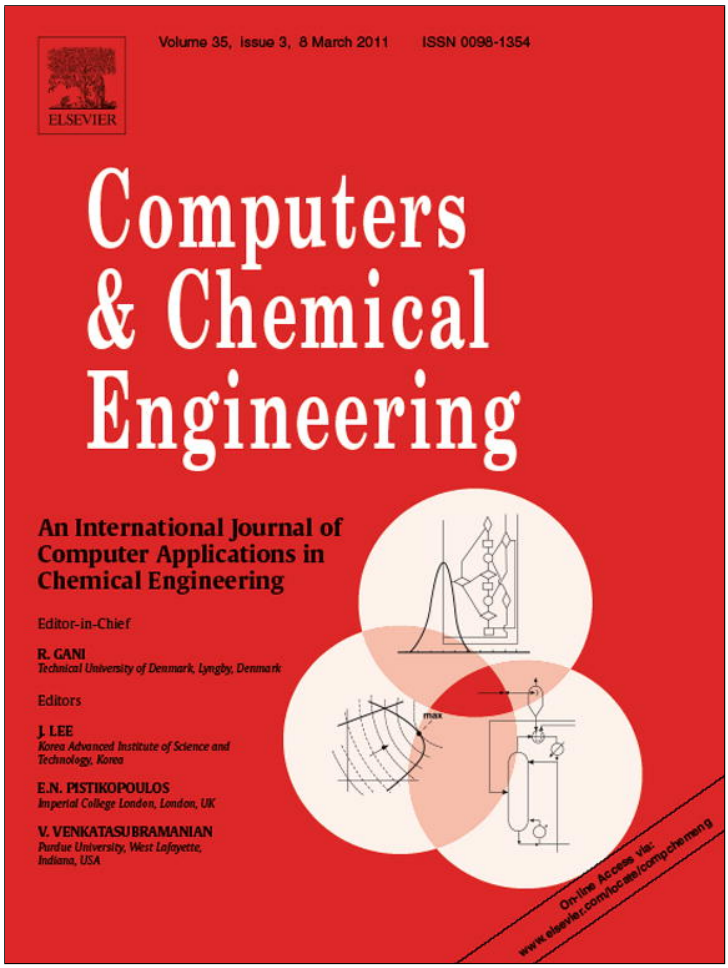

Computers and Chemical Engineering 35 (2011) 393–402

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Perspective

Thermophysical and thermochemical properties on-demand for chemical process and product design☆

Michael Frenkel*

Thermophysical Properties Division, National Institute of Standards and Technology, 325 Broadway, Mail Stop 838.01, Boulder, CO 80305-3337, USA

ARTICLE INFO

Article history:
Received 8 November 2010
Received in revised form
22 December 2010
Accepted 23 December 2010
Available online 30 December 2010

Keywords:
Dynamic data evaluation
Global Information Systems
On demand
Process design
Product design
Thermochemical properties
Thermophysical properties

ABSTRACT

The article provides a perspective of chemical process and product design on-demand, plus its implementation and impact in addressing modern challenges faced by the chemical industry. The concepts of Global Information Systems in Science and Engineering in application to the field of thermodynamics as well as Dynamic Data Evaluation for thermophysical and thermochemical properties are discussed as underlying principles for implementation of chemical process and product design on-demand.

Published by Elsevier Ltd.

Contents

☆ This contribution of the National Institute of Standards and Technology is not subject to copyright in the United States. Products or companies named here are cited only in the interest of complete technical description, and neither constitute nor imply endorsement by NIST or by the US government. Other products may be found to serve as well.

* Tel.: +1 303 497 3952; fax: +1 303 497 5044.
E-mail address: frenkel@boulder.nist.gov

0098-1354/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.compchemeng.2010.12.013

1. Introduction

Thermophysical and thermochemical property information represents a key resource for and critical component of chemical process and product design. In the last 20 years, intense research activities and rapid industrial growth in the fields of biotechnology, specialty chemicals, and medicinal chemistry created unprecedented demand for chemical engineering analysis and simulation of hundreds of new chemical processes and products on an annual basis. In addition, world population growth, reduction in natural resources, and climate change led to a societal recognition of the importance of the implementation of these new processes by the extensive use of energy-saving and environmentally sustainable technologies. Since the development of such chemical processes and products was hindered and, in many instances, impossible without information related to thermophysical and thermochemical properties of the chemical systems (pure compounds, multicomponent mixtures, and chemical reactions) involved, this trend, in turn, provided enormous demand for new thermophysical and thermochemical property data.

This demand in the last several decades has been partially met due to fundamental improvements in experimental measurement technologies resulting from enormous progress in material and computer sciences. That, in combination with new communication technologies and gradually increasing societal commitment to support public scientific research, has resulted in an unprecedented growth in the "production" of reported thermophysical and thermochemical experimental data. Indeed, the total number of experimental thermophysical and thermochemical data points published by the three major journals in the field (Journal of Chemical and Engineering Data; The Journal of Chemical Thermodynamics; Fluid Phase Equilibria) has tripled in the last ten years (from about 50,000 data points published in 1998 to more than 160,000 data points published in 2007, Frenkel, 2009).

However, this dramatic increase in the amount of information available in the public domain until very recently did not make any significant impact in addressing data needs for the chemical process industry. In the last several decades, overwhelmingly, the design of new chemical processes and modifications of existing ones were performed with the use of commercial process simulators traditionally containing very limited thermophysical and thermochemical property information. In our view, progress in further expanding these capabilities was hindered by a number of principal challenges associated with the collection, critical evaluation, quality assurance, and communication of thermophysical and thermochemical property data. In order to address these challenges in a comprehensive manner, the concept of the Global Information System in Science and Engineering (GISSE) has been developed and implemented for the field of Thermodynamics (Frenkel, 2009) at the Thermodynamics Research Center (TRC) of the U.S. National Institute of Standards and Technology (NIST). We define GISSE as information systems designed to collect, process, integrate, evaluate, and communicate the entire "body of knowledge" pertaining to a field and to support any application requiring this knowledge in an "on-demand" mode with definitive information quality assessments (Frenkel, 2009).

Software implementation of the elements of this system, called ThermoGlobe, provides opportunities for chemical process and product design on-demand (Frenkel, 2009; Frenkel et al., 2007). This article is prepared to provide a perspective related to the concepts, technologies, and software tools enabling these fundamentally new capabilities as well as to discuss the potential impact of their further implementation in the future. While implementation of the GISSE makes a significant impact on a broad number of areas of human activity such as the efficiency of information delivery, journal publication quality, molecular modeling and property prediction, strategic experiment planning, scientific discovery process, instrument calibration and validation, one of the major engineering applications of GISSE is chemical product and process design. The implementation of GISSE elements provides unique opportunities to eliminate or at least to reduce the "gap" between publicly available information for thermophysical and thermochemical property data and the information used in practical applications for chemical process and product design. In summary, the purpose of this perspective is to discuss new concepts, technologies, and software tools for critical evaluation and communication of thermophysical and thermochemical data on-demand enabling development of numerous new chemical processes with the use of high fidelity models which, in our view, is critical for a future progress of the process systems engineering.

2. Dynamic data evaluation for thermophysical and thermochemical properties

The evaluation of available experimental and predicted data is a critical element in "filtering" and validating property information. In order to define "evaluation" appropriately, we first have to define various types of data (Frenkel et al., 2004).

True data (hypothetical). *True* data are exact property values for a system of defined chemical composition in a specified state. These data have the following characteristics. They are (1) unique and permanent, (2) independent of any experiment or sample, and (3) a hypothetical concept with no known values. Experimental, predicted, and critically evaluated data may be considered approximations to the true values. The difference between the represented values and true values is defined as the *error*. The *error* is never known; however, it is a given that it is never zero. A measure of the quality or confidence in an experimental, predicted, or critically evaluated value is expressed in terms of the "uncertainty," which is a range of values believed to include the *true* value with a certain probability. All data types can and should have associated uncertainty estimates.

Experimental data. Experimental data are defined as those obtained as a result of a particular experiment on a particular sample by a particular investigator. The feature that distinguishes

experimental data from *predicted* and *critically evaluated* data is the use of a chemical sample including characterization of its origin and purity.

Predicted data. Predicted data are defined here as those obtained through application of a predictive model or method such as a particular molecular dynamics, corresponding states, group contribution method. Clearly, there is no sample associated with this type of property data.

Critically evaluated data. Like predicted data, there is no sample involved with critically evaluated data. The feature that distinguishes *critically evaluated* data from *predicted* data is the involvement of the judgment of a data evaluator or evaluation system. Critically evaluated data are recommended property values generated through assessment of available *experimental* and *predicted* data or both.

Derived data. Derived data can be defined as property values calculated by mathematical operations from other data, possibly including *experimental*, *predicted*, and *critically evaluated* data.

Thus, critical data evaluation can be defined as the process of generation of critically evaluated data based on available experimental and predicted data as well as on their uncertainties.

2.1. Concept

Traditionally, critical data evaluation is an extremely time- and resource-consuming process, which includes extensive use of labor in data collection, data mining, analysis, fitting, etc. Because of this, it must be performed far in advance of a need within an industrial or scientific application. In addition, it is quite common that by the time the critical data-evaluation process for a particular chemical system or property group is complete (sometimes after years of operation of data evaluation projects involving highly skilled data experts), it must be reinitiated because significant new data have become available. This type of slow and inflexible critical data evaluation, which was deployed in various fields of science and engineering for the last 200 years, can be defined as "static" (Fig. 1A, Frenkel, 2007; Frenkel, 2007, 2009). Another common problem associated with static data evaluation is the necessity of having intermediate data storage between every preceding and following stages resulting in a drastic reduction of the overall efficiency of the process. These shortcomings have become magnified dramatically within the last 5–10 years due to the significant increase in the rate of publication of experimental and predicted thermodynamic data that need to be analyzed during the critical data evaluation process.

To address the weaknesses of static evaluations, the concept of dynamic data evaluation was developed at NIST TRC. This concept requires large electronic databases capable of storing essentially all experimental data known to date with detailed descriptions of relevant metadata and uncertainties. The combination of these electronic databases with expert-system software, designed to automatically generate recommended data based on available experimental and predicted data, leads to the ability to produce critically evaluated data dynamically or 'to order' (Fig. 1B). This concept contrasts sharply with static critical data evaluation, which must be initiated far in advance of a particular need. The dynamic data evaluation process dramatically reduces the effort and costs associated with anticipating future needs and keeping static evaluations current.

2.2. Requirements

Implementation of the dynamic data evaluation concept consists of the solution of a number of major tasks (Frenkel, 2007): (1) design and development of a comprehensive database system structure based on the principles of physical chemistry and capable of supporting a large-scale data entry operation for the complete set of thermophysical, thermochemical, and transport properties for chemical systems, including pure compounds, binary mixtures, ternary mixtures, and chemical reactions; (2) development of software tools for automation of the data-entry process with robust and internally consistent mechanisms for automatic assessments of data uncertainty; (3) design and development of algorithms and software tools to assure quality control at all stages of data entry and analysis; (4) development of algorithms and computer codes to implement the stages of the dynamic data-evaluation concept; (5) development of algorithms to implement, target, and apply prediction methods depending on the nature of the chemical system and property, including automatic chemical structure recognition mechanisms; and (6) development of procedures allowing generation of output in a format suitable for application in commercial simulation engines for chemical-process and product design.

2.3. Implementation

2.3.1. Pure compounds

Software implementation of the dynamic data evaluation has been pursued for the last 10 years by the NIST Thermodynamics Research Center within its ThermoData Engine (Frenkel et al., 2010). The first software implementation of the dynamic data evaluation concept limited to pure compounds was reported in 2005 (Frenkel et al., 2005). In that implementation, all thermophysical properties were classified into four property groups or blocks: phase diagram properties, volumetric properties, energy-related properties, and other (including transport) properties. The algorithm of the software assured enforcement of mutual consistency of the thermodynamically related properties within determined ranges of the combined expanded uncertainties for each of the first three designated property blocks, followed by enforcement of thermodynamic consistency between properties belonging to different property blocks at the end.

2.3.2. Equations of state (EOS)

EOS deployment in dynamic data evaluation is particularly challenging for complex, high-precision equations, due to mathematical complexities, the sensitivity of results to data quality, the necessity to meet special validity criteria, and the need to function without human intervention. Dynamic evaluation provides new opportunities for generating EOS representations for a wide variety of chemical species. A key aspect is the establishment of criteria for deployment of a specific type of EOS based on a particular data scenario, i.e., the data quality and extent. The first software implementation for generation of EOS in an on-demand mode was reported in 2007 (Diky, Muzny, Lemmon, Chirico, & Frenkel, 2007). Four different equations of state of various complexity and precision were selected for implementation: Peng-Robinson (Peng & Robinson, 1976), PC-SAFT (Gross & Sadowski, 2001), Sanchez-Lacombe (Koak & Heidemann, 1996; Krenz, 2005) and fundamental equations based on the Helmholtz energy (Lemmon & Jacobsen, 2005; Span & Wagner, 2003). The selected equations were chosen, in part, to provide choices that would be serviceable for a broad variety of data scenarios ranging from extensive, high quality data, such as those available for the common hydrocarbon gases, to effectively no data, such as for new or hypothetical compounds.

2.3.3. Binary mixtures

The first software implementation of the dynamic data evaluation concept for binary mixtures was reported in 2009 (Diky, Chirico, Kazakov, Muzny, & Frenkel, 2009a). In this development, numerous challenges were addressed related to enforcement of consistency amongst the phase equilibrium properties of binary mixtures, which are expressed in a broad variety of ways in the experimental literature as well as enforcement of consistency

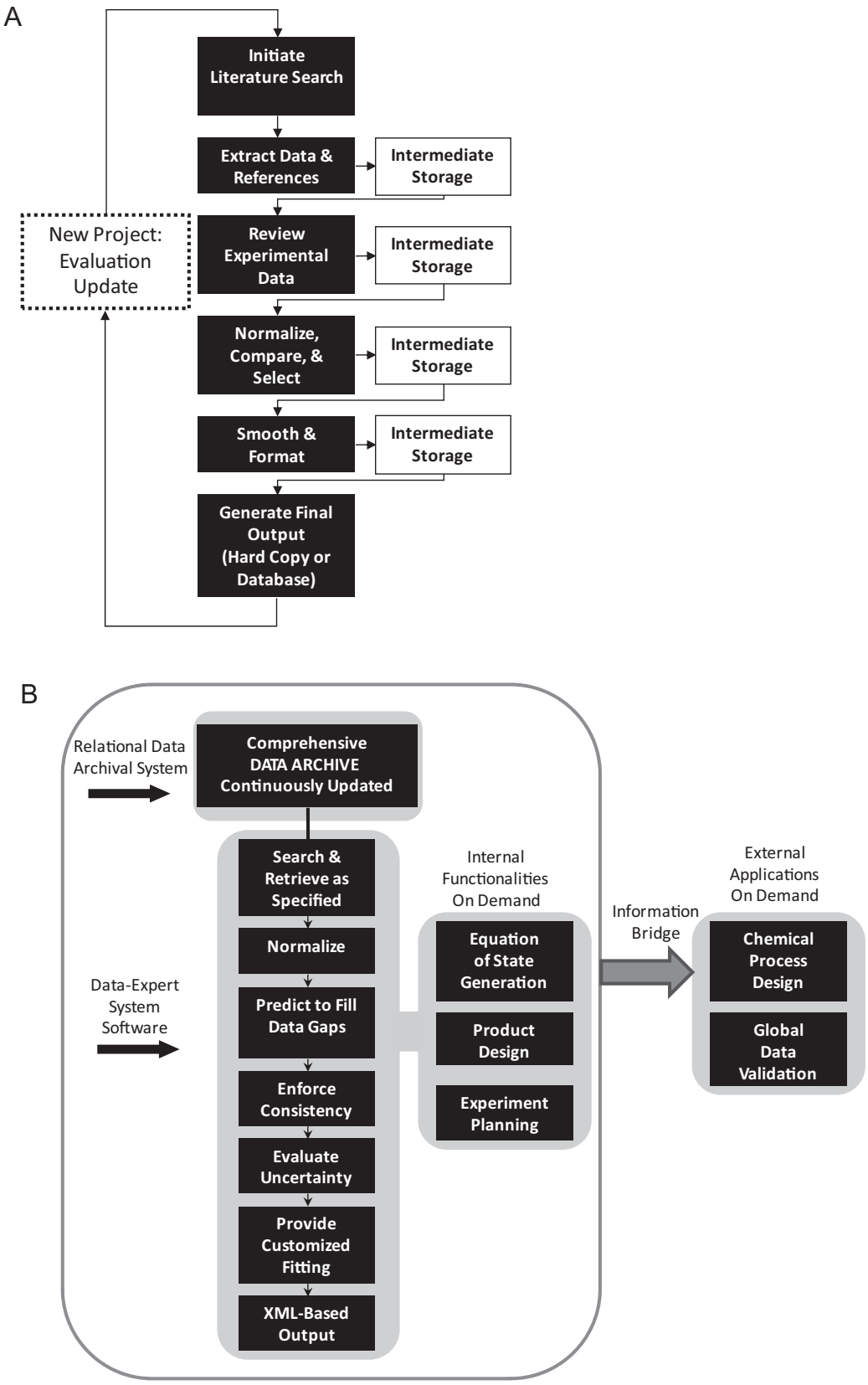

Fig. 1. (A) Schematic representation of static data evaluation performed by an evaluator in advance of use. (B) Schematic representation of dynamic data evaluation performed by a user on demand.

between properties of the pure components and their binary mixtures. This expansion is particularly important from the standpoint of industrial applications, including separation process design. Later (Diky et al., in press), this implementation was further improved by providing capabilities for comprehensive VLE data quality analysis (Jeong Won Kang et al., 2010).

2.3.4. Chemical reactions

The first software implementation of the dynamic data evaluation concept for chemical reactions has been reported in 2009 (Diky, Chirico, Kazakov, Muzny, & Frenkel, 2009b), including processing for both change-of-state and chemical equilibrium information. This implementation was designed to provide the enforcement of thermodynamic consistency between properties of the individual participants of a chemical reaction, experimental (commonly, calorimetric) data for chemical reaction change-of-state properties (such as enthalpies of combustion), and experimental chemical equilibrium data (such as equilibrium constants measured as functions of temperature).

3. Data communications

Efficient and interoperable capabilities of communicating thermophysical and thermochemical property data are critical for development of on-demand chemical process and product design. Such capabilities are essential for rapid delivery of data produced as a result of new experimental measurements and/or new applications of prediction/modeling techniques to their end-users, including thousands of process and plant engineers. On the other hand, lack of these capabilities often represents a "bottleneck" for engineering innovation in "bundling" data products and chemical process engines, hindering the process of information sharing within families of engineering and design applications, slowing cooperative projects between various groups within the same organization and between different organizations, and creating paramount difficulties in supporting team efforts between the end-users.

3.1. Challenges

The history of the development of technical information communication channels and software tools is fairly short and does not provide many examples of large-scale community-based information resources developed as the result of the deployment of such tools and shared by the industrial or scientific and engineering communities. Two well known success stories in this context are associated with the development of the Protein Data Bank (PTB) for storage and exchange of biological macromolecular structures and the Cambridge Structural Database for storage and exchange of small molecule crystal structures.

The challenges related to the establishment of robust communication channels are numerous and are primarily associated with the necessity of assuring their compliance with the "myriads" of currently existing and to be developed algorithmic languages, operation systems, and computational platforms. In addition, communicating thermophysical and thermochemical property data is further complicated by the enormous complexity of their metadata infrastructure including more than one hundred interrelated properties, and the significant number of variables, constraints, phases, and uncertainty measures. A combination of these challenges made the standardization of thermophysical and thermochemical property data communications an insurmountable task for quite a long time in spite of a number of projects initiated between 1985 and 2000 to accomplish this goal (DIPPR-991; Dewan, Embry, & Willman, 2000; Global Cape Open; IUCOSPED; COSTAT, Wilhoit & Marsh, 1987).

3.2. Advantages of the XML technology

Extensible Markup Language (XML) technology (Finkelstein & Aiken, 1999), fully developed within the last 10 years, provides significant advantages for the establishment of standards for data exchange, such as its native interoperability based on ASCII code, its modular nature, and transparent readability by both humans and computers. From a practical standpoint, it is also critical that this technology is currently supported by both software and hardware industries (see, for example, IBM XML Toolkit; Microsoft XML Downloads).

3.3. ThermoML—IUPAC standard

In 2006, ThermoML was established as the standard for the storage and exchange of thermophysical and thermochemical property data (Frenkel, Chirico, Diky, Marsh, et al., 2006), and was adopted by the International Union of Pure and Applied Chemistry (IUPAC). ThermoML, representing an application of XML technology, covers essentially all thermodynamic and transport property data (more than 120 properties) for pure compounds, multicomponent mixtures, and chemical reactions (including change-of-state and equilibrium reactions). The adopted standard included those aspects of ThermoML originally developed for representation of experimental data (Frenkel et al., 2003), uncertainties (Chirico, Frenkel, Diky, Marsh, & Wilhoit, 2003), predicted data, critically evaluated data, and fitting equations (Frenkel et al., 2004). Later, extensions were developed to represent properties of biomaterials (Chirico et al., 2010), speciation, and complex equilibria (Frenkel et al., in press).

3.4. Global data delivery process

The establishment of ThermoML as an IUPAC standard for thermophysical and thermochemical property data communications created new and, indeed, unprecedented opportunities for developing one of the first global data delivery processes (Fig. 2). First, an original thermodynamic data file is generated, then, it is converted to ThermoML format using ThermoML "writer" software, stored in a ThermoML archive, retrieved upon a query, converted into a desired application format using ThermoML "reader" software, and, finally, propagated into a target application. This process, fully implemented for experimental data of pure compounds, binary mixtures, ternary mixtures, and chemical reactions, is now adopted by the five leading journals in the field (Journal of Chemical and Engineering Data, The Journal of Chemical Thermodynamics, Fluid Phase Equilibria, Thermochimica Acta, and International Journal of Thermophysics; see, Cummings et al., 2008–2009; ThermoML Web Archive).

4. Thermophysical and thermochemical property data collection and storage

It is quite obvious that the volume and diversity of available thermophysical and thermochemical property information have a direct impact on the ability of using it for the purpose of chemical process and product design. On the other hand, the quality and reliability of this information is equally critical for it to be used in a variety of engineering applications including chemical process and product design ("garbage in–garbage out" problem well known for the development of software applications.) In the next few sections we shall review how these fundamental issues are being addressed using modern database management and software tools.

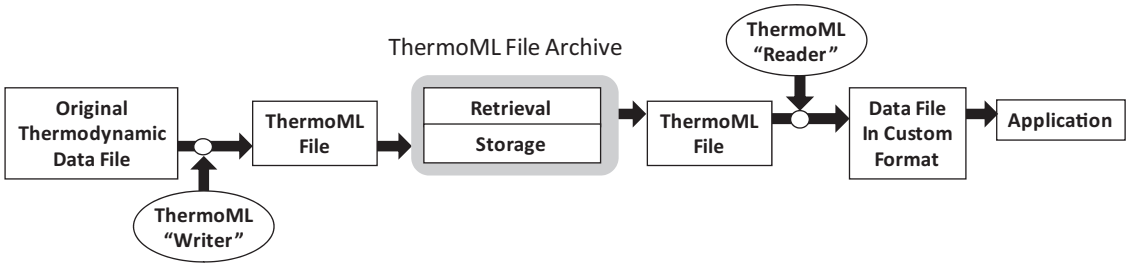

Fig. 2. Global delivery process for experimental thermophysical and thermochemical property data.

4.1. Challenges

Addressing the need of large-scale thermophysical and thermochemical property data collection under strict data quality guidance represents a number of principal challenges: (1) the ability to process an enormous amount of information, currently corresponding to about 300,000–500,000 experimental data points a year and continuously growing at a very high rate; (2) the interpretation of a highly complex metadata infrastructure; (3) the capability to detect "suspicious" data points reported in the original literature; (4) the assessment of combined expanded uncertainties as the only meaningful measure of the overall data quality; and (5) the preservation of the internal integrity of the data records.

4.2. Data archival systems

The first generation of data archival systems, in general, including those with an emphasis on thermophysical and thermochemical property data, was produced in the format of hard-copy books and could be tracked back to the end of the 19th century. The Landolt-Börnstein Series (1883–2010), CRC Handbook of Chemistry and Physics (1913), and International Critical Tables (1926–1930), represent the best examples of that generation which have made an enormous impact on the development of chemical engineering. A unique project, API 44, initiated at the U.S. National Bureau of Standards under the leadership of Frederick D. Rossini in 1942, resulted in the first broad-based hard-copy archival system for critically evaluated thermophysical and thermochemical property data focused on properties of hydrocarbons (Rossini, 1976; Rossini, Pitzer, Arnett, Braun, & Pimentel, 1953). This project was later continued until 2009 in the TRC Thermodynamic Tables (Hydrocarbons, 1942–2009; Non-hydrocarbons 1955–2009) by the Carnegie Institute of Technology, Texas A&M University, and the U.S. National Institute of Standards and Technology (NIST). At the end of 1970s, two other major projects were initiated as hard-copy data compilations. One of them, carried out by the Design Institute for Physical Properties (DIPPR), was focused on thermodynamic properties of pure compounds and operated first at Pennsylvania State University and later at Brigham Young University. The other one, the Dortmund Data Bank (DDB), was initially focused on the properties of mixtures and developed at the University of Oldenburg and later as a private enterprise in cooperation with DECHEMA (German Society for Chemical Engineering and Biotechnology). In the mid-1980s, following the dramatic development of main-frame and personal computer technology, many of these products were moved into electronic format, first as non-relational data archival systems. The TRC SOURCE Data Archival System (Frenkel, Dong, Wilhoit, & Hall, 2001) designed exclusively to store original experimental thermophysical and thermochemical property data and their uncertainties, was one of the first such systems and was developed at Texas A&M University. In the mid-1990s to the beginning of 2000s, following the development of reliable commercial relational data management systems, most of the principal thermophysical and thermochemical property data archival systems were transferred into a relational environment. Commercial relational data management systems provide significant built-in capabilities to preserve the internal integrity of data records, addressing one of the challenges associated with the collection and storage of thermophysical and themochemical property data. The TRC SOURCE Data Archival System (SOURCE), currently operated under the ORACLE data management system at the U.S. National Institute of Standards and Technology, has been designed to reflect the fundamental structural principles associated with the science of thermodynamics. In line with that, all major structural features of SOURCE follow the Gibbs Phase Rule concept and terminology. That, in turn, provides opportunities for the development of robust tools to address the challenge of interpretation of the highly complex metadata infrastructure for thermophysical and thermochemical property data.

4.3. Data capture software

In order to address the challenge of processing a high volume of incoming information pertaining to thermophysical and thermochemical properties under strict data quality guidance, Guided Data Capture (GDC) software was developed (Diky, Chirico, Wilhoit, Dong, & Frenkel, 2003). This software was originally developed to facilitate and enhance the process of data collection and storage for the TRC SOURCE Data Archival System. GDC serves as a data-capture expert system by guiding extraction of information from the literature, ensuring the completeness of the information extracted, validating the information through data definition, range checks, etc., and guiding uncertainty assessment to ensure consistency between compilers with diverse levels of experience. A key feature of GDC is the capture of information in close accord with customary original-document formats and leaving transformation to formalized data records and XML formats within the scope of the software procedures. GDC completely relieves the compiler of the need for knowledge related to the structure of the SOURCE data system or XML formats, thereby eliminating common errors related to data types, length, letter case, and allowable codes. The users of GDC are scientists with varying levels of experience but with competence in the fields of chemistry and chemical engineering. GDC was developed to serve as a powerful and comprehensive tool to be used for both SOURCE in-house data capture operations as well as a data-collection aid for authors of scientific and engineering publications. This software is now used by hundreds of scientists and engineers worldwide and is available for free downloading *via* the NIST TRC Web site (GDC).

4.4. Data quality assurance

Data quality assurance for thermophysical and thermochemical property data includes three principal tasks: (1) validation of the reported numerical and metadata; (2) audit of the collection and storage protocols designed to propagate originally reported property data into the storage facility; and (3) independent assess-

ment of the uncertainties associated with the numerical values to be stored.

Validation of new experimental data on a broad scale is an extraordinarily difficult task in science and engineering. Recently, with the development of the global data delivery process in the field of thermodynamics, discussed above, such a validation process was established (Frenkel, Chirico, Diky, Muzny, et al., 2006) for thermophysical and thermochemical property data. This process which now involves the authors of all relevant articles submitted to major journals in the field, allows for (a) checking of experimental data to be published for integrity (GDC communication line) and (b) checking consistency with the recommended data evaluated on the basis of the complete body of knowledge available to date (TDE communication line). Consistency is checked for a particular property as well as for those linked through thermodynamic identities and correlations. Implementation of this validation process provides capabilities for data reporting improvements even before articles are published. Nevertheless, it has been conservatively estimated that at least 10% of articles reporting experimental thermodynamic data for organic compounds contain some erroneous information (Frenkel, Chirico, Diky, Muzny, et al., 2006). In a great many instances, these errors would be extremely difficult to detect during the article preparation, submission, and peer-review process. It is a safe assumption that the situation described here with regard to thermodynamic data is typical for many other scientific fields dealing with large arrays of scientific experimental data.

A systematic approach to implementing database integrity rules was established through the use of modern database technology, statistical methods, and thermodynamic principles to assure reliable propagation of originally reported property data into a storage facility (Dong et al., 2002). Procedures and tools for error prevention, database integrity enforcement, scientific data integrity protection, and database traceability are incorporated in this approach. Examples of errors occurring in the process of information collection and propagation from the literature are as follows: (a) typographical errors; (b) unit-conversion errors; (c) report interpretation errors (i.e., misreading the original documents); (d) metadata compilation errors etc. Extensive use of the Guided Data Capture software is a critical part of this approach.

In accordance with the *Guide to the Expression of Uncertainty in Measurement* (GUM, 1993), the combined expanded uncertainty represents the most comprehensive measure of reliability for experimental scientific data. In 2003, the GUM was further interpreted for the field of thermodynamics (Chirico et al., 2003). However, although gradual and continuous progress has been made in the reporting of uncertainty information, it was shown that comprehensive uncertainty analyses remain rare (Dong, Chirico, Yan, Hong, & Frenkel, 2005). It is therefore important to establish algorithms for the independent assessment of the combined expanded uncertainty for thermodynamic data based on the reported information related to the purity of the sample(s) used for a measurement, the quality of the experimental apparatus, and propagation of the uncertainties from the variables and constraints to the property in question. This procedure holds true for properties of all types of chemical systems, including pure compounds, multicomponent mixtures, and chemical reactions. In addition, chemical family trends, consistencies between related properties of various types of chemical systems, and compliance with relevant thermodynamic relationships can serve as alternative measures of data quality. A recently developed algorithm for assessment of the data quality of vapor–liquid equilibrium data in the subcritical region based on a variety of tests related to compliance of reported experimental data with the Gibbs–Duhem equation, as well as vapor pressures of the pure components (Jeong Won Kang et al., 2010), represents an example of the determination of such an additional data quality measure.

4.5. Dynamic information update delivery systems

An essential aspect of the implementation of on-demand chemical process design is to assure that critically evaluated property data used by the process simulation engine are generated *via* the dynamic data evaluation process on the basis of a near comprehensive collection of experimental data, including the most recent data from the literature. For that purpose, the NIST/TRC Web-Oracle infrastructure for continuous dissemination of recently entered experimental thermophysical and thermochemical data stored in the SOURCE Data Archival System has been developed (Diky et al., 2007; Frenkel, 2009). This infrastructure, developed with the use of a multitier architecture is deployed to disseminate over the Web the SOURCE updates to the local TDE-SOURCE databases as components of the TDE software residing on the users' workstations worldwide, while TDE is run as a stand-alone application. These updates are provided periodically (quarterly) rather than on the basis of continuous access to NIST/TRC SOURCE. This reduces the variety of possible TDE SOURCE data conditions to a small number of well-defined states associated with particular dates and allows for unequivocal evaluation traceability, a key requirement for ongoing design of a chemical process (Tremblay & Watanasiri, 2010).

5. Building "information bridges"

Building reliable "information bridges", capable of supporting queries for thermophysical and thermochemical property data from a chemical process simulation engine to software products designed to generate these data on-demand, such as the NIST ThermoData Engine (Frenkel et al., 2010), is critical for implementation of the concept of chemical process and product design on-demand. Below we shall discuss several options to build these "bridges" (Frenkel, 2009). While the discussion below is kept primarily in conceptual terms, the reader can find numerous examples of the practical incorporation of the ThermoData Engine *via* various "information bridges" in the excellent recently submitted article by Watanasiri (in press) as well as in other relevant references provided in this section.

5.1. ThermoML "bridge"—single chemical system implementation

Such an approach can be used in the case where the chemical process software has no "built-in" access to the ThermoData Engine, but has capabilities to "read" ThermoML files (SimSci, 2008). In this case, if a chemical engineer has access to both the chemical process design software in question and ThermoData Engine independently, the ThermoML output file can be generated for a single desired chemical system from ThermoData Engine and saved as the first step. This file can then be used as an input file by the chemical process design software equipped with a ThermoML "reader" (Fig. 3).

5.2. ThermoML "bridge"—batch mode

There is the technical capability of running the calculational core of ThermoData Engine in a batch mode for a significant

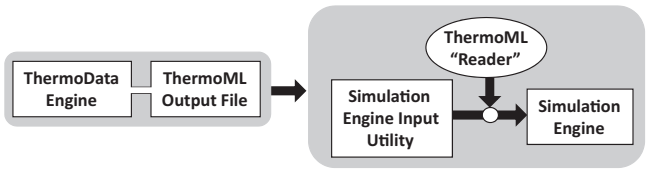

Fig. 3. Implementation of the ThermoML information "bridge" for a single chemical system.

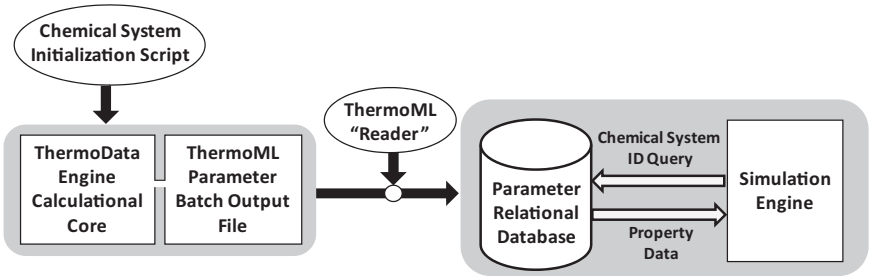

Fig. 4. Implementation of the ThermoML information "bridge" for a family of chemical systems in a batch mode.

group or family of chemical systems (particular class of chemical compounds, all systems characterized with reliable experimental data, etc.). As a result, a ThermoML output file containing thermophysical and themochemical property data for a variety of chemical systems can be generated. This file can then be interpreted by ThermoML "reader" software, and a data storage facility can then be designed to manage the information pertaining to the ThermoML file. This data storage facility can be incorporated into chemical process design software to support process simulation activities (Tremblay & Watanasiri, 2010; Watanasiri, 2010) (Fig. 4).

5.3. Dynamic link library "bridge"

The advantages of dynamic data evaluation are most beneficial when the "information bridges" are built using dynamic link library technology. This allows full incorporation of the calculational core of ThermoData Engine into the chemical process design software to generate critically evaluated thermophysical and thermochemical property data with their uncertainties "to order," and to immediately use them to support process simulation activities including hypothetical compounds and mixtures. In that case, communication between ThermoData Engine and the chemical process design software could be based on a molecular structure of interest, which could be generated by the user during a simulation session (Tremblay & Watanasiri, 2010; Watanasiri, 2010) (Fig. 5).

6. Future opportunities

Implementation of the Global Information System in Science and Engineering concept for the field of thermodynamics and the development of a fundamentally new expert-system software, such as the NIST ThermoData Engine, working on the premise of nearly the entire body of knowledge pertaining to thermophysical and thermochamical property data, creates unique opportunities in various areas of human activity. In our view, many of those opportunities will be explored in the next five to ten years. Below we shall briefly discuss just three of those opportunities associated with product design, strategic experiment planning, and equipment selection for chemical product design.

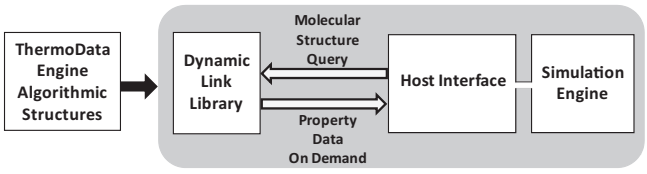

Fig. 5. Implementation of the information "bridge" based on dynamic link library (DLL) technology.

6.1. Product design

Since TDE operates, in principle, on the premise of analysis of the entire body of knowledge available to date, it is clear that it might be used as a powerful instrument to aid in the design of chemical products, where a goal is to solve the "inverse engineering problem" of determining chemical systems possessing desired values of thermophysical properties within defined ranges of tolerance (Diky et al., in press; Gani, 2004). Recently, initial implementation of product design in the TDE software was reported for pure compounds (Diky et al., in press). Properties supported by TDE product design are the melting temperature, enthalpy of fusion, normal boiling temperature, critical temperature, vapor pressure, density, viscosity, thermal conductivity, enthalpy of vaporization, and heat capacity. In addition, the user can restrict the search to specific chemical families and substructures. Multiple properties can be specified, each with a specified tolerance or limit. The distinguishing feature of this approach is the extensive use of experimental property data. A different approach pursued by Molecular Knowledge Systems, Inc. (MKS) and the Computer Aided Process-Product Engineering Center (CAPEC) of the Technical University of Denmark is based entirely on prediction models. We can envision that a combination of the two approaches can yield very powerful software tools that can be used, in particular, for solvent and extraction agent design.

6.2. Strategic experiment planning

Wakeham et al. (2007) envisioned that the future of experimental thermophysical property measurement science will represent a transformation "from accuracy to fitness for purpose." They emphasized that, in a great many instances, the desire for high accuracy in measurements, which was a driving force in the field for many years, did not lead to a better understanding of natural phenomena, nor did it help in the further advancement of theory and simulation. Consequently, one could conclude that the effort and resources associated with these measurements were, to a significant degree, inefficiently used, making the need for improved planning for experiments obvious.

Recently, an algorithmic approach to assist the process of experiment planning with assessment of the entire body of knowledge was implemented using the NIST ThermoData Engine (Diky et al., in press). This approach includes an analysis of availability of experimental thermophysical property data, variable ranges studied, associated uncertainties, the state of prediction methods, availability of parameters for deployment of prediction methods, how these parameters can be obtained using targeted measurements, etc., as well as how the intended measurement can address the underlying scientific or engineering problem under consideration.

It is our belief that expanding the capabilities of the described software and making it available as a free Web tool to experimen-

talists worldwide might, over time, significantly change the way thermophysical and thermochemical property measurements are planned, and ultimately, make a tremendous difference in the "production" of knowledge in the field. Chemical process and product design will, in turn, indirectly benefit from this development as major applications of this knowledge.

6.3. Equipment selection for process design

Whiting and colleagues (Vasquez & Whiting, 2000, 2004; Whiting, Vasquez, & Meerschaert, 1999) studied the impact of uncertainties in thermodynamic data and models on chemical process analysis. While these studies provided some guidance of how the uncertainties of thermodynamic data can be propagated into chemical process design, they are based on the assumption that all parties have a common understanding of what uncertainty means. It was shown that this assumption is difficult to justify on the basis of published literature (Dong et al., 2005). Indeed, experimental thermophysical and thermochemical property data often are characterized with a number of measures of data reliability even though the combined expanded uncertainty is the only comprehensive measure of overall data quality. Consequently, most data storage facilities for thermophysical and thermochemical property data either do not provide uncertainty characterization at all or provide various non-comprehensive precision metrics (most commonly, repeatability or reproducibility). The SOURCE Data Archival System supporting the NIST ThermoData Engine provides independently assessed estimates of combined expanded uncertainties for all experimental data stored. This represents an opportunity for implementation of the algorithms for propagation of the combined expanded uncertainties characterizing thermophysical and thermochemical properties into uncertainties of the properties of process "streams" on the basis of the NIST ThermoData Engine. Such an analysis, in combination with safety considerations, could lead to better optimized process equipment selection to mitigate the fundamental problem of overdesign in chemical process simulation.

7. Conclusions

The multitude of societal demands currently challenging chemical, petrochemical, agricultural, and pharmaceutical industries (environmental control, sustainability requirements, life-cycle analysis, low cost products, high cost of raw materials and energy resources) are impossible to address without dramatic further enhancements in chemical process and product design. In turn, the progress of chemical process design, to a significant degree, depends on the availability of reliable thermophysical and thermochemical property information. We believe that software implementation of the concepts of Global Information Systems in Science and Engineering in application to the field of thermodynamics, as well as Dynamic Data Evaluation for thermophysical and thermochemical properties discussed in this article, provide a foundation for the broad-based development of chemical processes and products on demand, creating opportunities for development of numerous new and improved chemical processes. We hope that this discussion involving major components and examples of such implementation, as well as future opportunities related to this development, outlined here, will promote further interest in chemical process design on-demand amongst process developers and industrial engineers. We strongly believe that extensive use and further exploration of the concepts, technologies, and software tools discussed in the article, will be critical for a future progress of the process systems engineering.

Acknowledgments

I thank Prof. Rafiqul Gani of the Denmark Technical University for the idea of putting together this perspective. I am grateful to Drs. Suphat Watanasiri (AspenTech, Burlington, MA), Marco Satyro (Virtual Materials Group, Calgary, Canada), and David Van Peursem (SimSci-Esscor Invensys, Lake Forest, CA) for their input regarding various ways of establishing communications between the NIST ThermoData Engine and chemical process design software. I am indebted to Dr. Robert D. ("Rob") Chirico and the rest of my colleagues on the scientific team at the NIST Thermodynamics Research Center for countless discussions that were very helpful in sharpening my vision of chemical process and product design on-demand, as articulated in this paper.

References

Cambridge Structural Database. <www.ccdc.cam.ac.uk/products/csd/> 2.09.10.
Chirico, R. D., Frenkel, M., Diky, V. V., Goldberg, R. N., Heerklotz, H., Ladbury, J. E., et al. (2010). ThermoML—An XML-based approach for storage and exchange of experimental and critically evaluated thermophysical and thermochemical property data. 4. Biomaterials. *Journal of Chemical and Engineering Data*, *55*, 1564–1572.
Chirico, R. D., Frenkel, M., Diky, V. V., Marsh, K. N., & Wilhoit, R. C. (2003). ThermoML—An XML-based approach for storage and exchange of experimental and critically evaluated thermophysical and thermochemical property data. 2. Uncertainties. *Journal of Chemical and Engineering Data*, *48*, 1344–1359.
Computer Aided Process-Product Engineering Center (CAPEC), Department of Chemical and Biochemical Engineering, University of Denmark. <http://www.capec.kt.dtu.dk/> 28.09.10.
Cummings, P. T., de Loos, T., O'Connell, J. P., Haynes, W. M., Friend, D. G., Mandelis, A., et al. (2008–2009). Joint statement of editors of Journals Publishing Thermophysical Property Data: Process for article submission for The Journal of Chemical Thermodynamics, Fluid Phase Equilibria, International Journal of Thermophysics, Thermochimica Acta, and Journal of Chemical Engineering Data. *Fluid Phase Equilibria (2009)*, *276*, 165–166.
International Journal of Thermophysics (2009), 30, 371–373; Journal of Chemical and Engineering Data (2009), 54, 2–3; Journal of Chemical Thermodynamics (2009), 41, 575–576; Thermochimica Acta (2008), 484, vii–viii.
Design Institute for Physical Properties. <http://www.aiche.org/dippr/> 22.09.10.
Dewan, A. K., Embry, D. L., & Willman, T. J. (2000). DIPPR/AIChE Project 991—Thermophysical property data exchange. In *Book of abstracts of the 14th symposium on thermophysical properties* Boulder, CO, (p. 169).
Diky, V., Chirico, R. D., Kazakov, A. F., Muzny, C. D., & Frenkel, M. (2009a). ThermoData Engine (TDE): Software implementation of the dynamic data evaluation concept. 3. Binary mixtures. *Journal of Chemical Information and Modeling*, *49*, 503–517.
Diky, V., Chirico, R. D., Kazakov, A. F., Muzny, C. D., & Frenkel, M. (2009b). ThermoData Engine (TDE): Software implementation of the dynamic data evaluation concept. 4. Chemical reactions. *Journal of Chemical Information and Modeling*, *49*, 2883–2896.
Diky, V., Chirico, R. D., Kazakov, A. F., Muzny, C. D., Magee, J. W., & Abdulagatov, I., et al. ThermoData Engine (TDE): Software implementation of the dynamic data evaluation concept. 5. Experiment planning and product design. *Journal of Chemical Information and Modeling, 51*, doi:10.1021/ci100373t, in press.
Diky, V. V., Chirico, R. D., Wilhoit, R. C., Dong, Q., & Frenkel, M. (2003). Windows-based guided data capture software for mass-scale thermophysical and thermochemical property data collection. *Journal of Chemical Information and Computer Science*, *43*, 15–24.
Diky, V., Muzny, C., Lemmon, E. W., Chirico, R. D., & Frenkel, M. (2007). ThermoData Engine (TDE): Software implementation of the dynamic data evaluation concept. 2. Equations of state on demand and dynamic updates over the web. *Journal of Chemical Information and Modeling*, *47*, 1713–1725.
Dong, Q., Chirico, R. D., Yan, X., Hong, X., & Frenkel, M. (2005). Uncertainty reporting for experimental thermodynamic properties. *Journal of Chemical and Engineering Data*, *50*, 546–550.
Dong, Q., Yan, X., Wilhoit, R. C., Hong, X., Chirico, R. D., Diky, V. V., et al. (2002). Data quality assurance for thermophysical property databases—Applications to the TRC SOURCE data system. *Journal of Chemical Information and Computer Science*, *42*, 473–480.
Dortmund Data Bank. <http://www.ddbst.com/en/ddbst/index.php> 22.09.10.
Finkelstein, C., & Aiken, P. (1999). *Building corporate portals with XML*. New York: McGraw-Hill.
Fluid Phase Equilibria. <http://www.sciencedirect.com/science/journal/03783812> 14.01.2010.
Frenkel, M. (2007). From molecular insights and chemical technologies to communications and expert systems: A few short thermodynamic stories. *Journal of Chemical Thermodynamics*, *39*, 169–191.
Frenkel, M. (2009). Global information systems in science: Application to the field of thermodynamics. *Journal of Chemical and Engineering Data*, *54*, 2411–2428.
Frenkel, M., Chirico, R. D., Diky, V. V., Dong, Q., Frenkel, S., Franchois, P. R., et al. (2003). ThermoML—An XML-based approach for storage and exchange of experimental and critically evaluated thermophysical and thermochemical property data. 1. Experimental data. *Journal of Chemical and Engineering Data*, *48*, 2–13.

Frenkel, M., Chirico, R. D., Diky, V. V., Marsh, K. N., Dymond, J. H., & Wakeham, W. A. (2004). ThermoML—An XML-based approach for storage and exchange of experimental and critically evaluated thermophysical and thermochemical property data. 3. Critically evaluated data, predicted data, and equation representation. *Journal of Chemical and Engineering Data*, *49*, 381–393.

Frenkel, M, Chirico, R. D., Diky, V. V., Marsh, K. N., Dymond, J. H., Wakeham, W. A., et al. (2006). XML-based IUPAC Standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture ThermoML: IUPAC recommendations 2006. *Pure and Applied Chemistry*, *78*, 541–612.

Frenkel, M., Chirico, R. D., Diky, V., Muzny, C., Dong, Q., Marsh, K. N., et al. (2006). New global communication process in thermodynamics: Impact on quality of published experimental data. *Journal of Chemical Information and Modeling*, *46*, 2487–2493.

Frenkel, M., Chirico, R. D., Diky, V. V., Muzny, C., Kazakov, A. F., Magee, J. W., et al. (2010). *NIST ThermoData Engine, NIST standard reference database 103b—Pure compounds, binary mixtures, and chemical reactions, version 5.0*. Gaithersburg, MD: National Institute of Standards and Technology, Standard Reference Data Program., <http://www.nist.gov/srd/nist103b.cfm> 8.10.10

Frenkel, M., Chirico, R. D., Diky, V., Muzny, C. D., Watanasiri, S., & Chen, C.-C. (2007). New technologies for product on-demand design. In *Book of abstracts, 11th international conference on properties and phase equilibria for product and process design* Hersonissos, Crete, Greece, (p. 99).

Frenkel, M., Chirico, R. D., Diky, V., Yan, X., Dong, Q., & Muzny, C. (2005). ThermoData Engine (TDE): Software implementation of the dynamic data evaluation concept. *Journal of Chemical Information and Modeling*, *45*, 816–838.

Frenkel, M., Diky, V. V., Chirico, R. D., Goldberg, R. N., Heerklotz, H., & Ladbury, J. E., et al. ThermoML—an XML-based approach for storage and exchange of experimental and critically evaluated thermophysical and thermochemical property data. 5. Speciation and complex equilibria. *Journal of Chemical and Engineering Data, 55*, doi:10.1021/je100999j, in press.

Frenkel, M., Dong, Q., Wilhoit, R. C., & Hall, K. R. (2001). TRC SOURCE Database: A unique tool for automatic production of data compilations. *International Journal of Thermophysics*, *22*, 215–226.

Gani, R. (2004). Chemical process design: Challenges and opportunities. *Computers and Chemical Engineering*, *28*, 2441–2457.

Global Cape Open. http://www-i5.informatik.rwth-aachen.de/lehrstuhl/projects/gco/. Web page contact: Alexander Kuckelberg (kuckelberg@informatik.rwth-aachen.de). 2.09.10.

Gross, J., & Sadowski, G. (2001). Perturbed-Chain SAFT: An equation of state based on a perturbation theory for chain molecules. *Industrial & Engineering Chemistry Research*, *40*, 1244–1260.

(1993). *Guide to the expression of uncertainty in measurement*. Geneva, Switzerland: International Organization for Standardization.

Guided Data Capture software. <http://trc.nist.gov/GDC.html> 24.09.10.

(1913). *Handbook of chemistry and physics*. Cleveland, OH: The Chemical Rubber Company.

IBM XML Toolkit. <http://www-03.ibm.com/systems/z/os/zos/tools/xml/> 3.09.10.

(1926–1930). *International critical tables of numerical data in physics, chemistry, and technology*. New York/London: McGraw-Hill.

IUCOSPED - Joint CODATA/IUPAC Project "Standardization of Physico-chemical Property Electronic Files". <http://iupac.org/web/ins/024-1-99> 2.09.10.

Jeong Won Kang, Diky, V., Chirico, R. D., Magee, J. W., Muzny, C. D., Abdulagatov, I., et al. (2010). A quality assessment algorithm for vapor–liquid equilibrium data. *Journal of Chemical and Engineering Data*, *55*, 3631–3640.

Journal of Chemical and Engineering Data. <http://pubs.acs.org/journals/jceaax/index.html> 14.01.10.

Journal of Chemical Thermodynamics. <http://www.sciencedirect.com/science/journal/00219614> 14.01.10.

Koak, N., & Heidemann, R. A. (1996). Polymer–solvent phase behavior near the solvent vapor pressure. *Industrial & Engineering Chemistry Research*, *35*, 4301–4309.

Krenz, R. A. (2005). Correlating the fluid phase behavior of polydisperse polyethylene solutions using the modified Sanchez-Lacombe equation of state, Ph.D. thesis, University of Calgary, Alberta, Canada.

(1883–2010). *Landolt-Börnstein Series: Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik; Numerical Data abd Functional Relationship in Science and Technology*. Darmstadt-Frankfurt: Springer-Verlag.

Lemmon, E. W., & Jacobsen, R. T. (2005). A new functional form and new fitting techniques for equations of state with application to pentafluoroethane (HFC-125). *Journal of Physical and Chemical Reference Data*, *34*, 69–108.

Microsoft XML Downloads. <http://msdn.microsoft.com/en-us/data/bb190600.aspx> 3.09.10.

Molecular Knowledge Systems, Inc., Bedford, New Hampshire. <http://www.molknow.com/index.htm> 28.09.10.

Peng, D.-Y., & Robinson, D. B. (1976). A new two-constant equation of state. *Industrial & Engineering Chemistry Fundamentals*, *15*, 59–64.

Protein Data Bank. <www.pdb.org/pdb/home/home.do> 2.09.10.

Rossini, F. D. (1976). Fifty years of thermodynamics and thermochemistry. *Journal of Chemical Thermodynamics*, *8*, 803–906.

Rossini, F. D., Pitzer, K. S., Arnett, R. L., Braun, R. M., & Pimentel, G. C. (1953). *Selected values of physical and thermodynamic properties of hydrocarbons and related compounds. API Research Project 44*. Pittsburgh: Carnegie Press.

SimSci-Esscor Thermo Data Manager 3.3. (2008). <http://www.cadfamily.com/download/Chemical/PROII-TDM/TDM%20User%20Guide.pdf> 27.09.10.

Span, R., & Wagner, W. (2003). Equations of state for technical applications. I. Simultaneously optimized functional forms for nonpolar and polar fluids. *International Journal of Thermophysics*, *24*, 1–39.

Thermodynamics Research Center. <http://www.trc.nist.gov/> 14.01.10.

ThermoML Web Archive. <http://trc.nist.gov/ThermoML.html> 13.09.10.

Frenkel, M. (Ed.). (1942–2009). *TRC thermodynamic tables—Hydrocarbons*. Boulder, CO/Gaithersburg, MD: National Institute of Standards and Technology.

Frenkel, M. (Ed.). (1955–2009). *TRC thermodynamic tables—Non-hydrocarbons*. Boulder, CO/Gaithersburg, MD: National Institute of Standards and Technology.

Tremblay, D., & Watanasiri, S. (2010). *Benefits of the NIST ThermoData Engine in aspenONE Engineering, V7.2, White Paper*. Burlington, MA: Aspen Technology, Inc., <http://www.aspentech.com/downloads/whitepaper_nist_benefits_aspenoneV72.pdf> 24.09.10

Vasquez, V. R., & Whiting, W. B. (2000). Uncertainty and sensitivity analysis of thermodynamic models using equal probability sampling (EPS). *Computers and Chemical Engineering*, *23*, 1825–1838.

Vasquez, V. R., & Whiting, W. B. (2004). Analysis of random and systematic error effects on uncertainty propagation in process design and simulation using distribution tail characterization. *Chemical Engineering Communications*, *191*, 278–301.

Wakeham, W. A., Assael, M. A., Atkinson, J. K., Bilek, J., Fareleira, J. M. N. A., Fitt, A. D., et al. (2007). Thermopysical property measurements: The journey from accuracy to fitness for purpose. *International Journal of Thermophysics*, 372–416.

Watanasiri, S. (2010). On-demand critically evaluated thermophysical properties data in process simulation. In *21st international conference on chemical thermodynamics* Tsukuba, Japan, (p. 62). Abstract Booklet.

Watanasiri, S. Development of on-demand critically evaluated thermophysical properties data in process simulation, *Pure and Applied Chemistry, 83*, in press.

Wilhoit, R. C., & Marsh, K. N. (1987). *COdataSTAndardThermodynamics. Rules for preparing COSTAT message for transmitting thermodynamic data (1987). Report to CODATA task group on geothermodynamic data and chemical thermodynamic tables*. Paris, France.

Whiting, W. B., Vasquez, V. R., & Meerschaert, M. M. (1999). Techniques for assessing the effects of uncertainties in thermodynamic models and data. *Fluid Phase Equilibria*, *160*, 627–641.