K.B. (2013), "Publishing scientific data as linked open data", *Scientific and Technical Information Processing*, Vol. 40 No. 4, pp. 253-263.

Fox, P. and Harris, R. (2013), "ICSU and the challenges of data and information management for international science", *Data Science Journal*, Vol. 12, pp. WDS1-WDS12.

Griffin, S., Peters, C. and Thanos, C. (2005), "Towards the new-generation digital libraries: recommendations of the NSF/EU-DELOS working groups", *International Journal of Digital Libraries*, Vol. 5 No. 4, pp. 253-254.

ISO/TR 11360 (2010), (2012), *Nanotechologies – Methodology for the Classification and Categorization of Nanomaterials*, ISO Technical Committee ISO/TC 229, Nanotechologies, available at: www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=55967 (accessed 5 August 2014).

Munro, R.G. (2003), *Data Evaluation, Theory and Practice for Materials Properties*, Special Publication 960-11, Materials Science and Engineering Laboratory, NIST, Washington, DC.

Murray-Rust, P. and Rzepa, H.S. (2002a), "STMML: a markup language for scientific, technical and medical publishing", *Data Science Journal*, Vol. 1 No. 2, pp. 128-192.

Murray-Rust, P. and Rzepa, H.S. (2002b), "Markup languages – how to structure chemistry-related documents", *Chemistry International*, Vol. 24 No. 4, available at: www.iupac.org/publications/ci/2002/2404/XML_markup.html (accessed 26 June 2015).

Rumble, J. and Freiman, S. (2012), "Describing nanomaterials: meeting the needs of diverse data communities", *Data Science Journal*, Vol. 11, pp. ASMD1-ASMD6.

Stones, R. and Matthew, N. (2005), *Beginning Databases with PostgreSQL: From Novice to Professional*, 2nd ed., Apress, New York, NY.

Trakhtengerts, M.S. (2009), "ABCD, an automated library system based on WinISIS", *Automatic Documentation and Mathematical Linguistics*, Vol. 43 No. 4, pp. 241-243.

Trakhtengerts, M.S. (2011), "The use of large-scale information services in gathering a collection of subject databases", *Scientific and Technical Information Processing*, Vol. 38 No. 3, pp. 186-188.

Trakhtengerts, M.S. and Nun, A. (1989), "Creation of thematic data-bases (TERMAL) by CDS/ISIS/EC", *Nauchno-tekhniceskaya informatsiya. Seriya 2 – Informatsionnye protsessy i systemy*, No. 11, pp. 16-18.

About the authors

Michael Trachtengerts, PhD in Thermophysics, now works on database development. Trachtengerts' research interests include electronic libraries and database management. Michael Trachtengerts is the corresponding author and can be contacted at: trachtengerts@yahoo.com

Adilbek Erkimbaev, PhD in Computer Science, now works on database design. Erkimbaev's research interests include database integration and Semantic Web technology.

Vladimir Zitserman, PhD in Physics, now works on categorization of nanomaterials. Zitserman's research interests include Semantic Web technology.

Georgii Kobzev, DSc in Physics, now works on theory of nanostructed materials. Kobzev's research interests include science informatics.