Data sources on the thermophysical properties of metals and alloys, a sketchy overview

I V Savchenko
Kutateladze Institute of Thermophysics Siberian Branch of Russian Academy of Sciences, 1, Academic Lavrentiev Avenue, Novosibirsk, 630090, Russia
E-mail: savchenko@itp.nsc.ru

Abstract. A brief review about data sources on the thermophysical properties of metals and alloys is presented. The basic advantages and disadvantages of popular information sources are shown. Approaches to the choice of the most reliable data are discussed in general. A simple algorithm for searching the most appropriate property information for specific cases is recommended.

1. Introduction
Thermophysical properties are the most important physical characteristics of substances and materials necessary when performing thermal calculations in the design of devices and installations. It is difficult to overestimate the importance of these parameters for the tasks of energy, metallurgy, mechanical engineering, and many other industries. Information on the properties of substances can be contained in various sources; moreover the data can be significantly different, both qualitatively and quantitatively. It is difficult to answer accurately which type of information source is better, or which source is better for solving a specific problem. Different sources provide different information, in different ways, and in varying degrees, in detail. The article is a brief overview of the most popular data sources on thermophysical properties, which shows their advantages and disadvantages, features, outlines the ways of their development in today rapidly changing world. On the one hand, the aim of the work is to attract the attention of the professional community to the question of the most appropriate way to publicly exchange information about properties and to involve the maximum number of interested parties in the discussion. The problem discussed in the article is especially relevant due to the rapid development of Internet services for collecting, storing and processing data about the material properties. On the other hand, acquaintance with the review will allow a person faced with the problem of finding reliable data on thermophysical properties to formulate the problem correctly, evaluate its capabilities and resources for its solution, and understand where to start and in which direction to move. Of course, the issues raised in the article require a more comprehensive and more detailed discussion.

2. Data sources

2.1. Experimental results
The most obvious way to obtain data about the parameters characterizing the material is a direct measurement of them. The experiment is the basis of the modern concept of scientific knowledge; all knowledge is based on it. From this point of view, the primary experimental results are the least
distorted source of information, which is an obvious advantage. Other significant advantages are the possibility of adapting experiments to specific tasks, the possibility of expanding the investigated interval of state parameters, the ability to select samples that meet certain criteria, increasing the varied parameters, and the like. The presence of primary experimental information, including recording signals of measuring sensors, it is possible to audit the accuracy of the results obtained, evaluate and analyze errors, refine the results by processing the primary data using a more advanced calculation model, it is even possible to extract additional information about the properties of the material under study by a more careful analysis.

The talk about the shortcomings can start with the issue price: the experiment has been and remains the most expensive way to obtain information about the properties of substances. Measurements require high-quality equipment and accessories, as well as highly qualified specialists. Only a few centers specializing in the study of properties can allow themselves both.

The results obtained in the experiment do not have absolute accuracy, there is always uncertainty, and most experimenters carefully evaluate it and give it along with the main results. Typical uncertainty values in measurements of thermophysical properties, in not too exotic areas of state parameters, are a few percents; there are exceptions both to the greater and to the lesser extent. A moment requiring a separate discussion is subjectivity associated with the views and personal qualities of the researcher. For example, the experimenter confidence in the correctness of any theory can lead to modernization of the experimental setup or methods for calculating the results and uncertainties, which in turn will lead to a distortion of the results. Often situations arise when the results of various studies performed by different groups diverge significantly beyond the limits of the total errors. This can indicate both the presence of unaccounted for uncertainties and the difference in the views of the researchers. Perhaps it is unpredictable subjectivity that is the main disadvantage of experimental information about properties.

2.2. Scientific articles
The results of experiments, calculations, and analytical reviews are often published in scientific journals of the corresponding profile. Information published in the articles is the result of a deeper analysis. As a rule, at the time of publication, the authors not only calculate the values of the properties and their errors from the primary experimental information, but also try to comprehend them, compare with the results of other researchers and theoretical ideas, and draw conclusions about the presence of correlations or artifacts. Such an analysis allows us to once again critically overview at the results of experiments, and it is possible to evaluate their accuracy and reliability more correctly.

Here, it should be noted that during the work on the publication it is subjected to critical evaluation not only by the authors, but also by reviewers and academic editors. The formers, as a rule, are specialists in the subject area, and the latter have wide erudition. Thus, the data obtained as a result of some kind of biased beliefs will not be allowed for publication. The process of reviewing scientific articles, on the one hand, is a significant advantage, reducing the impact of particular person judgments on published results, on the other hand, the drawback is the lack of the author ability to honestly express his/her position, especially if his/her views are very different from generally accepted ones. The advantages of scientific publications include the relatively low cost of the data obtained, especially in comparison with the experiment.

Disadvantages. Primary information is almost never published in articles; this refers to signals from measuring sensors, in the best case tables with the results of calculating a specific property and the estimated error are published. This makes it impossible to verify the calculations, or to recalculate something using a calculation model different from the original. Often, only approximation dependencies, or small-sized graphs, are published at all. There is no way to expand the measurement range. A certain understatement may take place in the publication, some experimental nuance may not be described, or an important parameter is not indicated, for example, the authors often do not give the purity of the samples or the composition of impurities; it is not always possible to clarify all these features. Finding an article containing the required information can take considerable effort, and take a
long time. The editorial policy of many journals often does not allow quick and efficient publication of the results, for example, some journals prefer to publish exclusively reviews, others do not allow the presentation of data in tabular form, and others introduce very stringent requirements and regulations for published information, and so on. Returning to the issue of subjectivity, for the publication of the article, subject to the requirements of editors and reviewers, authors can quite severely distort the original meaning and content of the articles.

2.3. Reviews
As well as regular articles, reviews are often published in scientific journals, rarely as separate reports or monographs. Typically, reviews are the result of large-scale analyses, containing the results of both experiments and calculations. The authors of the reviews, as far as possible, try to cover the widest range of known data sources about the properties of substances under consideration. Much attention is paid to the discussion of non-trivial issues. These include choosing the most reliable results and optimal methods for measuring and estimating errors, using statistical models to average the results of several studies, searching for correlations, and comparing theoretical concepts with practical observations. This approach allows us to get the most objective and reasonable view of the considered properties. Apparently, it is a carefully executed review, with elements of multivariate analysis that is the most convenient and reliable source of information about the properties of a particular substance or material.

Speaking about the shortcomings, we should start with the fact that any review is not perfect. Some data sources may be excluded from consideration, the influence of important factors for a particular task, and the like, is not considered. Often there are unprofessionally performed analytical studies. Writing a review becomes possible only for sufficiently well-studied substances, it is unlikely to find a review for a promising material whose properties have just begun to be investigated. The remaining advantages and disadvantages are similar to those for the articles.

2.4. Reference books and handbooks
Perhaps, the reference books and handbooks are the most popular, widespread, readily available, conservative and dangerous source of data on thermophysical properties to date. It is believed that only carefully selected, verified information falls into the handbooks. Creation of reference books is trusted by well-known researchers who have achieved success in their field, whose opinion is highly appreciated by the professional community. Often, reference books are published with the support, or by order of well-known organizations, such as CODATA, IAEA, OECD, NASA, NIST, NPL and others. The guarantee of such serious organizations adds credibility to the published information. Excessive trust in reference information is perhaps its main drawback. Users blindly trust the author big name, and the organization under the patronage of which the directory is published, refusing to critically evaluate the information received.

In fact, the results given in the handbooks do not have any advantage over other sources. In many cases, the data is borrowed from articles and reviews, usually not the most recent. In any case, reference recommendations are based on the results of experiments or calculations. An analysis of existing handbooks on the thermophysical properties of metals and alloys made it possible to formulate some general problems. Most popular today reference books provide recommendations based on the results of fairly old works, of the 50s - 80s of the last century. When reprinting directories, typos found are corrected, additional information can be added, new chapters, for example, however, revisions of previously published information are practically not carried out. There is no thorough review of the newly published results, and the addition of background information to the information received. Blind borrowing by the compilers of new reference books of information from old publications deserves special censure. This is very common, and does not lead to clarification of existing information; on the contrary, additional errors and errors due to elementary carelessness may arise during borrowing. Finally, the reference data can be burdened with errors associated with the author views on certain processes. For example, there are often attempts by handbook compilers to
extrapolate the results of experimental work to previously unexplored areas of state parameters, based on unverified hypotheses. Subsequent experiments show the failure of such forecasts, but users of the handbooks are unlikely to find out about this.

2.5. Databases
The most modern, perspective and poorly developed, in the sense of its ability limits, sources of information are databases. In a modern view, a database may contain completely arbitrary information. If the data is useful for us, we can keep them in a convenient format. Given that computer technology has reached enormous success in our time, modern data warehouses allow us to quickly access a giant amount of information. One of the flagship areas of modern information technology, briefly called “Big Data”, is the development of methods for storing and processing arbitrary amounts of information. Along with this, in the last couple of decades, the topic of artificial intelligence has been actively developing, the application of machine learning (ML) methods to the problems of materials science seems to be extremely promising [1]. For training ML-models, digitized data on thermophysical properties that can be easily extracted from databases are required. Returning to the current state of development of databases storing information on the thermophysical properties of metals and alloys, it should be recognized that most of them are very inconsistent with modern methods and approaches to their creation, maintenance and use. The last remark should not be taken as a reproach, an indication of job imperfection of materials scientists and thermal physicists: firstly, information technologies are developing too rapidly, and secondly, the creation of the most famous databases began long before the recent heyday of data science. Currently, most property databases use relational data models and resemble automated handbooks. Compared to conventional handbooks, such databases provide faster access to data, and a slightly more convenient way to use them. But already now there are the projects in which the use of more flexible data models is being promoted, allowing one to store more information and analyze it more thoroughly. Table 1 contains links to some interesting databases, data warehouses, and platforms for analyzing them.

| Link                                                                 |
|----------------------------------------------------------------------|
| https://www.grantadesign.com/                                         |
| https://www.materialsproject.org/                                    |
| https://nanohub.org/                                                 |
| https://www.totalmateria.com/                                        |
| https://app.knovel.com/web/                                          |
| https://citrine.io/                                                  |
| https://www.asminternational.org/materials-resources/online-databases |
| https://www.materialconnexion.com/                                   |
| https://www.materialsweb.org/                                       |
| https://www.ulprospector.com/                                        |
| https://www.chemspider.com/                                          |
| https://www.makeitfrom.com/                                          |
| http://www.matweb.com/                                              |
| https://www.matdat.com/                                              |
| http://iric.imet-db.ru/DB.asp?ido=0&idc=0&nameDB=&ABR=&pgn=1        |
| http://hbcponline.com/faces/contents/ContentsSearch.xhtml            |
| https://matmatch.com/                                               |
| https://materials.springer.com/                                      |
| https://www.nist.gov/data                                            |
| https://trc.nist.gov/metals_data/                                    |
2.6. Predictions
This category includes any methods for calculating properties, starting with simple assumptions about
the similarity of the properties of two substances, and ending with calculations of numerical models
and machine learning algorithms. Methods for predicting thermophysical properties are based on an
analysis of information available in other sources. As a result, they are less accurate. Experimental and
reference data serve as a reference for checking the accuracy of the forecast. The strong point of the
calculation methods is the possibility of their application for the evaluation of properties in the areas of
state parameters, and for such materials for which there are no data from other sources. Many well-
known forecasting methods are characterized by the problem of retraining. That is, the simple models
are quite universal, but inaccurate. The complication of the model leads to the fact that it well
describes a specific property of a particular substance, and becomes completely inapplicable to other
materials.

2.7. Standards
This type of information sources includes data on the basis of which the concepts and mechanisms of
our understanding of the world are created, and the abstractions with which we describe physical
processes are defined. An example of such data are the crystallization temperatures of many metals
(for example, tin, lead or copper), they are the reference points of the temperature scale adopted today
[2]. Such data cannot be clarified until the adoption of new standards.

2.8. Other sources
Technical progress, at present, is developing so rapidly that it is not possible to keep track of all the
innovations. Among the most intensively developing areas of knowledge today, there are also sciences
related to the storage, transmission, dissemination and processing of information. For this reason, in
recent years there are more promising sources of information. Even 30 years ago, information,
especially scientific, was disseminated exclusively in print media, now the trend is that more and more
information is distributed exclusively through the Internet. Scientific journals already have linked web
pages on which not only article lists are published, but also their contents. Many scientific journals are
already focused on the web version, insisting on presenting the results in a convenient way for
publication on the network. The electronic version has significant advantages. It is possible to publish
interactive information, videos, large amounts of data, there is an opportunity to share comments and
discuss with authors and editors. Speaking about discussions, we should mention the development of
social networks, including those for scientists (for example https://www.researchgate.net), in such a
social network you can easily exchange opinions with a rather large professional community. This
provides rich opportunities for analysis and examination of information. Other interesting ways of
presenting information, such as interactive notebooks https://jupyter.org, are also developing. In such a
notebook, it is possible to simultaneously present the primary experimental data, a program for their
processing, text, tables and graphs. The user (reader) of the notebook can borrow and transform any
piece of it. It is possible to apply the processing program created in notepad to your own data set. Or
vice versa, it is possible to process the primary results using your own calculation program. Using a
Jupiter notebook, you can supplement the code or data, build new graphs, compile information from
several notebooks; in general, an interactive notebook is a very flexible tool for data analysis. Other,
no less interesting, sources of information exist and are developing.

Conclusions
In conclusion, I will try to propose an approach for selecting optimal data on thermophysical
properties for specific cases. This is not a trivial task, given the heterogeneity of these sources, each
with its own advantages and disadvantages. The key to its solution is an adequate assessment of
importance of the problem being solved, and importance of information about the properties for its
solution. The customer of the study should answer these questions. If the problem being solved is not
very important, the results of its solution may be inaccurate or weakly depend on the thermophysical
properties, you can use any data that you can quickly find. Such a case is possible, for example: in solving educational problems, or in the initial rough estimation of some parameter. If the task is critically important, the cost of errors in the calculations is high, and their results strongly depend on data about the properties; the issue of selecting thermophysical properties should be approached as attentively as possible. For instance, in the design of nuclear power plants, information on the thermophysical properties of heat carriers and structural materials is critically important. In this situation, there are certain resources to perform a comprehensive review in order to select the most relevant data of the problem, and for the execution of large-scale experimental studies. In general, efforts and resources expended to obtain data on the properties should be proportional to their importance for the problem. And the question of choosing a specific data source is always individual, depends on many factors, and requires analysis of all available information.

Acknowledgments

The study was carried out as part of the state assignment of the IT SB RAS (AAAA-A17-117022850029-9) and the RFBR project No. 15-38-20223.

References

[1] Himanen L, Geurts A, Foster A, Rinke P 2019 Adv. Sci. 6 1900808
[2] Preston-Thomas H 1990 Metrologia 27 3