ISSN 0018-151X, High Temperature, 2016, Vol. 54, No. 2, pp. 215–222. © Pleiades Publishing, Ltd., 2016.
Original Russian Text © N.N. Kiselyova, V.A. Dudarev, A.V. Stolyarenko, 2016, published in Teplofizika Vysokikh Temperatur, 2016, Vol. 54, No. 2, pp. 228–236.

## THERMOPHYSICAL PROPERTIES OF MATERIALS

# Integrated System of Databases on the Properties of Inorganic Substances and Materials*

**N. N. Kiselyova, V. A. Dudarev, and A. V. Stolyarenko**

*Baikov Institute of Metallurgy and Materials Science, Russian Academy of Sciences, Moscow, 119991 Russia*

*e-mail: kis@imet.ac.ru*

Received January 26, 2015

**Abstract**—An integrated system of databases on the properties of inorganic substances and materials has been developed, currently combining the database of the A.A. Baikov Institute of Metallurgy and Materials Science, Russian Academy of Sciences, and an AtomWork database on the properties of inorganic substances, developed at the National Institute for Materials Science (Japan). This system is developed as an information service for specialists and for computer-aided design of new inorganic compounds, which was supported by an elaborate information-analytical system. The results of application of this system for compound design that have not yet been synthesized are presented.

**DOI:** 10.1134/S0018151X16020085

## INTRODUCTION

The increasing number of databases (DBs) containing extensive and diverse data about the properties of inorganic substances and materials is caused by an increasing volume of information and the invention and widespread use of relatively inexpensive and compact computing devices, the Internet, and convenient and affordable database management systems (DBMSs). Computer databases largely resolve the issue of how to obtain the necessary information promptly. Over the past decade, the number of databases in the field of inorganic chemistry and materials science has increased more than threefold. Information systems containing data on thermal properties traditionally occupy the first place by number (see, for example, [1–10]). The widespread use of databases with information on crystallography and crystal chemistry contributed to their intensive development [11–13]. In recent years, there has been a steady growth in the number of databases that contain information about the mechanical properties (strength, fatigue, creep, etc.) of inorganic substances and materials [14, 15]. Details about the databases on the properties of inorganic substances and materials (DB PISM) are given in [16] and the Information Resources of Inorganic Chemistry (IRIC) reference information system [17, 18].

DB PISM are widely used in fundamental and applied researches and in industry, but none of the developed information systems can provide comprehensive information about the totality of the properties of a particular substance or material. Most experts are compelled to view dozens databases to find the necessary parameters of a specific substance. To provide relevant and rapid search of data for a particular substance from different information systems, it is proposed to use virtual integration of DB PISM. The term "virtual" means that the data are not collected in a single data repository; they are located in developing organizations. This approach offers the most effective solution to provide users with a complete set of data about a specific inorganic substance from different databases located in different organizations and countries, based on different hardware and software.

Besides the basic function of the databases, that is, as an information service for specialists, these systems offer a much wider field for operating data. In particular, databases with thermal and crystal structure data are widely used for calculations. Systems like F*A*C*T [2], MALT2 [5], MTDATA [6], ThermoCalc [7], etc., can serve as examples of the systems combining the databases and the programs for thermodynamic calculations. A characteristic feature of such systems is the presence of the correct analytical description of dependencies. To perform the calculations, the expert needs to place the necessary information retrieved from the database into one of the selected models.

However, most of the problems solved in the practice of chemistry and, especially, materials science (prediction of new materials with desired properties, decoding of spectral information, selection of substances as materials for a particular purpose, development of optimal methods for manufacturing materi-

* Based on the Proceedings of the XIV Russian Conference on the Thermophysical Properties of Substances (RCTP-14), Kazan, October, 2014.

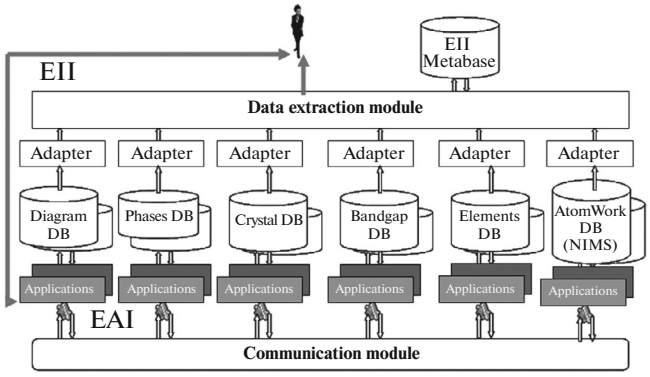


**Fig. 1.** Structure of integrated system of databases on the properties of inorganic substances and materials of the Institute of Metallurgy and Materials Science, Russian Academy of Sciences.

als, separation and identification of substances, etc.) are difficult to formalize at the level of those relatively simple algebraic structures that are used, for example, in thermodynamics. Currently, all these problems are often solved using empirical knowledge. To solve such problems that are difficult to formalize, information-analytical systems (IASs) have been developed that combine DB PISM and subsystems for analyzing information.

## PRINCIPLES OF INTEGRATION OF DB PISM

Basically, there are two possible approaches to database integration. In the first, information from different databases after the unification and elimination of obvious errors is loaded into a megadatabase, Data Warehouse. This approach is the basis for the methodology of ETL integration (Extract, Transform, Load), supported by software for extracting and transforming data from integrated databases, followed by loading the data into a repository [19]. However, the specific subject domain makes the full merging of information from the DB PISM and the creation of a central repository a highly difficult technical and organizational task requiring huge financial investments.

Regarding this, the second approach—a virtual integration of the DB PISM and the creation of a heterogeneous distributed information system—is more effective and less expensive, ensuring the independence of development of individual databases and providing access to the consolidated data on a specific substance or material from different databases.

There are two basic technological techniques in the virtual integration of the databases: (1) Enterprise Information Integration (EII) [20] and (2) Enterprise Application Integration (EAI) [21].

In the integration of information (EII), a software access interface is developed, with which one can extract the necessary data from different databases. Some central information system (metabase) is created, which interacts with the distributed data sources, retrieves the information about the requested substance, aggregated from various DB PISM, and delivers it to the user.

Using the second approach (EAI) is most appropriate when the DB PISM include applications. In implementing this approach, only the user interfaces of the databases are consolidated, accessing the calculation subsystems, rather than the databases themselves. These interfaces can be web-application of corresponding information systems. It should be noted that, in this approach, it is appropriate to use "cloud" computing for large volumes of data and complex calculations.

An analysis of the capabilities of existing database integration technologies in terms of the information needs of professionals in the field of inorganic chemistry and materials science has shown that none of the above approaches alone is able to solve all the problems arising from the integration of information systems and software applications. In this regard, in the creation of an integrated system of DB PISM, a complex approach was suggested that combines the integration at the levels of both data and user interfaces (EII + EAI) [22, 23]. It enabled, on the one hand, the integration of the user interfaces of web-applications of DB PISM and the data analysis applications (EAI) and, on the other hand, the consolidation of information extracted from the various distributed data sources (EII).

## INTEGRATED DB SYSTEM PISM OF THE INSTITUTE OF METALLURGY AND MATERIALS SCIENCE, USSIAN ACADEMY OF SCIENCES

The proposed approach to integration has been applied successfully to create an integrated system of the DB PISM, which currently brings together the information systems developed in the Institute of Metallurgy and Materials Science, Russian Academy of Sciences [16], on the phase diagrams of semiconductor systems (Diagram), the properties of the acousto-optical, electro-optical, and nonlinear optical substances (Crystal), the band gap of inorganic substances (Bandgap), the properties of inorganic compounds (Phases), and the properties of chemical elements (Elements), and the AtomWork database on the properties of inorganic substances, developed at the National Institute for Materials Science (NIMS, Japan) (Fig. 1).

The Phases database on the properties of inorganic compounds [16, 24] currently contains information on the properties of approximately 52000 ternary compounds and more than 31000 quaternary compounds, collected using more than 32000 publications. It includes brief information about the most common

properties of inorganic compounds: crystal chemical (the type of crystal structure with indication of the temperature and pressure above which this structure is implemented, the crystal system, the space group, the number of formula units in the unit cell, and the lattice parameters) and thermophysical (the melting type and temperature, the temperature of decomposition of the compound in solid or gaseous phases, and the boiling point at atmospheric pressure) data. In addition, the database contains information on the superconducting properties of compounds. The Phases database is based on the analysis of data extracted from periodicals, reference books, monographs, reports, and abstract journals; more than half of the sources are stored as pdf documents. This database is available on the Internet for registered users [25].

The Elements database [16] includes information about 90 of the most common properties of chemical elements: the thermal (the melting and boiling points at 1 atm and the standard values of thermal conductivity, molar heat capacity, enthalpy of atomization, entropy, etc.), size (the ionic, covalent, metal, and pseudopotential radii, the atomic volume, etc.), and other physical properties (the magnetic susceptibility, electrical conductivity, hardness, density, etc.); the position in the periodic table of elements; etc. The database is available on the Internet [26].

The Diagram database [16, 27, 28] contains dozens of phase $P,T,x$-diagrams of binary and ternary semiconductor systems and the physicochemical properties of phases formed in them, collected and evaluated by highly qualified experts. Figures with phase diagrams are stored as bitmap (format *.jpeg) and vector (format *.swf) images. It should be noted that the user can use only a standard software, which is a feature of all DB PISM developed at the Institute of Metallurgy and Material Science, Russian Academy of Sciences. In addition to the graphic information, computational models are stored in analytical reports, written in details, and special tables, obtained by thermodynamic self-consistency or statistical optimization of phase diagrams or their elements using various techniques. The user can work with them for his own calculations, by substituting the parameters included in research reports or stored in the appropriate tables of thermodynamic properties and descriptions of the simulation models of the database. In addition, the numerical data, evaluated by experts and self-consisted, for the elements of the phase diagrams are stored in the corresponding tables of the database. The Diagram database is available on the Internet for registered users [29].

The Bandgap database [30] includes information about the band gap of more than 3000 inorganic substances and is available on the Internet [31].

The Crystal database [16, 28] includes information about the piezoelectric (piezoelectric coefficients, elastic constants, etc.), nonlinear optical (nonlinear optical coefficients, the Miller tensor components, etc.), crystal chemical (the type of the crystal structure, crystal system, space and point groups, the number of formula units per unit cell, and the crystal lattice parameters), optical (refractive indices, the transparency band, etc.), and thermal (melting point, specific heat, thermal conductivity, etc.) properties of more than 140 acousto-optical, electro-optical, and nonlinear optical materials, collected and evaluated by highly qualified experts in the subject area. It has Russian and English versions available for registered users on the Internet [32].

The AtomWork Inorganic Material Database (NIMS, Japan) [33] contains information about more than 82000 crystal structures, 55000 values of the properties of materials, and 15000 phase diagrams; it is also available on the Internet [34].

A service-oriented architecture (SOA) was used to integrate the databases, based on the use of web-services for ensuring interoperability between heterogeneous information systems. To search for the relevant information in the context of information systems, a specially designed metabase is used [16], which describes the contents of the integrated databases in terms of a formalized hierarchy of concepts inherent to inorganic chemistry and materials science.

The integration of databases is a contemporary trend in the development of information systems in the fields of materials science. Integrated systems offer specialists the most reliable and complete data on the properties of substances and materials and deliver the aggregate information anywhere in the world through the Internet.

Further development of integrated DB PISM is directly related to the solution of the following types of problems:

—development of common standards and software for integrating DB PISM;

—further structuring of subject areas;

—creation of thesauri;

—evaluation of the reliability of data;

—arrangement of access to databases of different institutions and different countries.

## IAS FOR COMPUTER-AIDED DESIGN OF INORGANIC COMPOUNDS

One of the trends in the development of information systems is the incorporation of data analysis tools into their structure, from the simplest means of aggregation of information, processing of complex multicriteria queries, programs for statistical analysis and visualization of results to complex artificial intelligence systems. We developed a special IAS, which, in addition to the information service for professionals, is designed to search for regularities in big chemical data and computer design of inorganic compounds [16, 35,

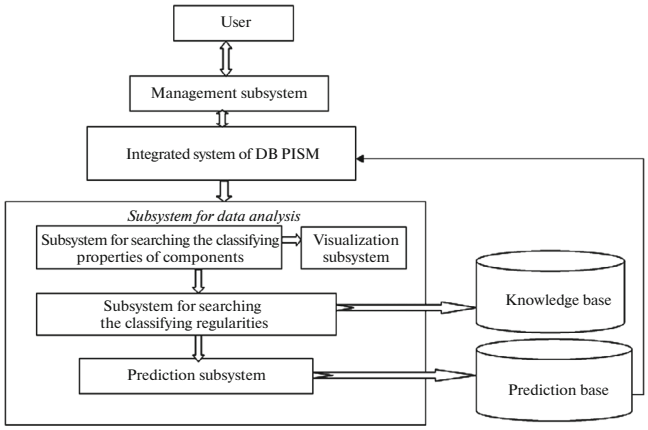


**Fig. 2.** Schematic diagram of the information-analytical system for the design of inorganic compounds.

36]. It includes (Fig. 2), along with an integrated DB PISM, a subsystem of information analysis and predictions, bringing together a set of programs of precedent-based pattern recognition methods, a base of found regularities (the knowledge base), a base of predictions of the possibility of forming and properties of inorganic compounds that have not been yet synthesized, and a management subsystem.

### *Data analysis subsystem*

**Subsystems for searching for classifying regularities and predictions.** In the development of this subsystem, the most important task was the selection of the most appropriate mathematical methods for searching regularities in chemical data. Typically, this task is performed by the trial-and-error method. In the selection of pattern recognition methods for analysis of chemical information, many years' experience in the application of these methods to design of inorganic compounds was taken into account [37]. The following methods and programs have been selected:

—a wide range of algorithms of the Recognition multifunctional system, developed at the Computing Center of the Russian Academy of Sciences [38] and bringing together, in addition to well-known techniques, the algorithms of pattern-recognition (based on calculation of estimates), voting algorithms based on deadlock tests, voting algorithms based on logical regularities, weighted statistical voting algorithms, etc.;

—a ConFor computer system for training a computer in the procedure for concept formation [39], which is based on an original organization of data in the computer memory in the form of growing pyramidal networks.

These programs of data analysis were selected for the following reasons.

(1) The versatility with respect to the problem solved. The above systems permit solution of both the problems of prediction of rare or unique events, phenomena, or processes, when the initial (training) information is small (tens of precedents), and large-scale problems (tens of thousands of precedents).

(2) The flexibility with respect to the type of data (numeric, binary, and nominal characteristics are accepted).

(3) The ability to handle incomplete and partially contradictory information, when the values of some attributes are unknown or known only approximately.

As a rule, it is not possible to specify in advance which algorithm is the most efficient for solving a particular problem. Therefore, it seems promising to apply the methods of recognition by ensembles of algorithms. In the creation of a collective decision, the possible errors of recognition with individual algorithms can be compensated for in many cases by correct results of other algorithms. Based on this, we included programs that implement different strategies for collective decision-making, for example, the Bayesian method; methods using clustering and selection, decision templates, logical correction, the method of a convex stabilizer, the Woods dynamic method, committee methods, etc., in the developed IAS [38].

The above algorithms of pattern recognition and collective decision-making are based on different principles, and they use different forms of representation of the desired regularities (algebraic functions, logical expressions, neural or growing pyramidal networks, etc.). To integrate such diverse software, we used an SOA, which takes into account the differences in the data and information structures used in integrated programs, as well as sophisticated mechanisms of their interaction. It provided the opportunity for a rather simple addition of new software for data analysis into the regularities search subsystem. In the integration of applications, we used an integrating environment, which plays the role of a universal software core that connects all applications, instead of creating specific interfaces between the individual programs [40]. The advantage of this technology using an integrating environment is primarily the simple support and extension of the system developed.

The initial information for computer analysis is extracted from the integrated system of DB PISM and has the form of a table in the Excel format. Each row in the table contains a set of values of the properties of components (elements or simpler compounds) of already known compounds indicating the class to which it belongs (for example, the type of crystal structure under normal conditions, or whether its temperature of transition into a superconducting state is above or below 4.2 K).

In predicting inorganic compounds that have not been synthesized and in assessing their properties,

only information about the properties of the components is used. The prediction procedure is fully automated. The user specifies only the designation of components, and the prediction subsystem itself generates a sample for recognition, places the obtained values of the properties of components in a classifying regularity, and calculates the prediction result.

**Subsystem for searching the classifying properties of components.** For the selection of informative properties of the component of chemical compounds, we included programs based on algorithms [41–43] in the IAS. The selection of the properties of the components, the most informative for the classification of substances, has a double meaning. On the one hand, it drastically decreases the volume of the information analyzed, which for multicomponent substances comprise hundreds of properties of elements and simpler compounds, as well as functions of these properties. On the other hand, the selection of properties of the components most important for the classification of chemical substances, enables the physical interpretation of the classifying regularities, which enhances the credibility of the predictions obtained and finding substantial causal links between the parameters of the objects and the development of the physical and chemical models of phenomena.

**Visualization subsystem.** The imaging subsystem facilitates the interpretation of the results, which constructs the projections of the points corresponding to the compounds in two-dimensional spaces of the properties of components, including not only the initial parameters but also user-specified algebraic functions of these parameters.

**Knowledge base**. The knowledge base contains the obtained classifying regularities. In its software implementation, there was a problem related to the fact that the presentation of knowledge in the computer training methods differs significantly. In this regard, a new software solution was proposed for storing the obtained regularities and related information about the application settings and objects under study [40]. The storage of this data is implemented by means of an SQL Server and file structures on the server disks. The server stores the obtained regularity in a special internal format of the data analysis programs, and operation information about these regularity, namely, the unique identifier of the regularity, the designation of the predicted characteristics, the formulaic composition of chemical compounds, the designation of the properties of the components used to describe the substances, the path to files on the disk, the name of the expert who carried out the evaluation of the data for training and search for regularities, the date of formation of the regularity, etc., is stored in the tables of the database in the SQL Server.

**Prediction base.** The database contains the results of previous computer experiments, as well as links to operation information stored in the knowledge base. Using the prediction base helped to improve the functionality of the database on the properties of inorganic substances and materials, developed at the Institute of Metallurgy and Materials Science, Russian Academy of Sciences, by providing the user with not only known data about already studied substances but also predictions for inorganic compounds not yet synthesized and evaluations of their properties. Now, this database is filled.

**Management subsystem.** The management subsystem organizes the computing process, ensures interaction between the functional subsystems of the IAS, and provides access to the system on the Internet. In addition, the management subsystem provides the user with software for data preparation for analysis, outputting reports, and implementation of other service functions. In particular, we developed a special subsystem to retrieve information from the database, which, after evaluation of the expert, is used to train the computer, and to prepare it for further analysis. It gives the expert the possibility to edit the found information and to form a sampling for analysis. In the latter case, the expert marks only the selected properties of the components in a special table (menu), and the subsystem for the sample preparation for analysis retrieves the selected property values from the Elements database. If needed, the algebraic functions of the initial properties are formed in the subsystem and the description of the compounds is assembled in the form of an Excel table, which is then input to the prediction subsystem. The subsystem of result delivery is intended to make predictions in a tabular form conventional to chemists and materials scientists.

An important feature of the software implementation of IAS is that the client side is fully built using a web-based interface [40]. Users work with the IAS on a web browser. The processes of training and recognition in the IAS are implemented using a special asynchronous web-service, which enables us to solve time-consuming tasks of training and prediction in the Internet environment, which may not function properly. The asynchronous web-service allows users to initiate the implementation of time-intensive operations, to control the degree of their implementation in an asynchronous mode, to receive notification of the results of calculations, to interrupt the execution of tasks, and to store intermediate results.

## USE OF THE IAS FOR PREDICTING NEW COMPOUNDS AND EVALUATION OF THEIR PROPERTIES

The results obtained using the created IAS are illustrated in the table; they give the predictions of the possibility of formation of $AB_3X_3$ compounds in the $A_2X_3$–$B_2X$ systems (A and B are different elements, and X = S, Se, or Te) under normal conditions [44], which could be promising for finding new semicon-

Prediction of the possibility of formation of $AB_3X_3$ compounds

X≡S

| A / B | B | Al | P | Sc | Ti | Cr | Fe | Ga | As | Y | Rh | In | Sn | Sb | La | Ce | Pr | Nd | Pm | Sm | Eu | Gd | Tb | Dy | Ho | Er | Tm | Yb | Lu | Bi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Li | #1 | | 1 | 1 | #1 | 1 | 1 | #2 | #1 | | | #2 | | #1 | 2 | 2 | 2 | 2 | | | | | | | | | | | | #2 |
| Na | #1 | #1 | 1 | 1 | 1 | 1 | #1 | #2 | #1 | 1 | #1 | #1 | 1 | #1 | | | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | #1 |
| K | #1 | 1 | 1 | 1 | 1 | 1 | © | #2 | #1 | 1 | #1 | 1 | 1 | #1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | #2 |
| Cu | 1 | #2 | #1 | #1 | | #2 | #1 | | #1 | #1 | 1 | #2 | #1 | #1 | #2 | #2 | 2 | #2 | | | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 |
| Rb | #1 | 1 | 1 | 1 | 1 | 1 | © | | 1 | 1 | 1 | #1 | 1 | © | 1 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | #1 |
| Ag | | #2 | | | | | #2 | #2 | #1 | | | #2 | #2 | #1 | #2 | #2 | #2 | #2 | 2 | 2 | 2 | | 2 | #2 | 2 | | | | | #2 |
| Cs | #1 | 1 | 1 | 1 | 1 | 1 | #1 | #2 | | | 1 | 1 | 1 | #2 | 1 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| Tl | #1 | | #2 | 1 | 1 | #1 | 1 | Θ | #1 | | 1 | #1 | #1 | #1 | #2 | 2 | #2 | 2 | 2 | 2 | 2 | | | | | | | 1 | 1 | |

X≡Se

| A / B | B | Al | P | Sc | Ti | Cr | Fe | Ga | As | Y | Rh | In | Sn | Sb | La | Ce | Pr | Nd | Pm | Sm | Eu | Gd | Tb | Dy | Ho | Er | Tm | Yb | Lu | Bi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Li | 1 | | 1 | 1 | 1 | 1 | 1 | 2 | #2 | 1 | 1 | 2 | | | 2 | 2 | 2 | 2 | | | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| Na | 1 | #1 | 1 | 1 | 1 | 1 | #1 | #1 | #1 | 1 | 1 | 1 | 1 | #2 | | | | | | | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| K | 1 | #1 | 1 | 1 | 1 | 1 | #1 | #1 | #1 | 1 | 1 | 1 | © | #1 | | 1 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | #1 |
| Cu | 1 | | 1 | #1 | 1 | 1 | 1 | #2 | #1 | #1 | 1 | #2 | #1 | #1 | 2 | 2 | 2 | 2 | | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 |
| Rb | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | © | 1 | 1 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | #1 |
| Ag | 1 | | 2 | | | 2 | 2 | #2 | #1 | 2 | | Θ | | #2 | Θ | Θ | Θ | 2 | 2 | Θ | Θ | | Θ | 2 | 2 | 2 | 2 | 2 | 2 | Θ |
| Cs | 1 | 1 | 1 | 1 | 1 | 1 | 1 | #1 | 1 | 1 | 1 | 1 | 1 | © | 1 | | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | #1 |
| Tl | #1 | | 1 | 1 | 1 | #1 | 1 | Θ | #1 | 1 | 1 | #2 | #1 | #1 | 2 | 2 | 2 | 2 | | | | | | | 1 | 1 | 1 | 1 | 1 | #2 |

X≡Te

| A / B | B | Al | P | Sc | Ti | Cr | Fe | Ga | As | Y | Rh | In | Sn | Sb | La | Ce | Pr | Nd | Pm | Sm | Eu | Gd | Tb | Dy | Ho | Er | Tm | Yb | Lu | Bi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Li | 1 | | | 1 | 1 | | 1 | 2 | | 1 | 1 | | | | 2 | 2 | 2 | 2 | 2 | | | | | | 1 | 1 | 1 | 1 | 1 | |
| Na | 1 | #1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | #1 | | | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 |
| K | 1 | #1 | 1 | 1 | 1 | 1 | 1 | #1 | 1 | 1 | 1 | 1 | #1 | #1 | 1 | | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | #1 |
| Cu | 1 | #2 | | 1 | 1 | 1 | 1 | #2 | #2 | #1 | 1 | #2 | | #2 | 2 | 2 | 2 | 2 | 2 | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #1 | #2 |
| Rb | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | © | 1 | 1 | 1 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Ag | 1 | | 2 | 2 | | 2 | 2 | #2 | 2 | 2 | | #2 | Θ | #2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | #2 | Θ | Θ | 2 | 2 | 2 | | | Θ |
| Cs | 1 | #1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | © | 1 | 1 | 1 | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Tl | 1 | | | 1 | 1 | | 1 | Θ | | | 1 | Θ | #1 | #2 | #2 | 2 | 2 | Θ | 2 | | | | | | | | | | | #2 |

ductor, nonlinear optical, electro-optical, and acousto-optical materials.

Experimental information on 117 examples of $AB_3X_3$ compounds formed and 58 examples when no compounds of such a composition were formed in the $A_2X_3$–$B_2X$ systems under normal conditions was used for computer analysis. To describe the compounds in the computer memory based on the physical and chemical concepts of the nature of substances of this type, we selected the properties of elements A, B, and X (the melting and boiling points; covalent, ionic (by Bokii and Belov), and pseudopotential (by Zunger) radii; the first three ionization potentials; electronegativity (by Pauling); the standard enthalpies of atomization and evaporation; thermal conductivity; molar heat capacities, etc.), the properties of simple chalcogenides $A_2X_3$ and $B_2X$ (standard entropy and enthalpy), and some algebraic functions of these prop-

erties (for example, the ratio of the covalent radius to the metal radius for elements A, B, and X [44]).

The table presents examples of predictions for the $AB_3X_3$ compounds and the results of their experimental verification. The following designations are adapted: 1, the prediction of formation of $AB_3X_3$ under normal conditions; 2, the prediction of an absence of $AB_3X_3$ under normal conditions; #, examples, the information on which is used for training the computer; empty cells, uncertain prediction; ©, the prediction of formation of $AB_3X_3$ matches new experimental data; and Θ, the prediction of the absence of a compound matches experimental data. All 25 tested predictions coincided with the experimental data [45–48].

The use of pattern recognition methods ensures prediction with an average accuracy of 80%, for example, of the type of crystal structure under predetermined conditions [36, 37] and the evaluation of some fundamental properties of the compounds, for example, the melting point [49], the band gap [30, 50], or the critical temperature of transition into the superconducting state [51].

## CONCLUSIONS

The information-analytical system solves two important problems. Firstly, it partially automates the analysis of the enormous amount of experimental data accumulated in chemical practice. This makes it possible to find regularities in the data and use them to design new compounds with desired properties; only the parameters of components can be used in the prediction stage of phases that have not yet been synthesized. Secondly, the IAS enhances the possibilities of conventional databases on the properties of substances and materials by offering the user not only information about substances that have already been studied but also predictions for not yet studied compounds with an assessment of their properties. A significant advantage of the IAS developed is its availability on the Internet [52]. The IAS has helped to obtain predictions for thousands of inorganic compounds not yet synthesized and to evaluate some of their properties.

## ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, projects nos. 12-07-00142, 14-07-00819, 14-07-31032, and 15-07-00980.

## REFERENCES

1. Belov, G.V., Iorish, V.S., and Yungman, V.S., *CALPHAD: Comput. Coupling Phase Diagrams Thermochem.*, 1999, vol. 23, no. 2, p. 173.
2. Bale, C.W., Chartrand, P., Degterov, S.A., et al., Factsage thermochemical software and databases, *CALPHAD: Comput. Coupling Phase Diagrams Thermochem.*, 2002, vol. 26, no. 2, p. 189.
3. Yamashita, Y., Yagi, T., and Baba, T., *Jpn. J. Appl. Phys.*, 2011, vol. 50, no. 11, p. RH03-1.
4. Xu, Y., Yamazaki, M., Wang, H., and Yagi, K., *Mater. Trans.*, 2006, vol. 47, no. 8, p. 1882.
5. Yokokawa, H., Yamauchi, S., and Matsumoto, T., *CALPHAD: Comput. Coupling Phase Diagrams Thermochem.*, 2002, vol. 26, no. 2, p. 155.
6. Huang, Z., Conway, P.P., Thomson, R.C., et al., *CALPHAD: Comput. Coupling Phase Diagrams Thermochem.*, 2008, vol. 32, no. 1, p. 129.
7. Andersson, J.-O., Helander, T., Hoglund, L., et al., *CALPHAD: Comput. Coupling Phase Diagrams Thermochem.*, 2002, vol. 26, no. 2, p. 273.
8. Eletskii, A.V., Erkimbaev, A.O., Zitserman, V.Yu., Kobzev, G.A., and Trakhtengerts, M.S., *High Temp.*, 2012, vol. 50, no. 4, p. 488.
9. Fokin, L.R. and Kalashnikov, A.N., *High Temp.*, 2009, vol. 47, no. 5, p. 643.
10. Erkimbaev, A.O., Zitserman, V.Yu., and Kobzev, G.A., *High Temp.*, 2010, vol. 48, no. 6, p. 830.
11. Faber, J. and Fawcett, T., *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, vol. 58, no. 3, p. 325.
12. Hellenbrandt, M., *Crystallogr. Rev.*, 2004, vol. 10, no. 1, p. 17.
13. White, P.S. and Rodgers, J.R., *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, vol. 58, no. 3, p. 343.
14. Over, H.H., Wolfart, E., Dietz, W., and Toth, L., *Adv. Eng. Mater.*, 2005, vol. 7, no. 8, p. 766.
15. Li, X., Su, H., Chen, X., et al., *Data Sci. J.*, 2007, vol. 6 (Suppl.), p. S467.
16. Kiselyova, N.N., Dudarev, V.A., and Zemskov, V.S., *Russ. Chem. Rev.*, 2010, vol. 79, no. 2, p. 145.
17. Information Resources on Inorganic Chemistry. http://iric.imet-db.ru
18. Kiselyova, N.N. and Dudarev, V.A., *Inf. Tekhnol.*, 2010, no. 12, p. 63.
19. Kimball, R. and Caserta, J., *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*, New York: Wiley, 2004.
20. Morgenthal, J.P., *Enterprise Information Integration: A Pragmatic Approach*, Morrisville: Lulu. com, 2005, p. 317.
21. Morgenthal, J.P., *Enterprise Applications Integration with XML and JAVA*, New York: Prentice Hall, 2000.
22. Kornyushko, V. and Dudarev, V., *Int. J. Inf. Theor. Appl.*, 2006, vol. 13, no. 2, p. 119.
23. Kiselyova, N., Iwata, S., Dudarev, V., et al., *Int. J. Inf. Technol. Knowl.*, 2008, vol. 2, no. 4, p. 366.
24. Kiselyova, N., Murat, D., Stolyarenko, A., et al., *Inf. Resursy Rossii*, 2006, no. 4, p. 21.
25. http://www.phases.imet-db.ru
26. http://phases.imet-db.ru/elements
27. Khristoforov, Yu.I., Khorbenko, V.V., Kiselyova, N.N., et al., *Izv. Vyssh. Uchebn. Zaved., Mater. Elektron. Tekh.*, 2001, no. 4, p. 50.
28. Kiselyova, N.N., Prokoshev, I.V., Dudarev, V.A., et al., *Inorg. Mater.*, 2004, vol. 42, no. 3, p. 321.
29. Diagram Database. http://diag.imet-db.ru

30. Kiselyova, N.N., Dudarev, V.A., and Korzhuev, M.A., *Inorganic Materials: Applied Research*, 2015, vol. 7, no.1, p. 3.
31. www.bg.imet-db.ru
32. http://crystal.imet-db.ru
33. Xu, Y., Yamazaki, M., and Villars, P., *Jpn. J. Appl. Phys.*, 2011, vol. 50, no. 11, p. RH02-1.
34. http://mits.nims.go.jp
35. Kiselyova, N.N., Stolyarenko, A.V., Ryazanov, V.V., et al., *Pattern Recognit. Image Anal.*, 2011, vol. 21, no. 1, p. 88.
36. Kiselyova, N., Stolyarenko, A., Ryazanov, V., *et al.*, in *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems*, Naidenova, X.A. and Ignatov, D.I, Eds., Hershey: IGI Global, 2012, p. 197.
37. Kiselyova, N.N., *Komp'yuternoe konstruirovanie neorganicheskikh soedinenii. Ispol'zovanie baz dannykh i metodov iskusstvennogo intellekta* (Computer Design of Inorganic Compounds: Use of Databases and Artificial Intelligence Methods), Moscow: Nauka, 2005.
38. Zhuravlev, Yu.I., Ryazanov, V.V., and Sen'ko, O.V., "Raspoznavanie." Matematicheskie metody. Programmnaya sistema. Prakticheskie primeneniya ("Recognition," Mathematical Methods, Software System, and Practical Applications), Moscow: FAZIS, 2006.
39. Gladun, V.P., *Protsessy formirovaniya novykh znanii* (Processes of Formation of New Knowledge), Sofia: Pedagog-6, 1995.
40. Stolyarenko, A.V., Kiselyova, N.N., and Podbel'skii, V.V., *Avtom. Sovr. Tekhnol.*, 2008, no. 9, p. 23.
41. Senko, O.V., *Pattern Recognit. Image Anal.*, 2009, vol. 19, no. 3, p. 465.
42. Yuan, G.-X., Ho, C.-H., and Lin, C.-J., *J. Mach. Learn. Res.*, 2012, vol. 13, p. 1999.
43. Yang, Y. and Zou, H., *J. Stat. Comput. Simul.*, 2014, vol. 84, no. 1, p. 1.
44. Kiselyova, N.N., *Inorg. Mater.*, 2009, vol. 45, no. 10, p. 107.
45. Schindler, L.V., Schwarz, M., and Röhr, C., *Z. Naturforsch., B: J. Chem. Sci.*, 2013, vol. 68, no. 12, p. 1295.
46. Olekseyuk, I., Tsisar, O., Piskach, L., and Parasyuk, O., *Nauchn. Vestn. Sredneevr. Nats. Univ. im. Lesi Ukrainki, Ser.: Khim. Nauki*, 2014, no. 20, p. 26.
47. Mucha, I., Wiglusz, K., Sztuba, Z., and Gawel, W., *Thermochim. Acta*, 2011, vol. 518, nos. 1–2, p. 53.
48. Schwarz, M. and Röhr, C., *Z. Kristallogr.*, 2012, vol. 32 (Suppl.), p. 96.
49. Kiselyova, N.N., Podbel'skii, V.V., Ryazanov, V.V., and Stolyarenko, A.V., *Tr. XII Ross. konf. po teplofiz. svoistvam veshchestv* (Proc. XII Russian Conf on Thermophysical Properties of Substances), Moscow: Interkontakt, Nauka, 2009, p. 133.
50. Kiselyova, N.N., Stolyarenko, A.V., Gu, T., et al., *Proc. 6th Int. Conf. on Computer-Aided Design of Discrete Devices (CAD DD 07)*, Minsk: Ob"ed. Inst. Probl. Avtomat., Nats. Akad. Nauk Belarusi, 2007, vol. 1, p. 236.
51. Savitskii, E.M. and Kiselyova, N.N., *Dokl. Akad. Nauk SSSR*, 1978, vol. 239, no. 2, p. 405.
52. http://ias.imet-db.ru

*Translated by O. Zhukova*