

# Ontologies and databases on thermophysical properties of nuclear reactor materials\*

Igor A. Chusov<sup>1</sup>, Pavel L. Kirillov<sup>1</sup>, Vladimir G. Pronyaev<sup>2</sup>, Nikolay A. Obysov<sup>3</sup>, Grigoriy E. Novikov<sup>3</sup>

1 Institute for Nuclear Power Engineering, NRNU «MEPhI», 1 Studgorodok, Kaluga reg., Obninsk, 249040, Russian Federation

2 Private Institution «Atomstandart», 17 Butlerova st., Moscow, 117342, Russian Federation

3 ROSATOM State Atomic Energy Corporation, 24 Bolshaya Ordynka st., Moscow, 119017, Russian Federation

Corresponding author: Igor A. Chusov ([igrch@mail.ru](mailto:igrch@mail.ru))

---

Academic editor: Yury Korovin ♦ Received 25 March 2019 ♦ Accepted 21 May 2019 ♦ Published 21 June 2019

---

Citation: Chusov IA, Kirillov PL, Pronyaev VG, Obysov NA, Novikov GE (2019) Ontologies and databases on thermophysical properties of nuclear reactor materials. Nuclear Energy and Technology 5(2): 145–153. <https://doi.org/10.3897/nucet.5.36476>

---

## Abstract

The study is dedicated to the information technologies for storage, systematization and distribution of thermophysical data for nuclear power engineering. The general trend existing in the areas involving wide use of scientific data is the shifting from conventional databases to the development of a consolidated infrastructure capable of overcoming sharply growing volumes of scientific data with continuously increasing complexity of the data structure due to the expansion of the range of materials. The above infrastructure ensures interoperability, including data exchange and dissemination. The general principle of data management for thermophysical properties of the nuclear reactor materials based on the subject-oriented ReactorThermoOntology (RTO) is suggested in the present paper. The ontology includes a unified glossary of all concepts, expanded through logical connections and axioms. The suggested RTO ontology combines the terms typical for reactor materials, their characteristics, as well as all types of information entities determining textual, mathematical and computer structures. In the coded form, the ontology becomes the control add-in capable to integrate heterogeneous data. Its most important feature is the possibility of its permanent expansion, which is necessary with introduction of new materials and terms related to them, e.g. nanostructures characteristics. Beside the ontology, description of the reactor materials, the possible scenarios for the use of the ontology during the phases of design, operation and integration of autonomous resources, primarily databases, are examined in the paper. The use of Big Data technology with diverse variations of logical structures of the data is suggested as the most efficient tool for data integration. Joint use of the technologies which before were applied separately, such as exchange standard in the form of the structured text documents, data control based on the ontology and platform for the work with big data, allows the conversion of multiple existing primary resources (databases, files, archives, etc.) to the standard JSON text format for the subsequent semantic integration.

---

## Keywords

Thermophysical properties, reactor materials, nuclear fuel, ontology, database, data integration, JSON-format

---

\* Russian text published: Izvestiya vuzov. Yadernaya Energetika (ISSN 0204-3327), 2019, n. 1, pp. 5–16.

## Introduction

Development of nuclear power generation is inextricably associated with vast infrastructure ensuring data storage and preservation of knowledge of technologies, work processes, materials and other aspects. The core of this infrastructure is composed by databases (DB) and, what is more, increase of the volume of data related to the expanding scope of objects necessitates the involvement in the operational practice of multiple non-homogeneous resources with markedly different formats, semantics, etc. This complicates joint use of the data in data processing and in calculations. As the result, the problem emerges of integration of heterogenous resources created under significant differences in the model and data representation (Dudarev 2016). Consciousness about this problem motivated large-scale research for refurbishment of the whole data infrastructure aimed at the increase of data volumes with plenitude of structures and models.

The key idea of the future infrastructure is the formation of space combining the data with different structure originally localized within autonomous DB (Bizer 2013, Frenkel 2009). One of the determining ideas in the organization of such space is the use of ontologies, semantically precise and machine-processable definitions of the information entities and their correlations (Uschold and Gruninger 1996). The ontology formalizes the data domain ensuring semantic unity for separate DB. Its function is much wider than the usual taxonomy because it supports logical links between concepts predetermined by the specifics of the data domain. By adding semantics (meaning) and logical connections to the data the ontology describes “knowledge” with possibility of its machine-based interpretation and, thus, becomes the indispensable superstructure in the creation of the integrating infrastructure.

Wide capabilities of ontologies motivated activities on their development for the formalization of multiple disciplines – chemistry, material studies, Earth sciences, etc. (Zhang et al. 2015, Brodaric and Gahegan 2010). Semantic system of knowledge management (Knowledge Organization Systems) close to ontologies in terms of underlying ideas since it is based on the set of glossaries and taxonomies ensuring semantic unity of resources working in the network, has been developed in nuclear power engineering sufficiently long ago.

The present paper is dedicated to the creation of ontology for thermal physics of reactor materials. Thermal physics properties of materials taking into account their significance in the modeling impose stricter requirements on the resources ensuring their collection and systematization. Simplified ontology singling out the basic concepts (substance, property, data) and accompanying terms (dimension, state, uncertainty, source) (Erkimbaev et al. 2015) was suggested earlier for integration of thermal physics data. It was possible within its framework to standardize the set of concepts and to introduce the limitations required by the context, for instance, conformance between the aggregation state of the substance and its proper-

ties. Construction of its analogue for reactor materials is possible only under significantly expanded discursive apparatus covering such previously ignored factors as the composition of the substance, characteristics of the specimen, information about technology, operational modes, etc. Prototype of such ontology (NuclThermo Ontology) establishing its structural elements including taxonomy of materials and properties, models of data representation and their uncertainties, extra factors of the type of porosity, burnup range, irradiation range, etc., axiomatics of the use of material properties as applied to certain classes of materials was outlined in (Chusov et al. 2017).

Implementation of the planned project specifying in details the hierarchy of classes and associative correlations determining logical connections between concepts is suggested. The following section is dedicated to the description of the ontology with possibility of permanent supplementing new concepts, for instance, in the listing of materials and in the nomenclature of material properties. Supplementary information in the form of diagrams illustrating taxonomy of objects and their characteristics, as well as OWL-file coding the contents of the ontology can be found on the website (Thermophysics 2018) because of significant volume.

The third section is dedicated to the issue which is most important from the viewpoint of practical use – how to ensure integration of several DB on thermal physical properties with different formats, data models and their semantics based on the existing ontology. Virtual connection with common interface and unified interrogation system is understood as the integration. Vast experience of application of ontologies for integration of DB in material studies (Zhang et al. 2015) allows considering the task of integration of data on reactor thermal physics to be fully realizable.

## The project of RTO ontology

Designing ontologies (Uschold and Gruninger 1996) begins, as a rule, from singling out fundamental concepts borrowed from domain-independent **BFO** ontology (Basic Formal Ontology) (Fig. 1).

As it is evident, the original **Entity** concept/class has two successors: **Continuant** for entities preserving their identity with time, and **Occurent** for entities having temporary parts and developing with time. The **Continuant** class has the following as the successors: **Independent Continuant** and **Dependent Continuant** for objects of the data domain and their attributes. The **Dependent Continuant** class also has two daughter classes: **Specifically Dependent Continuant** (combining characteristics of the objects) and **Generically Dependent Continuant** (summing data types and corresponding concepts) with one **Information\_content\_entity** daughter class. The **Specifically dependent continuant** class has two daughter classes: **Quality** (the whole population of properties of objects, processes and documents), as well as **Realizable**

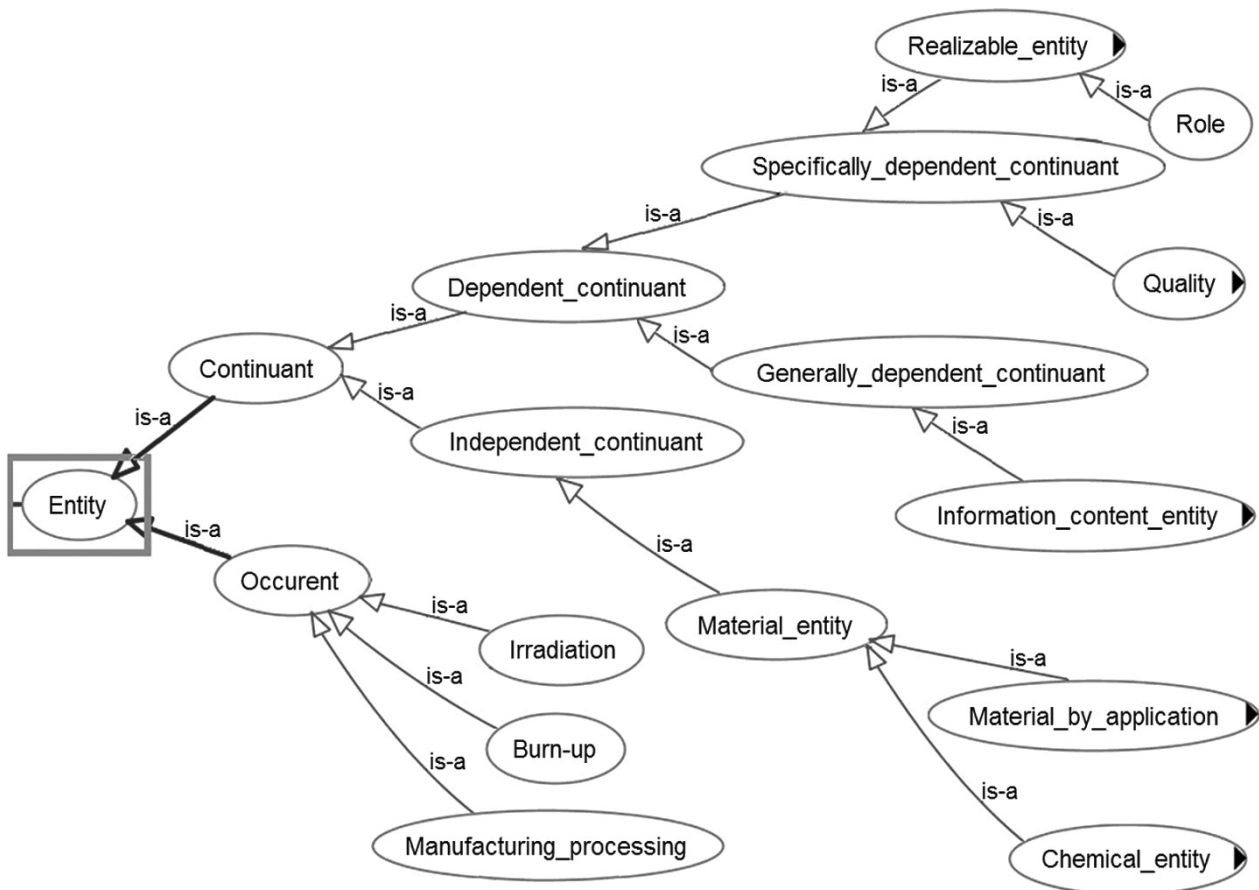


Figure 1. Classes of the upper level of the ontology.

**entity** for defining objects in terms of their possible roles or capabilities. Rigorous and universal structure adaptable to any data domain is thus formed. The following levels of the hierarchy including fundamental concepts of reactor thermal physics are also shown in the figure. All types of substances and materials are represented by the **Material\_entity** class; **Quality** class has the following daughter classes: **Material\_Quality**, **Sample**, **State**, **Information\_content\_quality**, **Process\_quality**. The first of these classes sums up all types of physical characteristics such as density, heat conductivity, etc. The **Sample** class determines characteristics of specific sample (manufacturer's brand, form and structure of the sample, etc.), the **State** class singles out the aggregation state (gas, liquid, etc.). The latter two classes characterize information entities (document, identifier) and processes.

Specifics of reactor thermal physics is determined first of all by the hierarchy of materials taking over the **Material\_entity** class (Fig. 2). Its daughter class **Material\_by\_application** combines in accordance with accepted classification (Kirillov (Ed.) 2006) materials of the following five categories: fuel, coolant, moderator, absorbing and structural materials. Another daughter class **Chemical\_entity** includes the substances or multi-component systems not belonging to the above categories.

The most typical classes are shown in Figure 2, more detailed diagram (Thermophysics 2018) contains cases

representing widely spread types of reactor materials and their numerous specimen. Composite materials and salt fuel compositions and their specimen, for instance, cermet  $\text{UO}_2\text{-Zr}$  fuel or  $\text{LiF-BeF-ThF}_4\text{-UF}_4$  molten salt, are indicated there. Metals and alloys from p-block of the periodic table (Pb, Pb-Bi) are shown as coolants, oxide systems  $\text{Dy}_2\text{O}_3\text{-TiO}_2$ ,  $\text{Dy}_2\text{O}_3\text{-HfO}_2$ , etc. are represented as moderators.

Investigation of materials applied in nuclear engineering often cover the objects not falling under the five categories specified above. Systems of U-O type including  $\text{UO}$ ,  $\text{UO}_2$ ,  $\text{UO}_3$ ,  $\text{U}_3\text{O}_7$ ,  $\text{U}_3\text{O}_8$ ,  $\text{U}_4\text{O}_9$  oxides in different phase states, solutions and compounds of Np, Am, Cm (the so-called minor actinides), fission products in mixtures with other materials, as well as numerous compounds and particles (atoms, molecules, isotopes) belong to such. All of them refer to the hierarchy initiated by the **Chemical\_entity** class furthermore reflecting the possible role or assigning in the upper-level **Role** class (see Fig. 1).

The taxonomy represented in Fig. 2 unambiguously illustrates the possibility of its expansion due to new materials or categories thereof. For instance, **CerMet\_fuel** class (class of composite fuels) of the  $\text{UO}_2\text{-Zr}$ ,  $\text{UO}_2\text{-Al}$  type can be added with its subsequent expansion due to  $\text{U-PuO}_2$ ,  $\text{U-PuN}$ ,  $\text{U-PuC}$  systems (Mishra et al. 2018) including the matrix from U instead of non-fissionable substances, for instance, Zr or Al. Another important moment

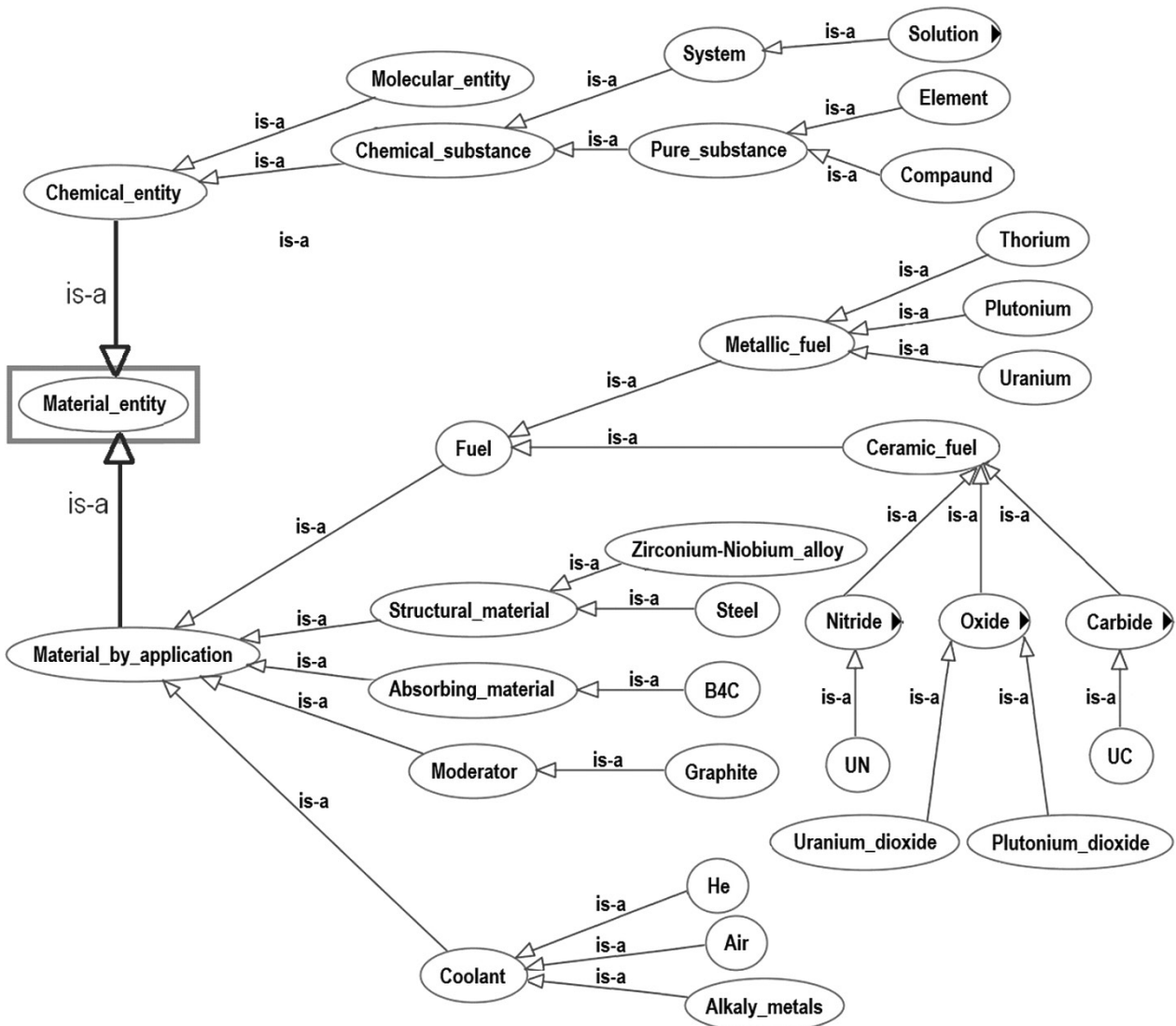
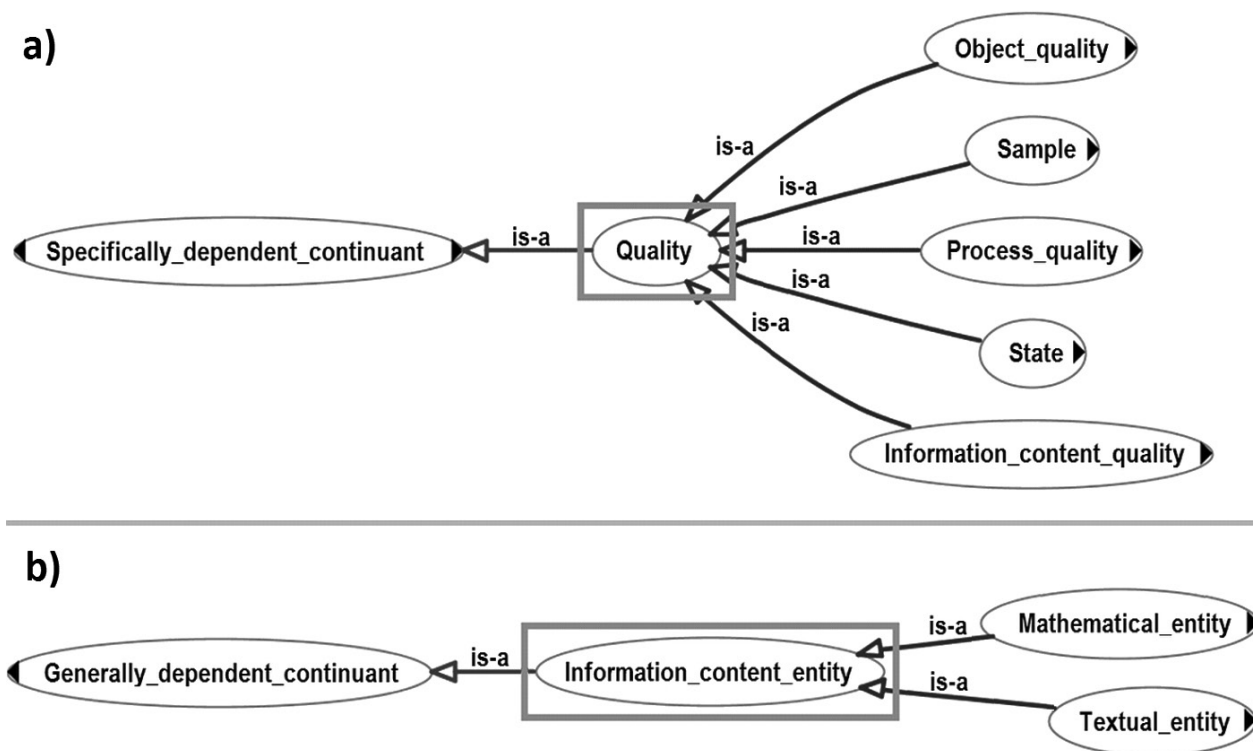


Figure 2. Classes determining the types of materials – abridged version of the hierarchy.

is the presence of the **Sample** class (see Fig. 1) providing detailed specification of specific specimen in addition to general properties of the material. As a rule, it summarizes such details as the manufacturer's brand (grade), form, dimensions, structure, deviation from stoichiometry, for instance, for  $\text{MO}_{2+x}$  oxides. Specific features of reactor materials are also reflected by the **Occurent** class (see Fig. 1) which includes the manufacturing technology (specimen preparation), fuel irradiation and burnup. For instance, for identification of absorbers of  $\text{Dy}_2\text{O}_3\text{-TiO}_2$ ,  $\text{Dy}_2\text{O}_3\text{-HfO}_2$  (Risovanny et al. 2005) type besides the composition it was necessary to specify the specimen shape (in the form of a pellet), its dimensions (diameter and height), as well as manufacturing technology (sintering or smelting). Both capabilities of the ontology permitting expansion of the hierarchy and detailed description of the specimen allow solving the problem of continuous adjustment of the data structure to the specific features of new objects and concepts (Erkimbaev et al. 2008).

Along with **Material\_entity** class the degree of completeness of data representation is ensured by two up-

per-level classes: **Quality** and **Information\_content\_entity** (Fig. 3). The first of these classes includes entities determining properties of materials, processes, information entities, as well as properties of the specimen and state of the matter. The second class - **Information\_content\_entity** – includes entities reflecting data representation in publications or in DB. These entities ensure identification of objects and information resources, representation of documents, formation of numerical data. The first two tasks are performed by the **Textual\_entity** class with **Identifier**, **Document**, **Document\_part** daughter classes. Keywords, names, ID, etc., are anticipated among identifiers. All concepts required for representation of numerical data are combined in the classes which are daughter in relation to **Mathematical\_entity** class. They include typical concepts – **number**, **quantity**, **variable** (with segregation of **independent** and **dependent**) as well as other concepts. The concept of **Data set**, i.e. the *container* with certain structure containing constants and the matrix of values of several physical quantities (density, heat capacity, enthalpy, etc.), is the leading concept. Indication of used measu-



**Figure 3.** Classes determining characteristics of: a) – material objects; b) – information entities.

rement units is made by the **Unit\_of\_measurement** class. Let us note that only the beginnings of those hierarchies are shown in Fig. 3 which are generated by the **Quality** and **Information\_content\_entity** classes, while more exhaustive picture can be found on the website (Thermophysics 2018). The whole population of physical properties, for instance, the data on phase transition diagrams and crystal structures included in the **Material\_quality** hierarchy can be found there. **Sample** and **State** classes combine the data for the specimen and phase state and **Property\_quality** class includes characteristics of processes, for instance burnup depth.

Special value is imparted to the data by the **Information\_content\_quality** hierarchy. It includes the **Uncertainty**, **Status\_descriptor** (subdividing data types – experimental, calculated, reference, recommended) and **Quality\_descriptor** classes, i.e. the overall assessment of the data quality taking into account their uncertainty, completeness of description of the specimen and the method, degree of repeatability, etc.

The potential realized here (Uschold and Gruninger 1996) implies active use of classes and separate concepts from other ontologies developed previously and represented on the web. This principle ensures coordination of heterogenous glossaries excluding nonidentity in the definition of one and the same concept.

Two ontologies **Semanticscience Integrated Ontology** (SIO) and **Chemical Information ontology** (ChemInf) from the **Ontobee** ([www.ontobee.org](http://www.ontobee.org)) repository were used along with BFO ontology for importing concepts. SIO ontology represents wide selection of types and correlations for description of scientific activities.

ChemInf ontology covers typical concepts in chemistry and material science, such as molecule, solution, substance, etc. Terms from external ontologies cover practically completely the set of concepts which are daughters as related to such classes as **Chemical\_entity**, **Information\_content\_entity**, **Information\_content\_quality** and, partially, **Material\_Quality**. All new concepts and classes corresponding to them are intended, mainly, for entities typical specifically for reactor thermal physics and, first of all, those, which are included in the **Material\_by\_application** hierarchy (see Fig. 2 and, in more comprehensive form, on the website (Thermophysics 2018)).

In full volume efficiency of importation from external ontologies is reflected in the selection of the so-called object properties or associative correlations. The above described hierarchy of classes per se is just a glossary in the form of taxonomy of concepts for objects, their characteristics, processes, documents, etc. Object properties establish logical connections between the object and its characteristic, the object and part thereof, document and the object of its description, etc. Already mentioned **Semanticscience Integrated Ontology** ([www.ontobee.org/ontology/SIO](http://www.ontobee.org/ontology/SIO)) is accepted here as a suitable source of object properties. The ontology includes a set of 207 object properties capable to adequately represent almost all imaginable connections between objects and their attributes. Limited selection from this set containing the most characteristic connections is presented in Table 1.

Each of the properties is defined by the ID included in the unique web address (URL). Thus, index SIO\_000028 defines the URL [http://semanticscience.org/resource/SIO\\_000028](http://semanticscience.org/resource/SIO_000028).

**Table 1.** Limited selection of typical object properties (associative correlations), imported from the SIO ontology

Item no.	Object Properties	ID
1	Is_related_to	SIO_000001
2	Denotes	SIO_000020
3	Has_part	SIO_000028
4	Is_denoted_by	SIO_000060
5	Is_part_of	SIO_000068
6	Is_contained_in	SIO_000128
7	Contains	SIO_000202
8	Is_connected_to	SIO_000203
9	Has_quality	SIO_000217
10	Is_quality_of	SIO_000218
11	Is_source_of	SIO_000219
12	Has_role	SIO_000228
13	Is_component_part_of	SIO_000313
14	Is_covalently_connected_to	SIO_000334
15	Is_weakly_interacting_with	SIO_000335
16	Has_component_part	SIO_000362
17	Is_described_by	SIO_000557
18	Describes	SIO_000563
19	Has_identifier	SIO_000671
20	Has_data_item	SIO_001277
21	Is_data_item_in	SIO_001278

rdf, where detailed description of the *has\_part* property is provided (Fig.4). Example of its use is provided by the proposition **molecule has\_part some atom**; the *is\_part\_of* (SIO\_000068) property is used for the converse proposition.

As the result the RTO ontology defines coordinated glossary of terms and specification of their meaning for reactor thermal physics, which allows conducting search and formulating logical conclusions. It is significant that it is possible to continuously expand classes associated with appearance of new materials or new factors, for instance, fuel nanostructure. Expandable listing of classes which are daughters with respect to **Sample** allows including originally not introduced specific features of the specimen such as configuration, impurities, porosity, etc. And, finally, enormous set of text and mathematical terms ensures the required adjustment of the data type, for instance, replacement of a single number with a data set or interval of values.

## Ontology as the means for designing and integration of DB

**Mutual complementarity of ontologies and DB.** New direction emerged presupposing synthesis of DB with ontologies for the purpose of utilization of advantages of both instruments. Clear correlation between elements of DB and ontology forms its basis. Thus, certain classes correspond in the ontology to the entities of the DB, properties of classes correspond to the attributes, and axioms correspond to the limitations. Their main differences amount to the fact that the ontology describes the structure of the data domain using formal language, while the DB conceptual layout describes certain DB without pretending to disseminate knowledge. In other words, DB is focused specifically on the data (numerical, text, etc.), while the ontology is oriented towards the interpretation of their meaning and realization of automated inferences. Loss of semantics is the major limitation of DB because the meaning of the entities is not accessible to persons not familiar with configuration of the DB.

Another critical drawback of DB is the impossibility to maintain evolution of the data schema. Practice of data systematization according to their properties demonstrates unsuitability of rigid structure for wide combination of substances taking into account the specifics of specimens (Erkimbaev et al. 2008). For instance, evolution of the data schema can be associated with specifics of operation of nuclear reactors which is reflected on the properties of materials.

At the same time DB ensure the highest productivity in the realization of complex queries not achievable for other types of architecture. This is exactly why implementation of systems combining accessibility of semantics with high productivity of operations with data is justified. **Database-to-ontology mapping** when possibilities of DB are augmented by the connection with ontology for queries using common semantics or integration of heterogeneous DB is realized in the DB design or integration tasks. Ontology ensures semantics, i.e. the glossary, correlations between concepts and the data structure, but,

### has part [SIO:000028]

**label** has part  
**identifier** SIO:000028  
**description** has part is transitive, reflexive and antisymmetric relation between a whole and itself or a whole and its part.  
**term type** ReflexiveProperty  
TransitiveProperty  
ObjectProperty  
**subproperty of** is location of  
**superproperty of** has proper part  
**inverse of** is part of  
**is defined by** <http://semanticscience.org/ontology/sio.owl>

**Figure 4.** Sample of representation of the *has\_part* object class in the SIO ontology.

nevertheless, it does not contain specimens with the role of the latter played by DB records. Combination of data schemas and the data per se is achieved in the process of integration, and the user formulates in terms of ontology the queries which are converted into queries to the DB. Thereby, the ontology plays the role of efficient intermediary between the user and the data. It was shown in (Erkimbaev et al. 2015) to what extent the ontology proves to be efficient in designing and integrating DB as applied to thermal physics and material studies data.

**Examples of efficient use of ontologies by the properties of substances.** Chemical DB ChEBI ([www.ebi.ac.uk/chebi/](http://www.ebi.ac.uk/chebi/)) is the most popular among DB by the properties of substances. Unique identifier ChEBI ID (for example, CHEBI:15377 for H<sub>2</sub>O) accessible for citation by the web user or by the software agent (it is sufficient to enter ChEBI ID of the required entry in Google to access the entry) is assigned to each of the molecular entities (as well as their groups and classes) included in the DB. Set of data includes elementary information about the molecule (chemical formula, mass, charge), structural data, hyperlinks to other DB, in particular, NIST DB ([webbook.nist.gov/chemistry/](http://webbook.nist.gov/chemistry/)) containing wide scope thermal dynamics information.

Each entry contains as well the fragment of ontology with possibility of navigation through parent and daughter classes. Associative correlations of the type of those represented in Table 1 indicate the objects functionally associated with the original object; for instance, correlation *has\_role* connects the entry for H<sub>2</sub>O with the entry for the term “greenhouse gas”, while the correlation *has\_part* indicates the hydrate for which H<sub>2</sub>O serves as a constituent part. Starting navigation from a certain entity it is possible to get access to a variety of other entries by defining their logical or role connections incorporated in multi-level taxonomies, which allows realizing complex queries involving search of substance according to the structure, functions, role, etc. As a whole, it means that along with the data for the specific substance ChEBI provides a fragment of “knowledge” from the domain area.

Mapping of external query generated in terms of ontology as a set of sub-queries to isolated data sources without their aggregation in a single repository is used for integration of several DB. The central component of such virtual system is the *intermediary* with unified access interface and unified model based on the ontology or the set of ontologies. Processor included in the intermediary performs the division of the query into sub-queries to data sources. The so-called *adaptors* are anchored to data sources for resolving problems emerging because of heterogeneity of models. Thus, the use of “intermediary-adapter” architecture of the global model in the form of ontology and data *federation*, i.e. their fusion in the course of execution of query to the system from autonomous DB, constitutes the basis of integration. The details of the technology for different approaches to integration are discussed in (Laallam et al. 2014, Kogalovsky 2012).

The potential of integration based on **Matinfo** ontology is examined in (Ashino 2010). For the “Structural materials” domain the ontology is built in the form of seven sub-ontologies among which the main concepts are furnished by the following four basic sub-ontologies: **Substance** for substances, mixtures and materials; **Property** for chemical, thermal and mechanical properties; **Process** for production and measurement methodologies; **Environment** for ambient characteristics (composition of the atmosphere, temperature, pH, etc.). **Materials Information** sub-ontology aggregating all terms and concepts characterizing the material and specific specimen, methods and conditions, data quality criteria, etc., is included in the overall ontology along with the four basic sub-ontologies. Two peripheral ontologies **Unit Dimension** and **Physical Constant** cover the needs in the representation of measurement units and constants.

Ontology resolves the problem of data exchange between three DB in Japan: AIST (Advanced Industrial Science and Technology), NIMS (National Institute of Material Science) and widely used structure **MatDB** (<https://odin.jrc.ec.europa.eu>) containing data on material testing. The main bulk of information in the above listed DB refers to thermal and mechanical properties. Data exchange is realized using the intermediate level occupied by **Matinfo** ontology.

Structure of NIMS data is adjusted to storage of measurement data with main metadata, **thermalConductivity** and **chemicalFormula** stored in relative fields. In contrast, metadata in AIST DB are specified by the user with a character string entered in the field named **property** which allows expanding or modifying the names of properties under fixed schema. Additional difficulty is introduced by the fact that AIST DB is capable to store both scalar and tensor values of heat conductivity typical for monocystals or materials with pronounced anisotropy of the structure. The same possibility is fixed in the ontology as well where both scalar value and arbitrary matrix can be reflected.

The system of data exchange between the above mentioned **MatDB** in the composition of Online Data & Information Network for Energy and the reference book **Gen IV Materials Handbook** (<https://gen4www.ornl.gov/>) prepared in Oak-Ridge National Laboratory is an example of such integration in nuclear engineering. Integration was implemented without physical aggregation of resources by conversion of two different formats into a unified data format (Lin et al. 2015). The technique of data import and export in both resources used XML representation not unifying the semantics which requires detailed knowledge of both conceptual frameworks. Noticeable complication of the problem with increasing number and variety of resources is the additional argument in favor of ontological approach to DB integration.

**Big Data technology of resource integration tasks.** Application of approaches intended for work with big data opens new emerging possibilities in the integration of scientific resources. As applied to materials properties

the source of big data is the flow of publications with data volume determined, moreover, not so much by the number of investigated objects than by the vast variety of conditions of synthesis and measurements, microstructural specifics, etc. It is specifically the latter among the above three attributes for assigning to the category of “big data” (3V – Volume, Velocity, Variety) which plays the decisive role as applied to the data on the properties. It is shown in (Erkimbaev et al. 2017) that this approach allows overcoming two main problems on the road to integration of resources with minimal expenditures, namely vast variety of schemes, terminology, data types and formats and the need of continuous adjustment of the created structure to variations of nomenclature of objects and concepts. The proposed solutions were based on the joint use of technologies which earlier were applied separately: interchange standard in the form of structured text documents; ontologically based data management; **Apache Spark** (<http://spark.apache.org/docs/>) big data handling platform. Application of these technologies allows converting the variety of primary resources (DB, file archives, etc.) into a standard JSON text format with subsequent use of ontologies for semantic integration. Tasks of storage of heterogeneous data, provision of access to the data and data analysis are delegated in this case to **Apache Spark** platform.

JSON-format proved to be one of the most convenient formats for data exchange due to the simplicity of reading and editing, friendly in terms of human comprehension and storage of hierarchical structures. JSON-format is the working unit for a number of platforms, for **Apache Spark** in particular, allowing organizing the exchange, storage and organization of queries for dispersed data.

Scenario developed in (Erkimbaev et al. 2017) provides for the conversion of each of the resources into JSON-documents connected to the repository including both domain specific and upper level ontologies. The role of ontologies consists of entering semantics in the documents, as well as of the possibility to enter corrections in the data structure. Coupling documents with ontologies allows conducting ontological search revealing information on the upper and lower levels (parents and daughter classes) and side connections (*related terms*) not knowing the schemes of the sources in question.

**Apache Spark** platform allows overcoming the expanding volume and dispersed nature of the data on the properties due to high productivity and manifested orientation towards data handling including data storage, processing, analysis in the distributed environment.

Among other technological features it is distinguished by the presence of inbuilt libraries for analytical processing, including for organization of SQL-queries by means of which access can be gained to the contents of structured JSON-documents. It is specifically the possibility of making SQL-queries to massive data arrays that plays the key role in the problem of their integration. Another feature determining the efficiency of Spark for storing and processing the data is the capability to support interrogation with numerous types of storage - from **HFDS** (*Hadoop Distributed Files System*) to conventional DB on local computers.

Advantages of **Apache Spark** amount to high speed of calculations, handling data of different origin (text, semi-structured and structured data) from different sources (files in different formats, DBMS and thread-specific data). In combination the proposed means open practically unlimited possibilities in terms of productivity and diversification of handling complex data which include the data on properties of complex compositions with which reactor thermal physics deals.

## Conclusion

Expediency of implementation of new approach to handling data on properties of materials based on ontologies as the means ensuring integration of heterogeneous data, is substantiated. Rich capabilities of ontologies in the areas associated with use of vast data arrays are supported by their wide implementation for the purposes of systematization, search and building logical architecture by computer. Option of ontology is suggested for thermal physics data characterized with strict requirements on the representation of uncertainty, details of experiment and specimen, etc. It is demonstrated that editing the ontologies allows overcoming two critical problems of data handling, namely, evolution of operated DB associated with expansion of nomenclature of objects and concepts and accounting for additional factors influencing properties of material (manufacturing technology, dimensions and structure of the specimen, influence of ambient conditions, operational conditions, etc.). Examples are examined of application of ontologies in chemistry and in material studies as the means for integration of autonomous resources with different structure and data format. It is suggested to use certain technologies of the so-called “big data” originally adapted to handling multiple heterogeneous sources as the most efficient means.

## References

- Ashino T (2010) Materials ontology: an infrastructure for exchanging materials information and knowledge. *Data Science Journal* 9: 54–61. <https://doi.org/10.2481/dsj.008-041>
- Bizer C (2013) Interlinking scientific data on a global scale. *Data Science Journal* 12: GRDI6-GRDI12. <https://doi.org/10.2481/dsj.GRDI-002>



- Brodaric B, Gahegan M (2010) Ontology use for semantic e-Science. *Semantic Web 1*: 149–153.
- Chusov IA, Kirillov PL, Bogoslovskaya GP, Yunusov LK, Obysov NA, Novikov GE, Pronyaev VG, Erkimbaev AO, Zitserman VYu, Kobzev GA, Trachtengerts MS, Fokin LR (2017) Development of the system of reactor thermophysical data on the basis of ontological modeling. *Journal of Physics Conference Series* 891: #012172. <https://doi.org/10.1088/1742-6596/891/1/012172>
- Dudarev VA (2016) Integration of information systems in inorganic chemistry and material science. KRASAND Publ., Moscow, 320 pp. [in Russian]
- Erkimbaev AO, Zitserman VY, Kobzev GA, Fokin LR (2008) The logical structure of physicochemical data: Problems of numerical data standardization and exchange. *Russian Journal of Physical Chemistry A* 82 (1): 15–25. <https://doi.org/10.1134/S0036024408010020>
- Erkimbaev AO, Zitserman VY, Kobzev GA, Kosinov AV (2015) Associating the ontologies with databases by properties of substances and materials. *Nauchno-Tekhnicheskaya Informatsiya (Ser. 2)* 12: 1–6. [in Russian]
- Erkimbaev AO, Zitserman VYu, Kobzev GA, Kosinov AV (2017) Standardization of Storage and Retrieval of Semi-structured Thermophysical Data in JSON-documents Associated with the Ontology. Proc. of the XIX International Conference “Data Analytics and Management in Data Intensive Domains” (DAMDID/RCDL’2017), Moscow, Russia, October 10–13. <http://ceur-ws.org/Vol-2022/paper36.pdf> [accessed Aug 10, 2018]
- Frenkel M (2009) Global Information Systems in Science. Application to the Field of Thermodynamics. *Journal of Chemical & Engineering Data* 54 (9): 2411–2428. <https://doi.org/10.1021/jc800877f>
- Kirillov PL (2006) Thermophysical Properties of Materials for Nuclear Engineering: Tutorial for students of specialty “Nuclear Power Plants”. IATE, Obninsk, 182 pp.
- Knowledge Organization Systems and Semantic Technology (2018) International Atomic Energy Agency. Nuclear Knowledge Management (NKM). <http://www.iaea.org/nuclearenergy/nuclearknowledge/> [accessed Aug 10, 2018]
- Kogalovsky MR (2012) Ontology-based data access systems. *Programming and Computer Software* 38(4): 167–182. <https://doi.org/10.1134/S0361768812040032>
- Laallam FZ, Kherfi ML, Benslimane SM (2014) A survey on the complementarity between database and ontologies: principles and research areas. *Int. J. Computer Applications in Technology* 49 (2): 166–187. <https://doi.org/10.1504/IJCAT.2014.060528>
- Lin L, Austin T, Ren W (2015) Interoperability of materials database systems in support of nuclear energy development and potential applications for fuel cell material selection. *Materials Performance and Characterization* 4(1): 115–130. <https://doi.org/10.1520/MPC20150003>
- Mishra S, Kaity S, Banerjee J, Nandi C, Dey GK, Khan KB (2018) U-PuO<sub>2</sub>, U-PuC, U-PuN cermet fuel for fast reactor. *Journal of Nuclear Materials* 499: 272–283. <https://doi.org/10.1016/j.jnucmat.2017.11.010>
- Risovanny VD, Zakharov AV, Muraleva EM (2005) New advanced absorbing materials for thermal reactors. VANT. Ser.: Fizika radiatsionnykh povrezhdeniy i radiatsionnoe materialovedenie 3: 87–93. [in Russian]
- Thermophysics (2018) Web Portal THERMOPHYSICS.RU. <http://www.thermophysics.ru/RTO/index.html> [accessed Aug 10, 2018]
- Uschold M, Gruninger M (1996) Ontologies: principles, methods and applications. *Knowledge Engineering Review* 11(2): 93–136. <https://doi.org/10.1017/S0269888900007797>
- Zhang X, Zhao C, Wang X (2015) A survey on knowledge representation in materials science and engineering: An ontological perspective. *Computers in Industry* 73: 8–22. <https://doi.org/10.1016/j.compind.2015.07.005>