Global Information Systems in Science: Application to the Field of Thermodynamics[†,‡,§]

Michael Frenkel*

Thermophysical Properties Division, National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305-3328

Global Information Systems can be defined 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. Recent advances in computer hardware technology, development of relational data management systems capable of reliably supporting storage of enormous amounts of information pertaining to specific areas of science, new interoperable means of standardizing data communications based on the Extensible Markup Language (XML) ontologies, new generation software expert systems, and secure online networks have created unprecedented opportunities for defining and implementing a new paradigm of Global Information Systems in Science. Global Information Systems have a profound impact on the scientific discovery process, industrial development, and knowledge communication. Within the last 10 years, one of the first Global Information Systems in Science has been developed for the field of Thermodynamics (ThermoGlobe) at the Thermodynamics Research Center (TRC) of the U.S. National Institute of Standards and Technology (NIST). The components of this system include software tools for mass-scale data capture (Guided Data Capture Software), a comprehensive data storage facility (SOURCE Data Archival System), the NIST/TRC Data Entry Facility, a data communication standard (ThermoML − IUPAC standard for thermodynamic data communications), data "reader" software (ThermoML opener into Microsoft Excel), expert system software (NIST ThermoData Engine), and a Web communication portal (NIST Web-Oracle data dissemination channel). The role and principle structure of all the components are discussed in this article, with emphasis on the profound impact of the ThermoGlobe implementation on various areas of research and engineering including, but not limited to, efficiency of information delivery, journal publication quality, and chemical process design.

Introduction

While scientific methods vary significantly in a multitude of research areas, they all require processing of available experimental data, and their analysis and interpretation eventually provide a foundation for the understanding of and, in many instances, the prediction of natural phenomena.

In the last several decades, enormous progress in material and computer sciences has led, in many scientific disciplines, to fundamental improvements in experimental measurement technologies. 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 the reported experimental data. Interestingly, that is true not only for relatively new and dynamically developing research fields such as biotechnology but also for such traditional fields of scientific research as chemical thermodynamics, kinetics, and thermophysics.

Indeed, the statistics accumulated by the Thermodynamics Research Center (TRC)[1] of the Thermophysical Properties Division at the National Institute of Standards and Technology in Boulder, Colorado, USA, specializing in critical data evaluation for thermophysical properties, indicate that the total number of experimental data points characterizing both thermodynamic and transport properties and published by the three major journals in the field (*Journal of Chemical and Engineering Data*,[2] *Journal of Chemical Thermodynamics*,[3] and *Fluid Phase Equilibria*[4]) increased by more than a factor of 3 within the 10-year period between 1998 and 2007 (Figure 1). Such a dramatic growth of information contributes to a number of major challenges related to our ability to process it in order to support the very scientific research that generates it in the first place. Here are just four principal questions associated with these challenges:

- How can one navigate the entire body of knowledge pertaining to the scientific field to find relevant information? (i.e., the needle-in-a-haystack problem)
- How can one develop experimental plans assuring "significant return on investment"?
- Is it possible to validate new experimental data?
- How can one efficiently deliver information from "data producers" to "data users"?

In order to provide adequate solutions to these and other principal questions, the concept of the Global Information Systems in Science (GISS) has been developed at the TRC within the last 10 years and implemented for the field of Thermodynamics. In this paper, we shall discuss the definition of the concept, the critical components for its implementation,

* To whom correspondence should be addressed. E-mail: frenkel@boulder.nist.gov.
† Part of the "William A. Wakeham Festschrift".

§ 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.

10.1021/je800877f
Published on Web 01/22/2009

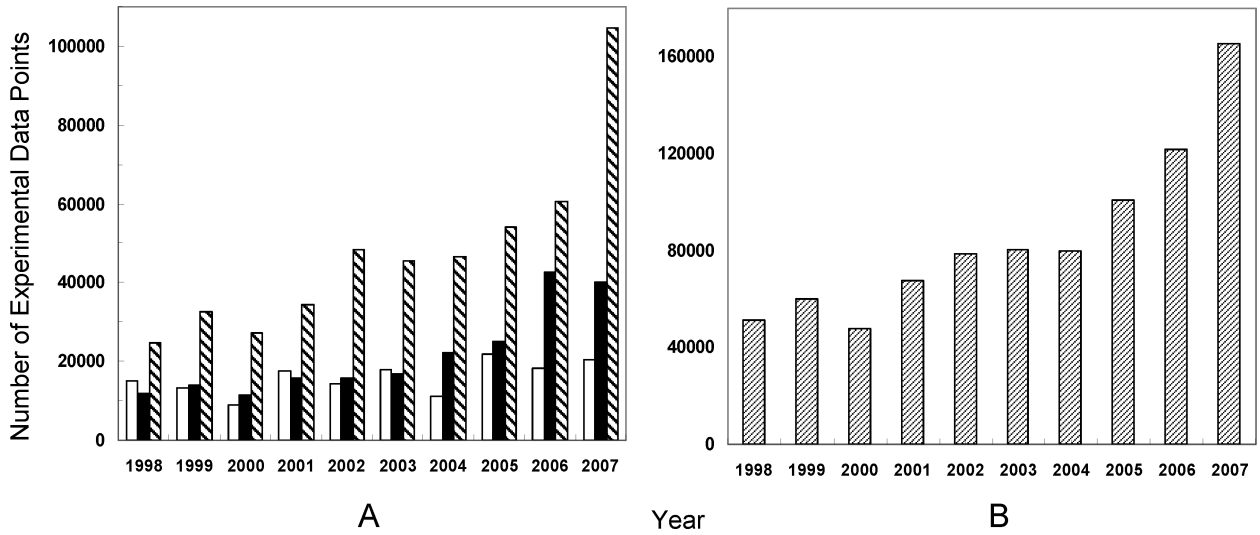

Figure 1. Number of experimental thermophysical property data points published by the *Journal of Chemical and Engineering Data* (A, striped), *Journal of Chemical Thermodynamics* (A, black), and *Fluid Phase Equilibria* (A, white) as well as cumulatively in the three journals (B) on a yearly basis within the period of 1998 to 2007.

and its impact on various areas of human activity with emphasis on the field of thermophysical and thermochemical properties of pure compounds, their mixtures, and chemical reactions. This discussion will summarize our experience in the implementation of the GISS concept for the field of Thermodynamics and the development of the ThermoGlobe system, as well as illustrate examples of its impact.

1. Concept Definition and Principal Areas of Impact

We define Global Information Systems in Science as information systems designed to collect, process, integrate, evaluate, and communicate the entire "body of knowledge" pertaining to a scientific field and to support any application requiring this knowledge in an "on-demand" mode with definitive information quality assessments.

It is quite obvious that practical implementation of the GISS concept is very challenging and depends significantly on the state of computer technology, both hardware and software. Recent advances in computer hardware technology, development of relational data management systems capable of reliably supporting storage of the enormous amounts of information pertaining to the specific areas of science, new interoperable means of standardizing data communications, and new generation software expert systems have created the technological foundation for the full-scale implementation of the GISS concept.

Being implemented, GISS have an enormous impact on a broad range of areas of applications. Indeed, we identified a number of areas of impact for ThermoGlobe, the first system fully implementing the GISS concept for the field of Thermodynamics. They include, but are not limited to:

- Efficiency of information delivery
- Journal publication quality
- Chemical process design
- Strategic experiment planning
- Molecular modeling and property prediction
- Scientific discovery process
- Instrument calibration and validation

In this paper, we shall provide illustrations of the ThermoGlobe impact on a number of these areas.

2. Critical Components

Implementation of the GISS concept consists of a number of critical components. They include:

- Software tools for mass-scale data capture

Software tools designed to support a high-volume data capture process under strict quality guidelines

- Comprehensive data storage facility

Relational data storage facility capable of storing all elements of the meta- and numerical data infrastructure for properties, variables, and constraints pertaining to the research field

- Data processing facility

Organizational facility designed to implement high-volume data capture and data storage facility loading processes

- Data communication standard

Interoperable standard for communicating all elements of the meta- and numerical data infrastructure pertaining to the research field

- Data "reader" software

Software tools designed to convert data from a standardized format to one or more of the commonly used application formats

- Software expert system

Expert system designed to generate on-demand critically evaluated data based on a comprehensive up-to-date combination of experimental and predicted data and their uncertainties

- Web communication portal

Web facility designed to communicate continuously experimental data updates to support software expert systems functionality, as well as to store, display, and communicate the contents of the archives of the experimental data in a standardized format.

Below we describe the structure and the functionality of these components within the ThermoGlobe system.

Software Tools for Mass-Scale Data Capture: NIST Guided Data Capture (GDC) Software. NIST Guided Data Capture software (GDC)[5−7] serves as a data capture guide for extraction of experimentally measured thermophysical and thermochemical property data for pure compounds, binary and ternary mixtures, and chemical reactions from the literature, assuring 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

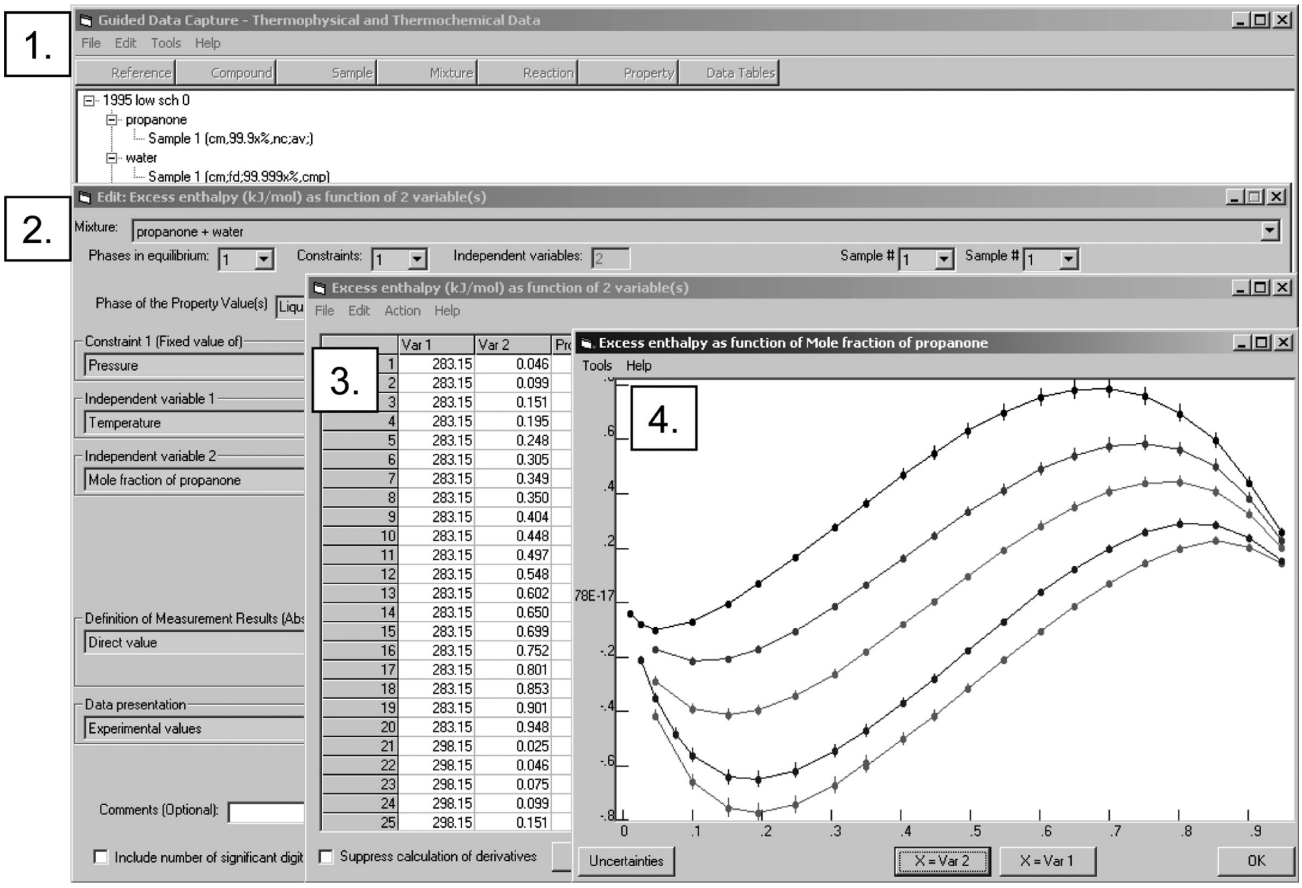

Figure 2. Collage of the four major screens of the Guided Data Capture (GDC)[5,6] software in processing excess enthalpy experimental data for the binary mixture of propanone + water as a function of mole fraction of propanone at different temperatures:[8] 1, navigation tree; 2, metadata information (phases, variables, constraints, units, uncertainties); 3, numerical data; 4, graphical representation.

of experience. Version 5.0, released in 2007, extends the scope of GDC to biothermodynamic properties, such as properties of enzyme-catalyzed reactions, reaction properties determined with titration calorimetry, properties determined with differential scanning calorimetry (DSC), and solubilities in complex media. A key feature of the GDC software is the capture of information in close accord with customary original document formats. The GDC architecture is designed to detect inconsistencies and errors in reported data (erroneous compound identifications, typographical errors, etc.), resulting in improved integrity of the captured data over that given in the original sources.

The compiler's main interactions with GDC involve a navigation tree, which provides a visual representation in accord with the hierarchical structure of the batch data file as it is created. Each node of the tree corresponds to a record in the batch data file structure. Management of records, including deletion, addition, and editing, is accomplished through interactions with the navigation tree. Numerical values are not shown explicitly in the tree but may be accessed through the property-specification nodes. Lists of established field values (journal title abbreviations, compound identifiers, properties, units, phases, experimental methods, etc.) are stored in a local database that is a part of the GDC software. Selection of field values by the data compiler is achieved through single-value or multiple-selection lists of predefined values, which prevent many simple errors. All predefined lists are prioritized to speed access. Keyboard input is never used for direct input of coded information, which eliminates typographical errors. Keyboard input to GDC is provided exclusively for entry of isolated numerical values, general comments, document titles, and new chemical and author names. Most numerical values are captured through electronic means by use of existing HTML, PDF, ASCII, EXCEL, and WORD files or by digitizing hard-copy documents and rarely require manual input. All other input is accomplished through predefined menus, check boxes, or other controlled selection processes. The GDC data processing operation encompasses both metadata and numerical data and provides graphical representation of the numerical data (Figure 2). GDC provides capabilities for automatic conversion of the produced data files to the ThermoML format discussed briefly in the present paper.

Comprehensive Data Storage Facility: NIST SOURCE Data Archival System. The SOURCE Data Archival System,[9] created and maintained by TRC, is a large, general-purpose archive of experimental data covering thermodynamic, thermochemical, and transport properties for pure compounds and mixtures of well-defined composition, as well as for chemical reactions. The database contains numerical values for various kinds of thermodynamic and thermochemical properties of systems in all phases and values of transport properties of fluids. SOURCE does not include properties whose values depend upon the history of samples. It is critically important that the SOURCE database includes the estimated combined expanded uncertainties[10−13] for practically all the numerical data stored. This feature allows, in principle, determination of the quality of recommended data based upon the original experimental data collected in SOURCE.

SOURCE is a relational data system (Figure 3). The information about the authors of the original sources of data is stored in the table AUTHOR linked via the AUTHREF to the

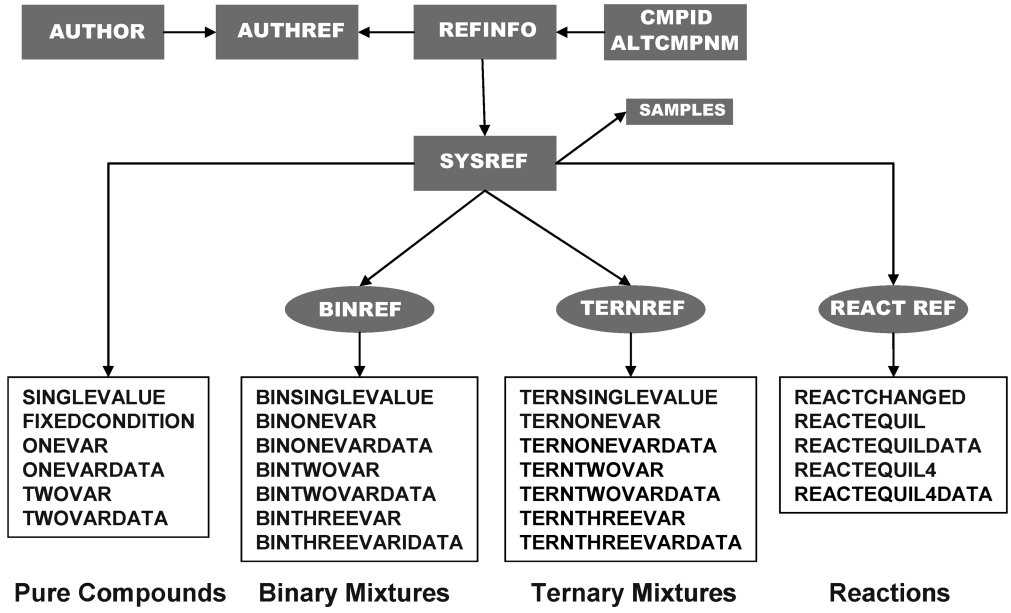

Figure 3. Principal structural elements of the NIST/TRC SOURCE Data Archival System.

information identifying these sources and stored in the table REFINFO. The unique chemical identifiers and their alternative chemical names are populated into the tables CMPID and ALTCMPNM. The table SYSREF links reference and compound information. It identifies compounds whose properties are reported in the document, either in the pure state, as components of mixtures, or as participants in a chemical reaction. Tables BINREF, TERNREF, and REACTREF identify components of binary mixture, ternary mixture, or chemical reaction, correspondingly, whose properties are reported in the cited document. The detailed description of the sample used to perform the measurement includes the source of sample, method of purification, and initial and final purity and the method of their determination as reported by the authors of the document (table SAMPLES). This information is critical for independent assessment of the measurement uncertainties.

The structure of SOURCE follows the application of the Gibbs phase rule. SOURCE groups tables that contain numerical data according to the number of components for pure compounds, binary mixtures, ternary mixtures, and chemical reactions. Within each of these groups, tables express the "effective degrees of freedom" for the data sets. The effective degrees of freedom determine the number of state variables in a data record. The Gibbs phase rule establishes the total degrees of freedom for a system according to the number of independent components and number of phases. The effective degrees of freedom are less than the total degrees of freedom if the system contains additional constraints for a particular data set. Examples of additional constraints are a state variable held constant, a property code that includes a constraint in the definition, and a special state such as a critical or an azeotropic state. Although the Gibbs phase rule does not apply to transport properties, the concept of effective degrees of freedom applies to them as well. Properties for systems with zero degrees of freedom appear in tables corresponding to the number of components in the system. These properties are constants. Data sets for these properties contain only one data record, and each row in these tables contains one data set. In these cases, the data records include all the descriptive information and metadata (tables SINGLEVALUE and FIXEDCONDITION for pure compounds, BINSINGLEVALUE for binary mixtures, TERSINGLEVALUE for ternary mixtures, and REACTCHANGE for chemical reactions with thermodynamic state change, Figure 3). Properties of systems with one or more effective degrees of freedom appear in a pair of tables constructed for the number of components and effective degrees of freedom. Data sets for these properties contain one or more data records. One table contains descriptive information, while another table contains the associated data sets (ONEVAR and ONEVARDATA for one-variable properties of pure compounds, etc., including REACTEQUIL and REACTEQUILDATA for thermodynamic constants characterizing reaction equilibria and REACTEQUIL4 and REACTEQUIL4DATA for apparent equilibrium constants, Figure 3).

SOURCE employs five mechanisms for data integrity: data type definitions, primary key constraints, foreign key constraints, check constraints, and database triggers. Each table or each column defines its mechanisms to prevent invalid data entry. The mechanisms also enforce predefined data rules associated with information in SOURCE. If a data transaction violates an integrity constraint, it is rejected, and an error message is displayed. SOURCE is maintained via the ORACLE Database Management System and currently resides on the SPARC T5240 Sun server.

As of September 2008, the SOURCE Data Archival System contained more than 3.4 million experimental data points for thermophysical and thermochemical properties of more than 19 000 pure compounds, 31 000 binary mixtures, 7000 ternary mixtures, and 5000 chemical reactions.

Data Processing Facility: NIST/TRC Data Entry Facility. The NIST/TRC Data Entry Facility (Figure 4) was opened at the NIST/Boulder campus in 2001. It was originally designed to process about 300 000 experimental thermophysical and thermochemical property data points per year under strict data quality assurance guidelines. Between 2001 and 2008, these capabilities were further increased to about 500 000 data points per year to process experimental data obtained as a result of cooperation with major journals in the field (*Journal of Chemical and Engineering Data*,[2] *Journal of Chemical Thermodynamics*,[3] *Fluid Phase Equilibria*,[4] *Thermochimica Acta*,[14] and *International Journal of Thermophysics*[15]), international cooperation with other data centers worldwide, as well as in-house data collection and critical evaluation operations. International

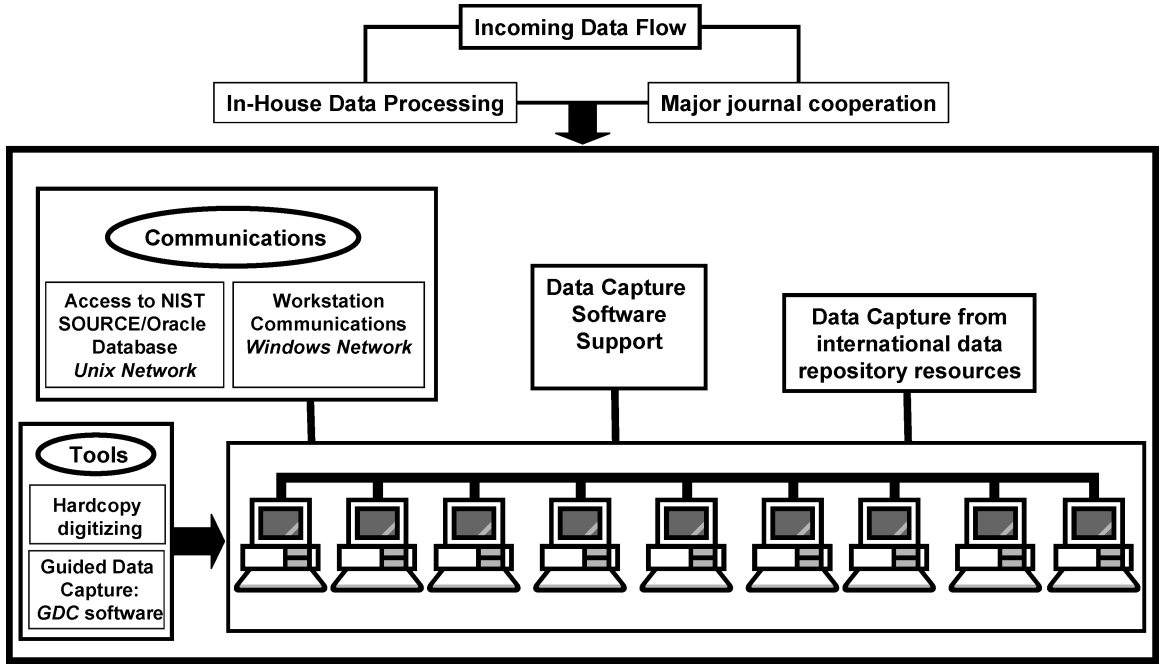

Figure 4. Operational structure of the NIST/TRC Data Entry Facility.

cooperation is particularly critical to support data processing from the original sources based on other than the Latin alphabet. This cooperation includes submission of all relevant experimental data published in the Chinese language and Russian language for the last 50 years by the Center of Thermal and Fluid Science of the Xi'an Jiaotong University, China[16] (targeting such journals as *Journal of Chemical Engineering*, *Applied Acoustics*, *Technical Acoustics*, *Acta Physico-Chimica Sinica*, *Acta Petroleim Sinica* (Petroleum Processing Section), *Journal of Chemical Engineering of Chinese Universities*, *Oil and Gas Storage and Transportation*, *Specialty Petrochemicals*, *Acta Polymerica Sinica*, *Engineering Chemistry and Metallurgy*, *Journal of Chemical Industry and Engineering* (China), *Journal of Fuel Chemistry and Technology*, *Chemical Engineering of Oil and Gas*, *Journal of Sun Yat-Sen University*, *Chemical Journal of Chinese Universities*, *Journal of Beijing Institute of Petrochemical Technology*, *Journal of Shanghai Jiaotong University*, *Journal of Petrochemical Universities*, *Journal of Refrigeration*, *Chinese Journal of Inorganic Chemistry*, *Chinese Journal of Analytical Chemistry*, *Journal of Engineering Thermophysics*, *Journal of Daqing Petroleum University*, *Chemical Reaction Engineering and Technology*, *Acta Petrolei Sinica*, and others), and by the Moscow Power Engineering Institute, Russia[17] (targeting such journals as *High Temperature Thermal Engineering*, *Physical Engineering Journal*, *Journal of Physical Chemistry (Russ.)*, *Journal of Applied Chemistry (Russ.)*, *Journal of Inorganic Chemistry (Russ.)*, *Geochemistry*, *Journal of Chemical Physics (Russ.)*, *Journal of Organic Chemistry (Russ.)*, *Izvestiya VUZov ser. Oil and Gas*, *Transactions (Doklady) of Azerbaidjanian SSR Academy of Sciences*, *Bulletin of Azerbaidjanian Instititue of Oil and Chemistry*, *Izvestiya VUZov ser. Power Engineering*, *Izvestiya VUZov ser. Physics*, *Transactions (Doklady) of Academy of Sciences of USSR*, *Journal of Experimental and Theoretical Physics*, *Ukrainian Physical Journal*, *Chemistry and Technology of Water*, *Theoretical Foundations of Chemical Technologies*, *Journal of Applied and Technical Physics*, *Fuels*, and others).

The NIST/TRC Data Entry Facility data processing workstations are operated by students from the University of Colorado at Boulder and Colorado School of Mines in Golden, Colorado. The data processing operation is supported by two networks (a UNIX network provides access from every workstation to the SOURCE Data Archival System, and a Windows network supports communications between the workstations as well as between the workstations and character-recognition facility used to digitize relevant information from the hard-copy matter). All elements of data collection and processing at the NIST/TRC Data Entry Facility are supported by the Guided Data Capture software described above.

Data Communication Standard: IUPAC Standard for Thermophysical and Thermochemical Property Data Communications—ThermoML. ThermoML—an XML (X-Markup Language[18])-based IUPAC Standard for storage and exchange of experimental thermophysical and thermochemical property data—was previously fully described.[19,20] Supporting Information for the article[19] includes several examples illustrating the use of ThermoML to process experimental data for pure compounds, mixtures, and chemical reactions, as well as the initial ThermoML specification. ThermoML capitalizes on the fact that XML files are essentially textual files and can, in principle, be interpreted without customized software. ThermoML files also are assured to be interoperable between any variety of computer platforms and operation systems compliant with the standard ASCII (American Standard Code for Information Interchange[21]) code.

ThermoML covers essentially all experimentally determined thermodynamic and transport property data (more than 120 properties) for pure compounds, multicomponent mixtures, and chemical reactions (including change-of-state and equilibrium). Although the focus of ThermoML is properties determined by direct experimental measurement, ThermoML does cover key derived property data such as azeotropic properties, Henry's Law constants, virial coefficients (for pure compounds and mixtures), activities and activity coefficients, fugacities and fugacity coefficients, and standard properties derived from high-precision adiabatic heat-capacity calorimetry.

The ThermoML structure represents a balanced combination of hierarchical and relational elements. The ThermoML schema

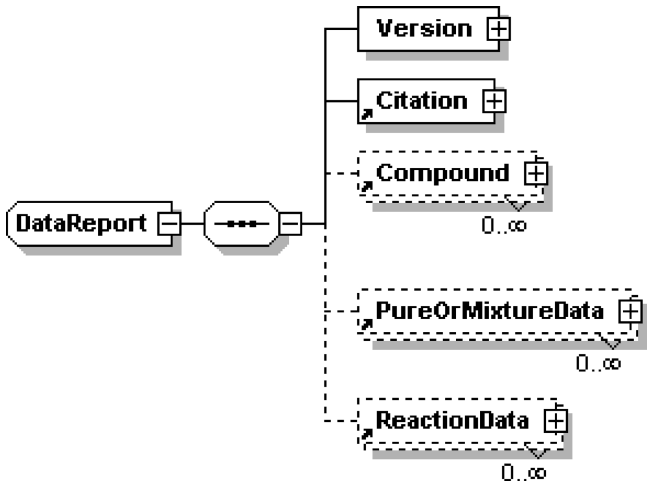

Figure 5. Major blocks of the ThermoML standard.[19]

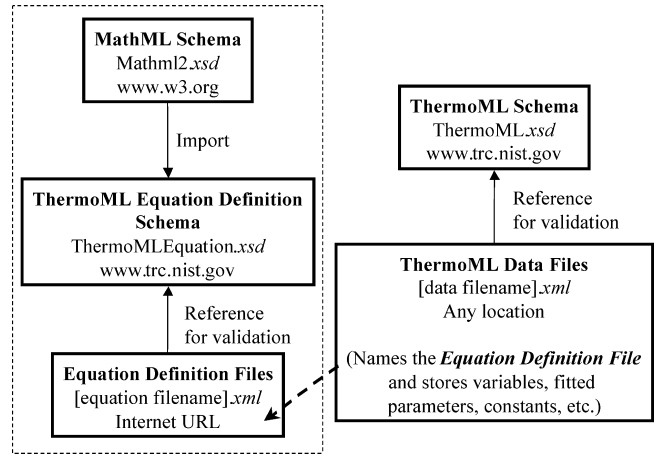

Figure 6. Components of equation representation with ThermoML.[23] The components within the dotted line are required to give meaning to the equation information stored in the ThermoML data files (shown in the lower right). Each component is listed with its file name and/or type ('xsd' identifies a schema and 'xml' identifies an xml file) and general location.

structure explicitly incorporates structural elements related to basic principles of phenomenological thermodynamics: thermochemical and thermophysical (equilibrium and transport) properties, state variables, system constraints, phases, and units. Meta- and numerical data records are grouped into "nested blocks" of information corresponding to data sets. The metadata records precede numerical data information, providing a robust foundation for generating "header" records for any relational database where ThermoML-formatted files could be incorporated. The structural features of the ThermoML metadata records ensure unambiguous interpretation of numerical data as well as data quality control based on the Gibbs phase rule.

Generally, all three major types of thermodynamic data[7]—experimental, predicted, and critically evaluated—are within the scope of ThermoML. The framework of ThermoML[22] and the extensions to the ThermoML schema for the expression of uncertainties[10] as well as extensions for representation of critically evaluated data, predicted data, and equation representation[23] have previously been described. ThermoML consists of four major blocks, as shown in Figure 5: *Citation*, describing the source of the data; *Compound*, characterizing the chemical system; *PureOrMixtureData*, providing information for meta- and numerical data for a pure compound or multicomponent mixture; and *ReactionData*, providing information for meta- and numerical data for a chemical reaction with thermodynamic state change or in a state of chemical equilibrium.

The ThermoML structure contains the elements necessary to store and exchange information related to fitting equations as well as associated covariance matrixes providing the measure of uncertainty for parameters of the equations. To accomplish this, the modular nature of the XML technology was taken advantage of in establishing communication between two different XML languages, ThermoML and MathML[24] (Figure 6). The ThermoML data file includes the identities of all variables, fitted parameters, and constants that are required for a particular equation representation but contains no mathematical expressions. One element of the ThermoML data file is a URL used to specify the Internet location (URL) of the full equation definition. The ThermoMLEquation schema is designed for storage and exchange of equation definitions with full mathematical content included through importation of the MathML schema. ThermoMLEquation is a general schema for the definition of any type of equation for representation of thermophysical and thermochemical properties. An equation definition file is created for definition of a particular equation. Care must be taken by the ThermoML file creator to ensure that the identities of the property, variables, constraints, and equation parameters and constants are correctly matched in the ThermoML data file and the ThermoMLEquation file.

Data "Reader" Software: NIST/TRC ThermoML Opener into Microsoft Excel. Although a variety of ThermoML software "readers" have now been developed by users to convert thermophysical and thermochemical property information from the standardized format to engineering application formats, we developed ThermoML Opener,[25] a software product that allows one to view ThermoML files in Microsoft Excel. Such a translation significantly eases visualization and interpretation of the information contained in the original ThermoML files. The example of the conversion of the ThermoML file containing property data reported in ref 26 to Microsoft Excel is provided in Figure 7. ThermoML Opener uses XSLT (eXtensible Stylesheet Language for Transformations) technology[27] to translate a ThermoML file into an HTML (HyperText Markup Language)[18] file suitable for opening by Excel. The algorithm of the ThermoML Opener implemented as Visual Basic software program includes a sequence of the commands encompassing opening of the ThermoML file chosen by the user, the XSLT translation, opening of the Excel, and imports of the translated ThermoML file. ThermoML Opener is a component of the software infrastructure designed to support the ThermoML standard.

Software Expert System: NIST ThermoData Engine. NIST ThermoData Engine (TDE)[28−32] is the first product fully implementing all major principles of the concept of dynamic data evaluation formulated at NIST/TRC.[7,33,34] This concept requires the development of large electronic databases capable of storing essentially all "raw" experimental data known to date with detailed descriptions of relevant metadata and uncertainties. The combination of these databases with expert software designed to generate recommended data based on available raw experimental data and their uncertainties leads to the possibility of producing data compilations automatically "to order" forming a dynamic data infrastructure (Figure 8). This concept contrasts sharply with static critical data evaluation, deployed essentially in all critical evaluation projects for the last 200 years and that must be initiated far in advance of need. The dynamic data evaluation process dramatically reduces the effort and costs associated with anticipating future needs and keeping static

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="http://www-i.boulder.nist.gov/div838/trc/journals/jced/ThermoMLtoHTML.xsl" type="text/xsl"?>
<!--Created by Guided Data Capture (GDC) Software (http://www.trc.nist.gov)-->
<DataReport xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://www.trc.nist.gov/ThermoML.xsd">
<Version>
<nVersionMajor>2</nVersionMajor>
<nVersionMinor>0</nVersionMinor>
</Version>
<Citation>
<TRCRefID>
<yrYrPub>2005</yrYrPub>
<sAuthor1>ash</sAuthor1>
<sAuthor2></sAuthor2>
<nAuthorn>0</nAuthorn>
</TRCRefID>
<eType>journal</eType>
<eSourceType>Original</eSourceType>
<sAuthor>Ashour, I.[Ibrahim]</sAuthor>
<sPubName>J. Chem. Eng. Data</sPubName>
<yrPubYr>2005</yrPubYr>
<dateCit>2005-02-10</dateCit>
<sTitle>Liquid-Liquid Equilibrium of MTBE + Ethanol + Water and MTBE + 1-Hexanol + Water over the Temperature Range of 288.15 to 308.15 K</sTitle>
<sAbstract>Liquid-liquid equilibrium data for methyl tert-butyl ether (MTBE) + ethanol + water and MTBE + 1-hexanol + water have been experimentally measured over the temperature range of 288.15 to 308.15 K. The equilibrium data of this work, in addition to the available MTBE-containing LLE data in the technical literature, are analyzed using UNIQUAC, NRTL, UNIFAC-LL, and UNIFAC-DMD models as programmed by the Aspen Plus simulator. On the basis of analyses of the experimental data of this work, UNIFAC-LLE and UNIFAC-DMD showed the best predictive performance for the mole fraction of the target species (MTBE) in both the organic and aqueous phases.</sAbstract>
<sKeyword>Liquid-Liquid Equilibrium</sKeyword>
<sKeyword>MTBE</sKeyword>
<sKeyword>Ethanol</sKeyword>

A

Liquid-Liquid Equilibrium of MTBE + Ethanol + Water and MTBE + 1-Hexanol + Water over the Temperature Range of 288.15 to 308.15 K

I. Ashour

J. Chem. Eng. Data 50 113-118

Compounds studied in this report		
Name	Molecular Formula	Registry Number
methyl tert-butyl ether	C5H12O	1634044
ethanol	C2H6O	64175
1-hexanol	C6H14O	111273
water	H2O	7732185

Result Set #1

Measurement of:
Specific density, kg/m3
Method: Pycnometric method
Phase: Liquid mixture 1

For compounds:
ethanol (C2H6O)
methyl tert-butyl ether (C5H12O)
water (H2O)

With the constraint:
Pressure, kPa = 101.3

Temperature, K	methyl tert-butyl ether Mole fraction	Specific density, kg/m3
288.15	.9217	746

B

Figure 7. Example of the conversion of the ThermoML file containing the property data reported in ref 26 to Microsoft Excel. A, a fragment of the original ThermoML file; B, a fragment of the resulting Microsoft Excel file.

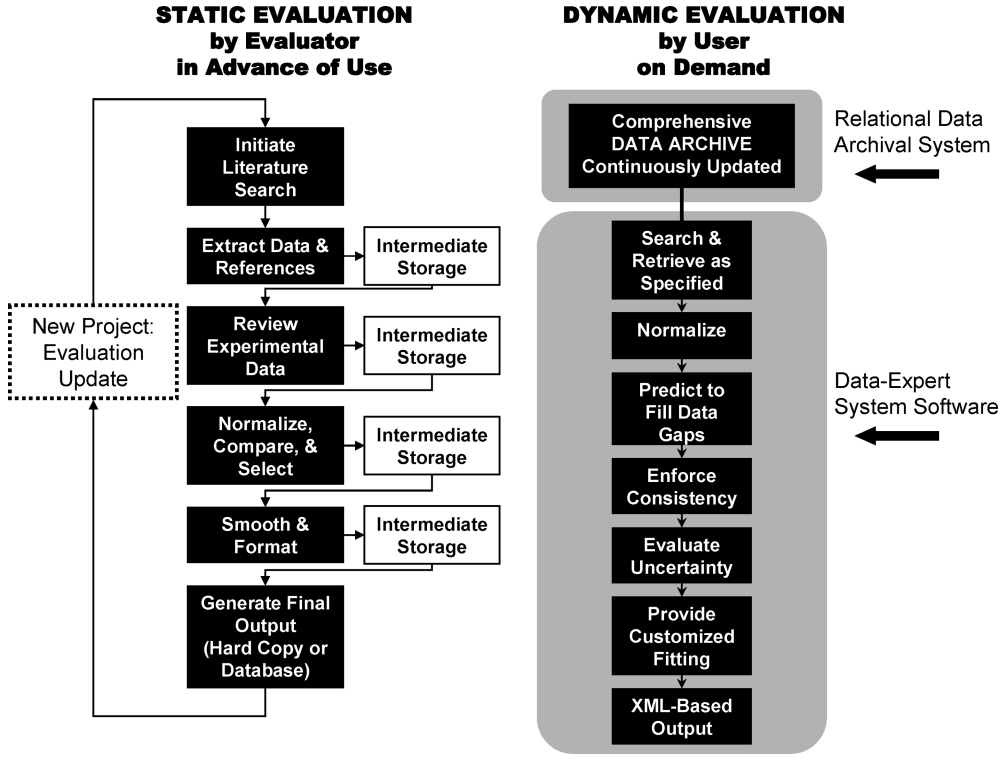

Figure 8. Schematic representation of the static and dynamic data evaluation concepts.

evaluations current. Indeed, it eliminates a necessity of intermediate information storage between every preceding and following steps of the process typical for the static data evaluation as well as reinitiation of the critical evaluation updates to be performed by the evaluator, saving an enormous amount of time and effort. The dynamic data evaluation concept brings the user to the center of the process, providing an opportunity to perform critical data evaluation in an "on-demand" mode.

Implementation of the dynamic data evaluation concept consisting, in principle, of the development of the comprehensive data archival system and data-expert system (Figure 8) as well as the infrastructure for their management and maintenance, is very challenging. In the development of TDE, SOURCE, described here, was used to generate a local version of the comprehensive data archival system, TDE-SOURCE. The first version of the TDE, released in 2004, was limited to property-

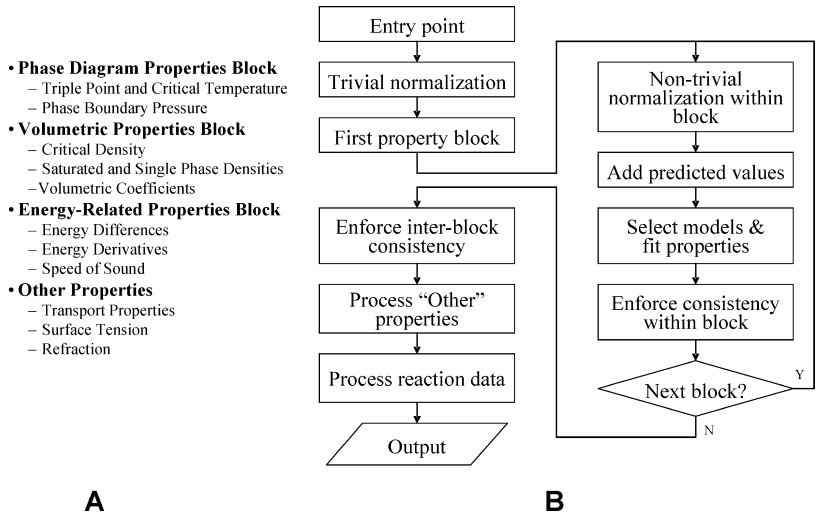

Figure 9. Property blocks (A) and principal algorithm of the NIST ThermoData Engine (B) limited to pure compounds.

by-property critical evaluation for pure compounds (more than 50 thermophysical properties).[30] The developed algorithm incorporated all major stages of the concept implementation, including data retrieval, grouping, normalization, sorting, consistency enforcement, fitting, and prediction for all major thermophysical properties. TDE fills the gaps in experimental data space by deployment of automated group-contribution and corresponding-states property predictions, emphasizes enforcement of consistency between related properties (including those obtained from predictions), provides for flexibility in selection of default data models depending on the particular data scenario, incorporates a large variety of models for secondary fitting, evaluates the uncertainties (covariance matrixes for the equation parameters), and allows saving of critically evaluated data in ThermoML format (Figure 9, B). All thermophysical properties are divided into four property blocks (phase-diagram properties, volumetric properties, energy-related properties, and other properties such as transport properties, surface tension, and refraction, Figure 9, A) and are processed block-by-block with the enforcement of the consistencies between related properties first within the blocks and then between properties belonging to different blocks.

In the second version of TDE, released in 2006, the implementation of dynamic data evaluation was further expanded for pure compounds to generation of equations of state on demand.[31] The four principal equations of state (original and modified volume-translated Peng−Robinson, Sanchez−Lacombe, PC-SAFT, and Span−Wagner) were included to ensure adequate experimental and predicted data fitting depending on various "data scenarios". Periodical Web updates of the local TDE-SOURCE database maintain its up-to-date status by use of the Web-Oracle infrastructure described in the present paper, providing new data to users of TDE soon after original publication in the literature. In 2008, the capabilities of the first two versions of the TDE were combined within the NIST Standard Reference Database 103a.[28]

The operational model deployed by the NIST ThermoData Engine for critical data evaluation of the pure compound properties is shown on Figure 10. The example of simultaneous generation of the variety of thermophysical properties as functions of temperature by the deployment of the Span−Wagner equation of state for pyridine within the NIST ThermoData Engine is illustrated in Figure 11.

The third version of TDE, released in 2008, expanded the implementation of the concept to the thermophysical properties

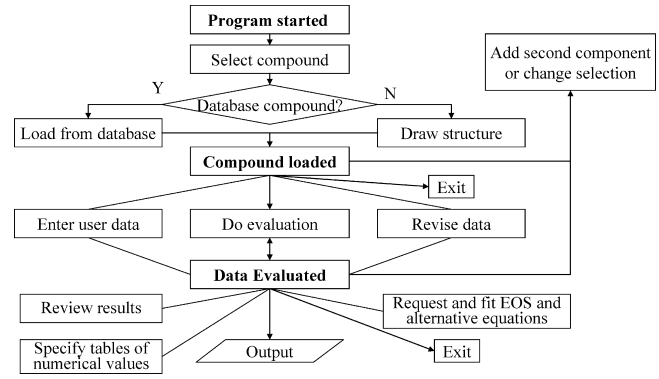

Figure 10. Operational model deployed by the NIST ThermoData Engine for critical data evaluation of pure compound properties.

of binary mixtures.[29,32] TDE provides access to single-phase thermodynamic and transport property data and VLE, LLE, and SLE data for more than 30 000 mixtures and does automated evaluation of most of those properties. Certain properties such as densities and critical and transport properties are described by special fitting equations; phase equilibria data are described by activity coefficient models selected by the user from the set of supported models: Margules, NRTL, Redlich−Kister, UNIQUAC, van Laar, and Wilson. UNIFAC predictions are generated for mixtures covered by the UNIFAC method, including those for which experimental data are currently not available. Phase diagrams, isotherms, and isobars based on those models can be calculated and drawn for the user's convenience. Proprietary data can be entered for inclusion in the evaluation, and the user can influence the evaluation process by changing relative data weights or by rejecting particular data sets. Properties, models (A), and principal algorithm (B) deployed by the TDE for thermophysical properties of binary mixtures are shown in Figure 12. The operational model of TDE for binary mixtures is presented in Figure 13. The example of critical data evaluation of the vapor−liquid equilibria performed by TDE for the ethanol + 1-bromobutane mixture is illustrated in Figure 14.

Future expansions of TDE will cover thermochemical properties of chemical reactions and thermophysical properties of ternary mixtures.

Web Communication Portal: NIST/TRC Web-Oracle Infrastructure. To make the critical data evaluation process fully dynamic, the NIST/TRC Web-Oracle infrastructure for continu-

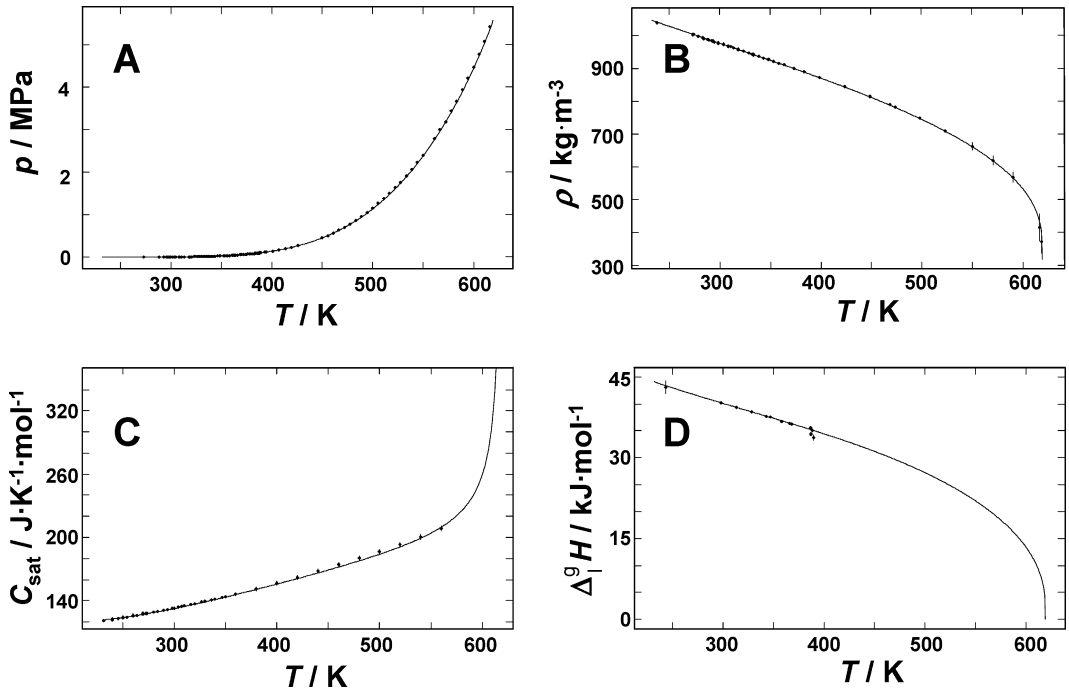

Figure 11. Simultaneous generation of a variety of thermophysical properties by the deployment of the Span−Wagner equation of state for pyridine within the NIST ThermoData Engine (A, p, vapor pressure; B, ρ, density; C, c_{sat}, heat capacity at the saturation line; D, $\Delta_l^g H$, enthalpy of vaporization). All properties presented as functions of temperature.

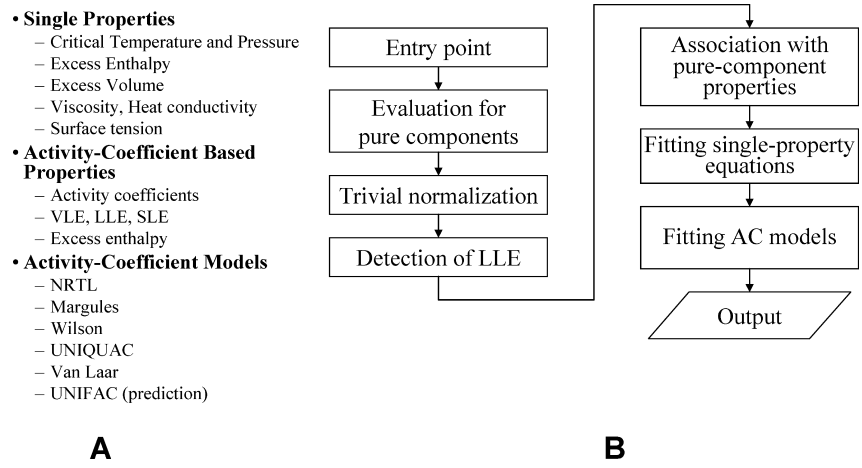

Figure 12. Properties, models (A), and principal algorithm (B) deployed by the NIST ThermoData Engine for thermophysical properties of binary mixtures.

ous dissemination of recently entered experimental thermophysical and thermochemical data stored in the SOURCE Data Archival System has been developed. This infrastructure 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 many engineering applications. A local database of differences (TDE UPDATES) is generated for this purpose, allowing each instance of TDE to set or restore the working database (TDE-SOURCE) to any of the available discrete states rather than maintaining traceability information for every individual data point. The process by which TDE checks for updates and maintains current TDE-SOURCE information is described in detail elsewhere.[31]

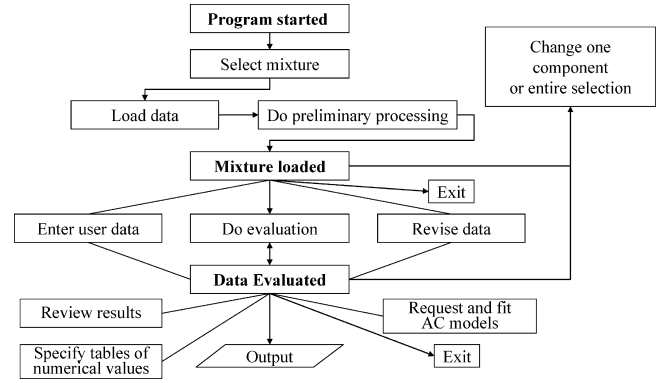

Figure 13. Operational model deployed by the NIST ThermoData Engine for critical data evaluation of the binary mixture properties.

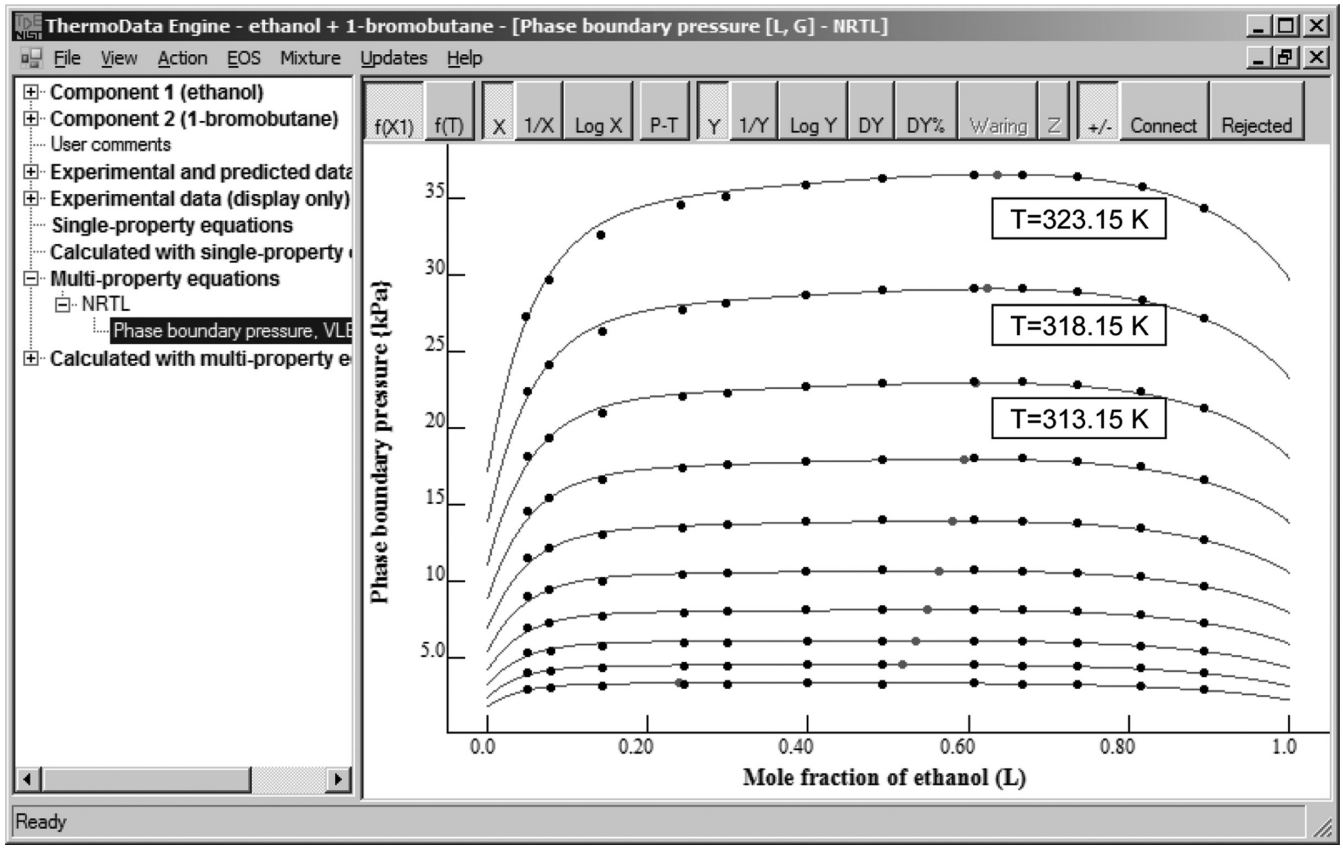

Figure 14. Critically evaluated by the NIST ThermoData Engine, vapor−liquid equilibria (pressure versus mole fraction of ethanol at various temperatures) data for the ethanol + 1-bromobutane mixture. The deployed activity coefficient model is NRTL.

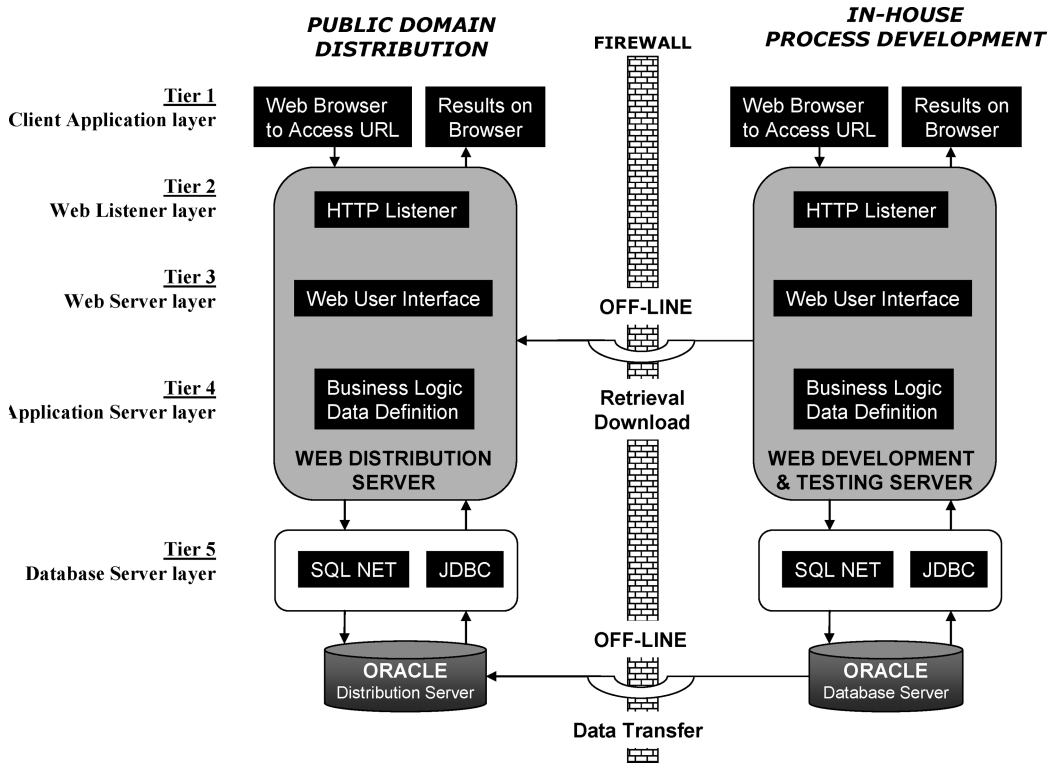

Figure 15. NIST/TRC Web-Oracle infrastructure schema.

The NIST/TRC Web-Oracle infrastructure developed is based on a multitier architecture[35] containing five layers and consists of two identical components (one behind the firewall and one outside the firewall (Figure 15)) to ensure robust security for communications. The data and retrieval codes are communicated from the testing component (behind the firewall) to the disseminating component (outside the firewall) offline. The five layers of infrastructure include:

- Tier 1: Client Application layer
- Tier 2: Web Listener layer (Client Controlled)

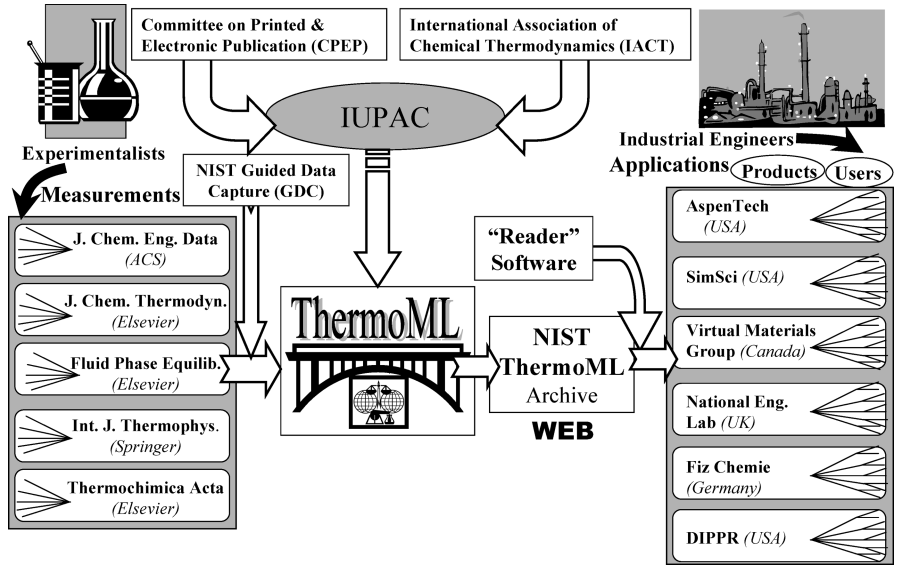

Figure 16. Schematic representation of the global data delivery process in thermodynamics. Currently, there is a significant number of the organizations that have already developed ThermoML "readers". The organizations identified here are examples since these organizations approached TRC for advice during the development of their ThermoML readers. These organizations are cited only in the interest of complete technical description, and neither constitute nor imply endorsement by NIST or by the US government.

- Tier 3: Web Server layer
- Tier 4: Application Server layer
- Tier 5: Database Server layer

Communication is initiated through a request generated by the end-user at the Client Application layer. This request is sent over the network to one of several ports on the NIST/TRC Web-hosting server depending on the protocol used (such as File Transfer Protocol (FTP) or Hypertext Transfer Protocol (HTTP)[36]). A Web Listener layer processes incoming network communication and forward data requests to the Web Server layer. This software then accesses an application programming interface (API) contained within the Application Server layer. The API functions to reformulate the data request into a database query, which is submitted to the Database Server layer. The Database Server layer then searches the tables of the SOURCE Data Archival System to collect the appropriate information. The Database Server layer then returns this resulting set to the Application Server layer. The functions defined within the Application Server layer process this input and use it to populate a data file of appropriate format, such as the format associated with the HyperText Markup Language (HTML). This file is then translated by the Web Server layer into the appropriate response format, usually HTTP, and delivered via the network to the Client Application layer on the end-user's machine, which translates the incoming document into the appropriate visual representation.

This infrastructure is used not only to disseminate experimental data updates to support TDE but also to provide access to other NIST/TRC databases such as the Ionic Liquids Database, ILThermo,[37,38] and Web Thermo Tables (WTT).[39,40]

Free-access, open-domain dissemination of the ThermoML-formatted data files containing thermophysical and thermochemical property data obtained via cooperation with five major journals in the field is supported by the NIST/TRC Web Archive.[41]

3. Impact of the Implementation of the ThermoGlobe on Principal Application Areas

Below we shall illustrate examples of the impact of the implementation of the Global Information System in Science concept for the field of thermodynamics on various principal application areas.

Impact Area: Efficiency of Information Delivery. It is quite obvious that the efficiency of delivery of thermophysical and thermochemical property information from "data producers" to "data users" plays a critical role in an effective use of this information in a broad range of applications from the scientific discovery process to chemical process and product design. While most of the scientific journals publishing this information provide access to their articles for the readers over the Web, scientists and engineers worldwide who are dependent on this information cannot take full advantage of this opportunity because it is very difficult to interpret the property data as automated large-scale computer-to-computer communication from the PDF (Portable Document Format) or HTML formats used by the journals and publishers online. As a result, new experimental thermophysical and thermochemical property data reported in the scientific literature are "migrated" to the engineering application software and databases over a very long period, of months or, in many instances, years. Intellectual and economic consequences of this delay are enormous.

This problem has to a significant degree been addressed with the use of the ThermoGlobe components in establishing the first global data delivery process in the field of thermodynamics.[42] Five components of the ThermoGlobe (GDC, NIST/TRC Data Entry Facility, ThermoML, ThermoML Opener, Web-Oracle infrastructure) are used to create the delivery infrastructure for this process (Figure 16), while the other two components (SOURCE and ThermoData Engine) are instrumental for data quality control (see next section Impact Area: Journal Publication Quality). Five major journals in the field (*Journal of Chemical and Engineering Data*, *Journal of Chemical Thermodynamics*, *Fluid Phase Equilibria*, *Thermochimica Acta*, and *International Journal of Thermophysics*) have endorsed this process[43−47] and encouraged their authors to be immediately involved in it. In this process (Figure 16), the authors of the articles submitted to the participating journals download GDC software freely available from the NIST/TRC Web site,[5] as a part of the overall article submission procedure. GDC is then used by the authors to generate GDC data files containing

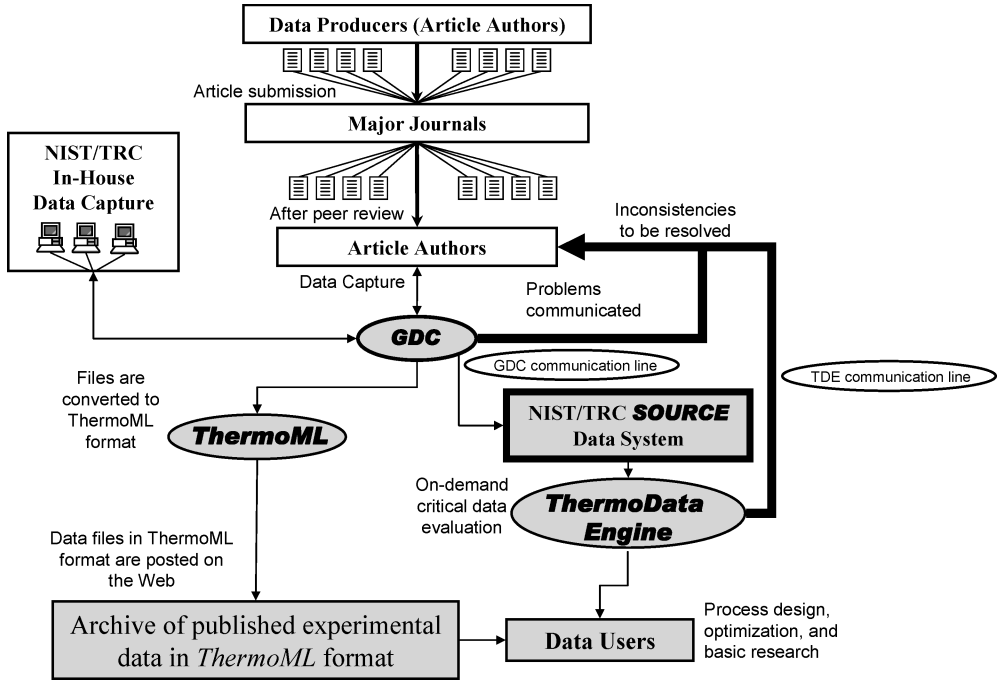

Figure 17. Data flow diagram at the NIST/TRC Data Entry Facility.[42]

experimental thermophysical and thermochemical property information pertaining to the subject articles. These files are sent over e-mail to the NIST/TRC Data Entry Facility, processed, and after analysis for internal thermodynamic consistency, converted to the ThermoML format by use of ThermoML schema,[48] as the IUPAC standard for thermodynamic data communication. This standard has been approved by the IUPAC Committee on Printed and Electronic Publications (CPEP)[49] and endorsed by the International Association of Chemical Thermodynamics (IACT).[50] These ThermoML-formatted data files are posted within the ThermoML Web archive,[41] keeping the file structure similar to that for PDF and HTML files of the articles of the participating journals. Scientists and engineers worldwide have unlimited capabilities to download the data files from the ThermoML Web Archive. The downloaded ThermoML-formatted data files can then be converted to a format suitable for a particular software or database engineering application by use of either the ThermoML Opener or other specifically designed ThermoML "reader" software. This data delivery process significantly shortens the period of information propagation from experimentalists "producing" new thermophysical and thermochemical property data to scientists and engineers using these data in a broad range of applications. These applications include, but are obviously not limited to, chemical process design (see section Impact Area: Chemical Process Design), development of new property modeling and prediction methods,[51−53] and database production.[54]

While the process illustrated in Figure 16 emphasizes "horizontal" data propagation from "data producers" to "data users", similarly, a number of "vertical" data communication processes have been implemented, providing efficient data propagation between various experimentalists, journals, organizations, products, and users.

Impact Area: Journal Publication Quality. The dramatic increase in the volume of reported technical information within the last two decades creates challenges in establishing data quality protocols capable of preventing erroneous data from coming to the public domain with a "stamp" as peer-reviewed and being considered therefore as highly reliable. This situation is, in particular, true for published thermophysical and thermochemical property data. Indeed, while the total number of experimental data points published continues to grow significantly (Figure 1), the nature of the peer-review process has not been changed to any significant degree over the last 50 years. That mismatch has in many instances made it virtually impossible to detect, during the peer-review process, mutual thermodynamic inconsistencies of the data soon-to-be-published or even violations of the fundamental laws of nature. On the other hand, new technologies for data communication including the Internet are extremely efficient in propagating both correct and erroneous information. This "efficiency" in propagating erroneous information is further aggravated by the subsequent use of it in the development of new property models and, frequently, in designing new industrial applications. Accumulation of the total amount of "error" is happening on many stages of the scientific data production and communication, from original measurement to uncertainty analysis, to data recording and processing, to the preparation of the article to be published, to the article transformation during the publication process.[10,55]

The implementation of the ThermoGlobe as the Global Information System in Thermodynamics provides unique opportunities for curing this problem in a systematic and structural manner. The data flow diagram[42] shown in Figure 17 illustrates the two distinct components of the data processing at the NIST/TRC Data Entry Facility. One of them is associated with the in-house data processing, and the other one results from cooperation with the major journals in the field. In principle, the new data to be published undergo two-stage testing of integrity and consistency. The first stage, performed with the use of Guided Data Capture software, targets detection of any internal data inconsistencies, while the second stage is designed for consistency analysis between the new data and the "entire body of knowledge" by use of the ThermoData Engine. Each stage of data testing corresponds to a particular communication channel with authors and the editors (Figure 17), developed to communicate potential inconsistencies and to further resolve them. Below we illustrate this process with several examples.

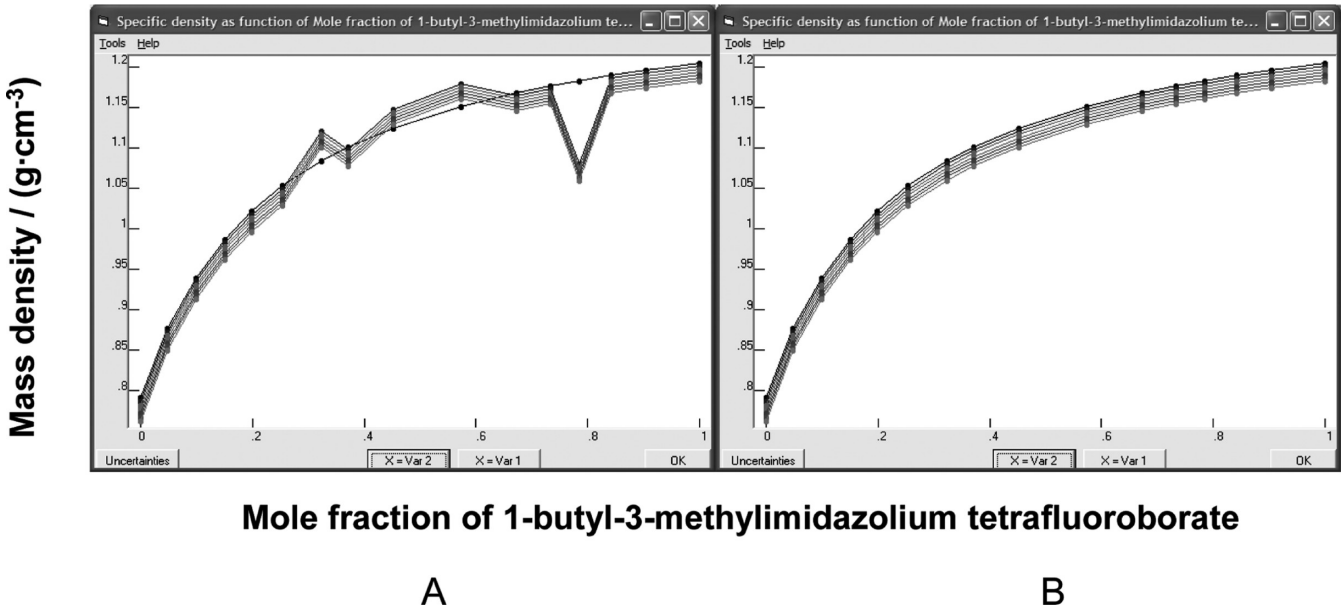

Figure 18. Mass (specific) density as a function of mole fraction for mixtures of 1-butyl-3-methylimidazolium tetrafluoroborate and methanol. The lines represent results for various temperatures. The figures show the reported experimental data before (A)[56] and after (B)[57] correction. The graphs were created with the GDC software.[6]

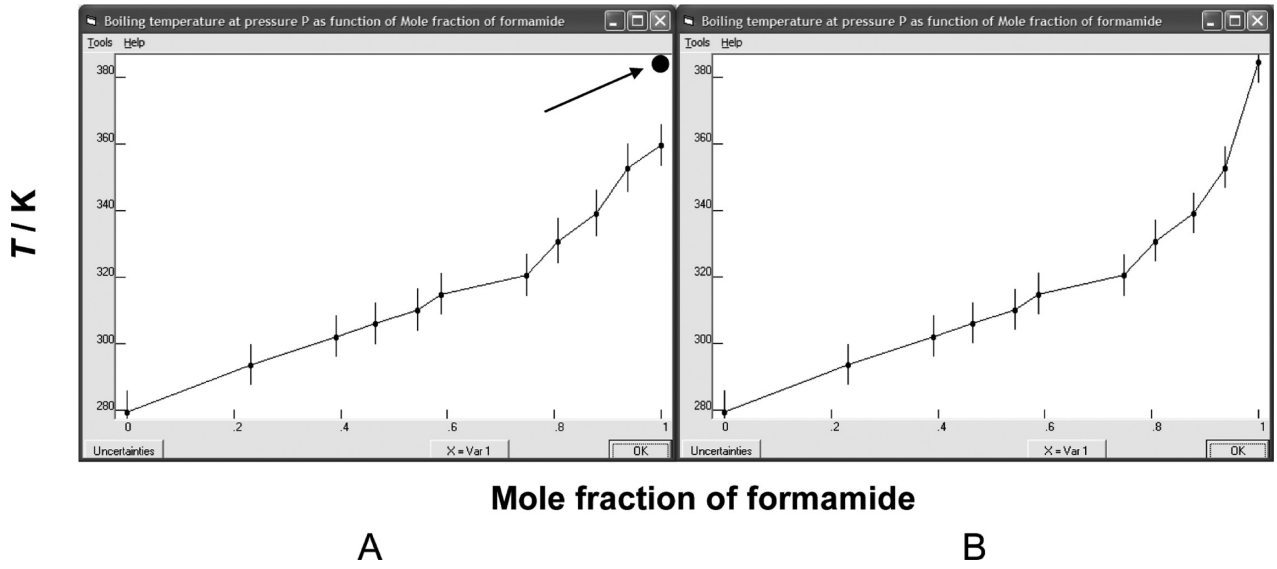

Figure 19. Boiling temperatures T at constant pressure ($p = 2.5$ kPa) for (formamide + water). The figures show the reported experimental data before (A)[58] and after (B)[59] correction. The arrow indicates the boiling temperature for pure formamide critically evaluated with the ThermoData Engine (TDE) software: $T = (385 \pm 2)$ K at $p = 2.5$ kPa for formamide. The graphs were created with the GDC software.[6]

The first example (Figure 18) constitutes a resolution of internally inconsistent data by use of the GDC communication channel (Figure 17). Originally published data for mass density of the mixture of the ionic liquid 1-butyl-3-methylimidazolium tetrafluoroborate and methanol as a function of the mixture composition (mole fraction of 1-butyl-3-methylimidazolium tetrafluoroborate) at various temperatures[56] when visualized with the GDC graphical representation tools indicate nonmonotonous behavior near 0.8 mol fraction (Figure 18, A). After this inconsistency had been communicated, the authors corrected the data, eliminating the anomaly[57] (Figure 18, B).

The second example illustrates the use of both GDC and TDE communication channels (Figure 17) simultaneously. Originally published data for the boiling temperature of the mixture (formamide + water) at constant pressure (2.5 kPa) as a function of the composition (mole fraction of formamide)[58] when visualized with the GDC graphical representation tools indicate somewhat nonmonotonous behavior (Figure 19, A). In addition, the boiling point of the pure formamide (the end point of the curve) is far beyond the sum of the uncertainties, inconsistent with the critically evaluated value generated by the TDE ((385 ± 2) K). After these findings were communicated, the authors corrected their data[59] (Figure 19, B), eliminating the inconsistency.

The third example[42] illustrates the use of the TDE communication channel (Figure 17). Figure 20 illustrates the TDE analysis of published experimental density data in the liquid phase for tetrahydrofuran as a function of temperature and pressure. Preliminary GDC analysis did not detect any anomalies in the originally published data[60] (Figure 20, A). However, comparison of the published values with those along the saturation line obtained with on-demand critical data evaluation by TDE showed that the reported single-phase data crossed the saturation line into the two-phase region, which is not physically possible. The originally published experimental data were

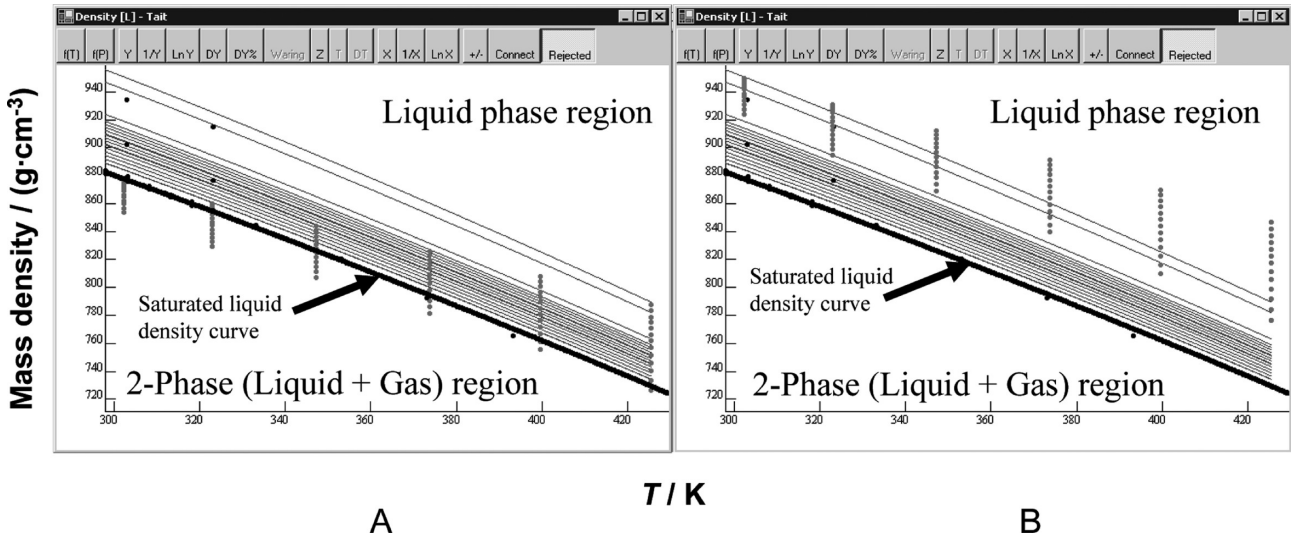

Figure 20. Experimental density data (dots) for tetrahydrofuran are reported as a function of temperature and pressure in the liquid phase before (A)[60] and after (B)[61] correction. The saturation line (i.e., the line of separation between the one- and two-phase regions) critically evaluated with TDE is shown in the plot. The lines above the saturation density curve correspond to the isobars evaluated by TDE for the single-phase region. The graphs were created with the TDE software.[6,42]

corrected[61] (Figure 20, B) after the results of the TDE analysis were reported to the authors. These data now appear correctly within the single-phase liquid region.

On the basis of two years of experience with the new data delivery process, we estimate that 10 % of articles reporting experimental thermodynamic data for organic compounds contain some erroneous information (the analysis is based on approximately 1000 articles).[42] 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.

Taking these findings into account, the five major journals in the field of thermodynamics (*Journals of Chemical and Engineering Data*, *Journal of Chemical Thermodynamics*, *Fluid Phase Equilibria*, *Thermochimica Acta*, and *International Journal of Thermophysics*) have agreed to establish the ThermoGlobe data quality protocols as a required component of their overall article submission process effective January, 2009.

Impact Area: Chemical Process Design. Recent progress in laboratory synthesis of numerous specialty chemicals, drug and biofuel development, as well as the increasing cost of energy and environmental compliance have created new challenges for chemical process design. These challenges are numerous and multifaceted. One of the most difficult and principal challenges is to create software capabilities to support chemical process design for products of a given chemical structure or, in other words, to create new technologies for on-demand product design. Development of such technologies was impossible for many years due to deployment of the established "static" protocols for generating critically evaluated thermophysical and thermochemical property data "feeding" process simulators.

The ThermoData Engine (TDE), one of the ThermoGlobe components, implementing the concept of dynamic data evaluation for thermophysical and thermochemical property data, provides for the first time an opportunity for the development of a new generation of simulator software capable of designing chemical processes and products on-demand.[62]

One of the principal ways of communicating data from TDE to the process simulator is illustrated in Figure 21.[7] As shown in Figure 21, this communication is based on ThermoML data file processing, while the local TDE-SOURCE database on the

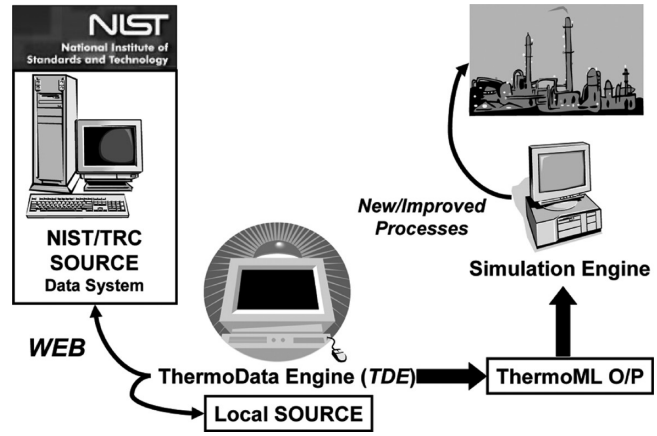

Figure 21. Schematic representation of the communication between the ThermoData Engine and commercial simulators on the basis of ThermoML-formatted file input−output.

user end is continuously updated from the NIST/TRC central server over the Web by use of the NIST Web-ORACLE communication infrastructure. Such a process can be organized as sequential running of the TDE and the simulation engine on the selected compound/mixture basis. Thus, the ThermoML-formatted TDE data output file can be interpreted and processed by the designed ThermoML "reader" incorporated within the simulation engine. The other option is to run TDE in the "batch" mode, generating a ThermoML output file of the parameters of the property models for a large group of compounds/mixtures (for example, for all the compounds characterized with reliable experimental data). This file can further be interpreted by the designed ThermoML "reader", and the data can be propagated into a customized database within the simulation engine (Figure 22). The simulation engine then provides options to the user to query the database for a particular compound/mixture and use the generated data to support the chemical process/product design activities. If such a process is deployed for all pure compounds characterized with reliable experimental data, the simulation engine is capable of supporting chemical process design projects encompassing about 20 000 compounds.

Full advantage of the dynamic nature of the TDE in application to chemical process design is achieved by the direct

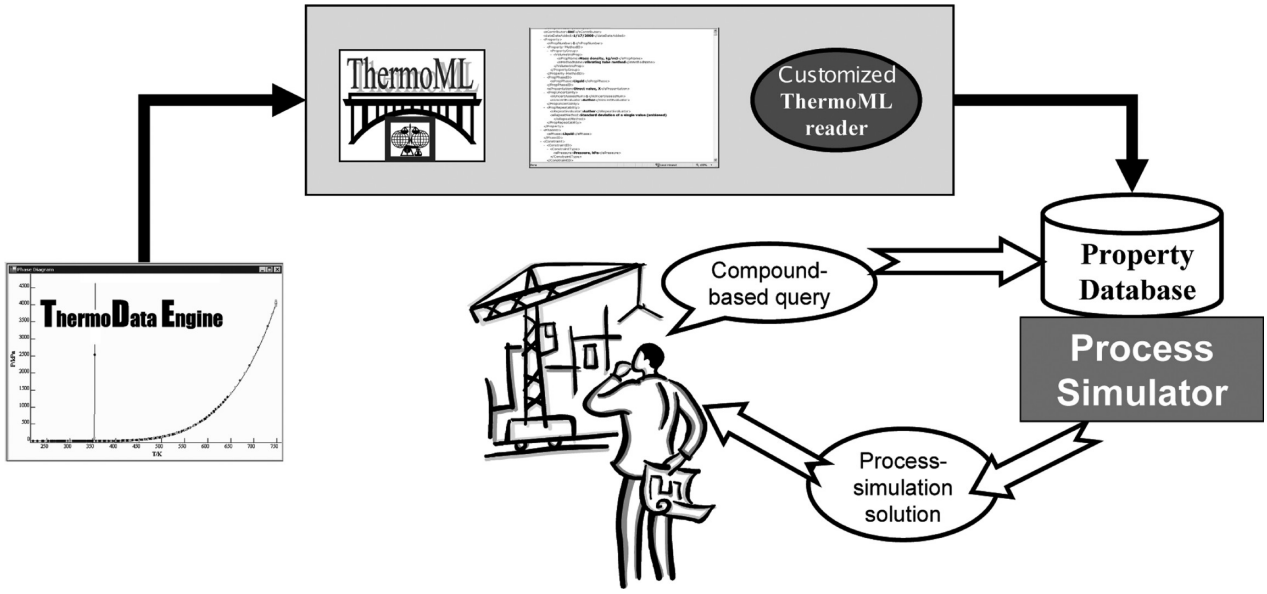

Figure 22. Process simulator access to the ThermoData Engine via the database of the property equation parameters.

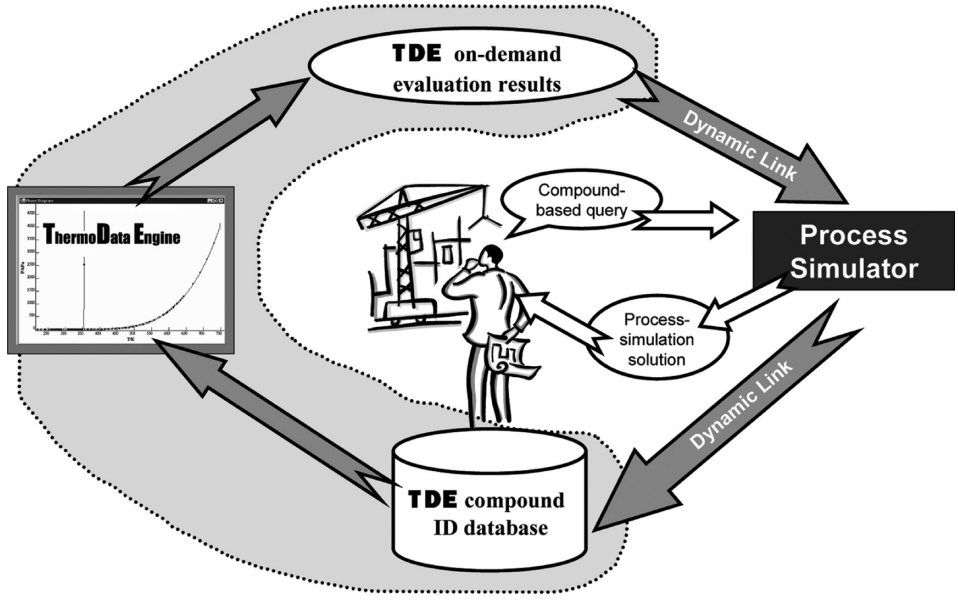

Figure 23. Process simulator access to the ThermoData Engine via a dynamic link based on compound identifier indexes.

"bundling" of the TDE and commercial process simulators using Dynamic-Link Library (DLL) technology.[63] Application of this technology makes it possible to use a graphic interface developed as a part of the process simulator to be populated with data generated by TDE. Such a "bundling" of TDE and the process simulator truly places the user at the focal point of the process (Figures 23 and 24). Indeed, the user generates a compound-based query depending on the nature of the chemical process to be designed or modified. This query is communicated from the process simulator to the compound identifier database of TDE over the dynamic link (Figure 23). The compound identifier is then used by TDE to generate all critically evaluated data for the requested system and return the results to the process simulator via another dynamic link. The process simulator then propagates the generated property data to support chemical process/product activities. Such an implementation of the TDE technology makes it possible to support chemical process design projects for more than 100 000 pure compounds. Further enhancement of the DLL bundling by use of chemical structures results in additional and rather dramatic capabilities related to chemical process design, making it possible to analyze processes encompassing millions of pure compounds (Figure 24). Various options for communication between TDE and commercial simulators described above are currently deployed by a variety of flagship chemical process design software products and used by tens of thousands of customers worldwide, leading in essence to a pronounced change in the very nature of chemical process design.

In the future, this impact might be even more significant, when TDE capabilities in the assessment of the combined expanded uncertainties for thermophysical properties will be propagated into the uncertainties of the reaction streams and used as a factor in selecting appropriate equipment elements for technological schemas.

Other Impact Areas. In this section, we will outline various aspects of the ThermoGlobe impact on such areas of research activities as strategic experiment planning, molecular modeling and property prediction, the scientific discovery process, and instrument calibration and validation.

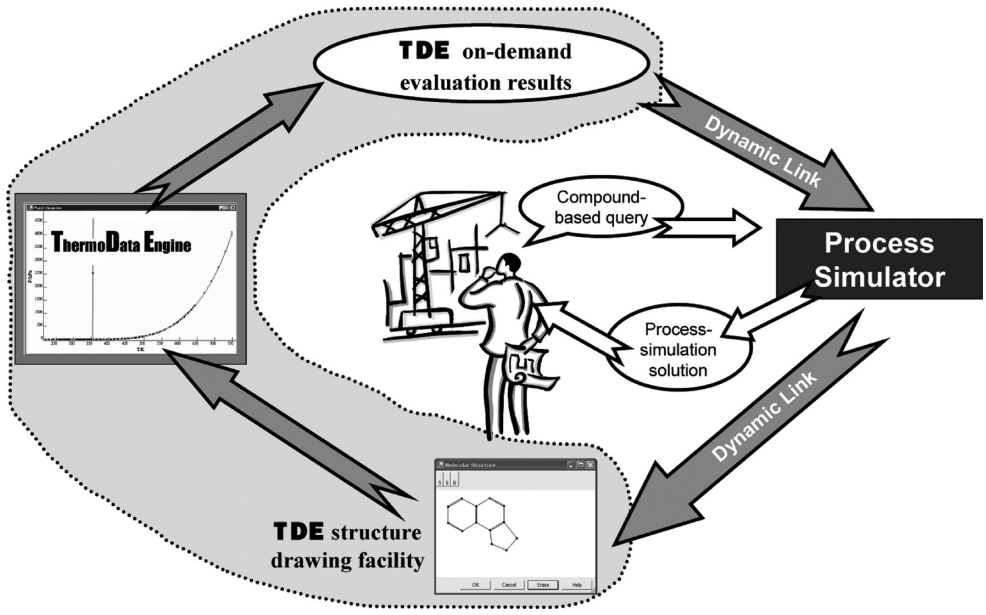

Figure 24. Process simulator access to the ThermoData Engine via a dynamic link based on compound structure.

Wakeham et al.[64] envisioned that the future of experimental thermophysical property measurement science will represent a transformation "from accuracy to fitness for purpose". They indeed emphasized that, in a great many instances, the sole desire for high accuracy of 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 further advancement of theory and simulation. Therefore, one can legitimately conclude that the effort and resources associated with these measurements were, to a significant degree, inefficiently used. The validity of this point, in our view, is beyond doubt. However while it is true that frequently this was caused by "a tendency to validate the instrument by first proving its operation using well-known properties",[64] we believe that, more often than not, the fundamental reason for such inefficiency lays in a complexity of the analysis similar to what is defined in economics as "significant return on investment". Such an analysis, which we would call here *strategic experiment planning*, is impossible now without an algorithmic approach, including assessment of the entire body of knowledge, availability of experimental thermophysical property data, the variable ranges studied, the associated uncertainties, the state of the prediction methods, availability of the parameters for deployment of the prediction methods, and how these parameters can be obtained using targeted measurements, etc., and, indeed, how the intended measurement can address the underlying scientific or engineering problem under consideration. Since the very definition of the Global Information System in Science implies operation over the entire body of knowledge in the field, we intend to use ThermoGlobe to develop a software product for *strategic experiment planning* for thermophysical properties. We hope that such a software product, when made available to the scientific and engineering communities, can serve as a guide for experimentalists in choosing the systems and properties to be measured as well as appropriate conditions of these measurements such as variable ranges and thermodynamic constraints.

Development of high-quality *molecular modeling and property prediction* methods is impossible without basis on an extensive set of experimental or critically evaluated data of defined quality. ThermoGlobe provides a foundation for the development of group contribution,[65] corresponding states,[66] and high-level prediction methods such as molecular simulation[67] and ab initio-based Quantitative Structure−Property Relationship (QSPR) techniques.[68] It also provides guidance in establishing the limits of applicability of a variety of property prediction methods toward particular classes of chemical compounds and their mixtures. In some instances, ThermoGlobe deployment makes it possible to determine inconsistencies of the experimental data used as key parameters for the development of prediction methods and to communicate a necessity for new measurements, leading to a potential resolution of the determined inconsistencies.[69,70]

It is clear that Global Information Systems in Science have tremendous potential in supporting the *scientific discovery process*. In application to ThermoGlobe that constitutes supporting the process of establishing relationships between various thermophysical properties, as well as relationships between thermophysical and other properties (magnetic, electrical, mechanical, etc.). It is, in particular, promising to search for such relationships between various thermophysical and physiological properties. Establishing such relationships might be very beneficial for progress in the pharmaceutical industry in further enhancing the processes of drug development.

Finally, similar to being "bundled" with various chemical process design software products, the ThermoGlobe components can be bundled with software supporting commercial instruments for thermophysical property measurements, serving as an automated tool for ongoing *instrument calibration and validation*.

Conclusions

1. The concept of the Global Information System in Science is defined.

2. The major components of the Global Information System in the field of Thermodynamics, ThermoGlobe, have been identified and developed.

3. A detailed analysis of the impact of the ThermoGlobe on efficiency of information delivery, journal publication quality, and chemical process design is given.

4. The potential of impact of ThermoGlobe on other areas such as strategic measurement planning, molecular modeling

and property prediction, scientific discovery process, and instrument calibration and validation is outlined.

5. Development of ThermoGlobe has facilitated development of Global Information Systems in other scientific fields outside the field of Thermodynamics and, in particular, in application to chemical kinetic properties.[71]

6. Global Information Systems in Science will play a critical role in the future development of the scientific cyberspace infrastructure.

Acknowledgment

This article is prepared as a contribution for The Festschrift issue honoring our friend and colleague for many years, Prof. William Wakeham of the University of Southampton (UK) who made outstanding contributions to the fields of Thermodynamics and Thermophysics. We join many others in congratulating Prof. Wakeham on the occasion of his retirement and wish him to enjoy his "additional degrees of freedom" for many years to come. We also express our appreciation to our colleagues at the NIST Thermodynamics Research Center in Boulder, Colorado, Drs. Rob Chirico, Vladimir Diky, Chris Muzny, Andrei Kazakov, Kenneth Kroenlein, and Ilmutdin Abdulagatov, without whose devoted and creative efforts the concept of the Global Information System in the field of Thermodynamics would still remain just a "beautiful dream".

Literature Cited

(1) http://www.trc.nist.gov/.
(2) http://pubs.acs.org/journals/jceaax/index.html.
(3) http://www.sciencedirect.com/science/journal/00219614.
(4) http://www.sciencedirect.com/science/journal/03783812.
(5) http://www.trc.nist.gov/GDC.html.
(6) Diky, V. V.; Chirico, R. D.; Wilhoit, R. C.; Dong, Q.; Frenkel, M. Windows-Based Guided Data Capture Software for Mass-Scale Thermophysical and Thermochemical Property Data Collection. *J. Chem. Inf. Comp. Sci.* **2003**, *43*, 15–24.
(7) Frenkel, M. Global Communications and Expert Systems in Thermodynamics: Connecting Property Measurement and Chemical Process Design. *Pure Appl. Chem.* **2005**, *77*, 1349–1367.
(8) Lowen, B.; Schulz, S. Excess Molar Enthalpies of Acetone + Water, Cyclohexane, Methanol, 1-Propanol, 2-Propanol, 1-Butanol, and 1-Pentanol at 283.15, 298.15, 323.15, 343.15, 363.15 K. *Thermochim. Acta* **1995**, *262*, 69–82.
(9) Frenkel, M.; Dong, Q.; Wilhoit, R. C.; Hall, K. R. TRC SOURCE Database: A Unique Tool for Automatic Production of Data Compilations. *Int. J. Thermophys.* **2001**, *22*, 215–226.
(10) Chirico, R. D.; Frenkel, M.; Diky, V. V.; Marsh, K. N.; Wilhoit, R. C. ThermoML − an XML-based Approach for Storage and Exchange of Experimental and Critically Evaluated Thermophysical and Thermochemical Property Data. 2. Uncertainties. *J. Chem. Eng. Data* **2003**, *48*, 1344–1359.
(11) *Guide to the Expression of Uncertainty in Measurement* (International Organization for Standardization, Geneva, Switzerland, 1993). This *Guide* was prepared by ISO Technical Advisory Group 4 (TAG 4), Working Group 3 (WG 3). ISO/TAG 4 has as its sponsors the BIPM, IEC, IFCC (International Federation of Clinical Chemistry), ISO, IUPAC (International Union of Pure and Applied Chemistry), IUPAP (International Union of Pure and Applied Physics), and OIML. Although the individual members of WG 3 were nominated by the BIPM, IEC, ISO, or OIML, the *Guide* is published by ISO in the name of all seven organizations.
(12) *U. S. Guide to the Expression of Uncertainty in Measurement*, ANSI/NCSL Z540−2-1997, ISBN 1−58464−005−7. NCSL International: Boulder, CO. 1997.
(13) Taylor, B. N.; Kuyatt, C. E. *Guidelines for the Evaluation and Expression of Uncertainty in NIST Measurement Results*, NIST Technical Note 1297; NIST: Gaithersburg, Maryland. 1994.
(14) http://www.sciencedirect.com/science/journal/00406031.
(15) http://www.springer.com/materials/journal/10765.
(16) http://www.chinatefl.com/shaanxi/study/xajd.htm.
(17) http://www.mpei.ru/LANG/ENG/Main/About/General/General.asp.
(18) Finkelstein, C.; Aiken, P. *Building Corporate Portals with XML*; McGraw-Hill: New York, 1999.
(19) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Marsh, K. N.; Dymond, J. H.; Wakeham, W. A.; Stein, S. E.; Königsberger, E.; Goodwin, A. R. H. XML-based IUPAC Standard for Experimental, Predicted, and Critically Evaluated Thermodynamic Property Data Storage and Capture (ThermoML): IUPAC Recommendations 2006. *Pure Appl. Chem.* **2006**, *78*, 541–612.
(20) http://old.iupac.org/namespaces/ThermoML/index.html.
(21) http://www.asciitable.com/.
(22) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Dong, Q.; Frenkel, S.; Franchois, P. R.; Embry, D. L.; Teague, T. L.; Marsh, K. N.; Wilhoit, R. C. ThermoML − an XML-based Approach for Storage and Exchange of Experimental and Critically Evaluated Thermophysical and Thermochemical Property Data. 1. Experimental Data. *J. Chem. Eng. Data* **2003**, *48*, 2–13.
(23) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Marsh, K. N.; Dymond, J. H.; Wakeham, W. A. 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. *J. Chem. Eng. Data* **2004**, *49*, 381–393.
(24) Sandhu. P. *The MathML Handbook*; Charles River Media: Hingham, MA, 2003. See also: http://www.w3.org/Math/.
(25) http://www.trc.nist.gov/ThermoML_Opener.html.
(26) Ashour, I. Liquid-liquid Equilibrium of MTBE + Ethanol + Water and MTBE + 1-Hexanol + Water over the Temperature Range of 288.15 to 308.15 K. *J. Chem. Eng. Data* **2005**, *50*, 113–118.
(27) http://www.w3.org/TR/xslt.
(28) Frenkel, M.; Chirico, R. D.; Diky, V.; Muzny, C. D.; Lemmon, E. W.; Yan, X.; Dong, Q. *NIST ThermoData Engine, NIST Standard Reference Database 103a − Pure Compounds, Version 2.1*; NIST, Boulder: Colorado, USA, 2008.http://www.nist.gov/srd/nist103a.htm.
(29) Frenkel, M.; Chirico, R. D.; Diky, V.; Muzny, C. D.; Kazakov, A.; Lemmon, E. W. *NIST ThermoData Engine, NIST Standard Reference Database 103b - Pure Compounds and Binary Mixtures, Version 3.0*; NIST: Boulder, Colorado, USA, 2008.http://www.nist.gov/srd/nist103b.htm.
(30) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Yan, X.; Dong, Q.; Muzny, C. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. *J. Chem. Inf. Model.* **2005**, *45*, 816–838.
(31) Diky, V.; Muzny, C. D.; Lemmon, E. W.; Chirico, R. D.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 2: Equations of State on Demand and Dynamic Updates over the Web. *J. Chem. Inf. Model.* **2007**, *47*, 1713–1725.
(32) Diky, V.; Chirico, R. D.; Kazakov, A.; Muzny, C. D.; Frenkel, M. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 3. Binary Mixtures. *J. Chem. Inf. Model.* **2009**, *49*, in press.
(33) (a) Wilhoit, R. C.; Marsh, K. N. Future Directions for Data Compilation. *Int. J. Thermophys.* **1999**, *20*, 247–255. (b) Frenkel, M. Dynamic Compilation: A Key Concept for Future Thermophysical Data Evaluation. In *Forum 2000: Fluid Properties for New Technologies − Connecting Virtual Design with Physical Reality*; Rainwater, J. C., Friend, D. G., Hanley, H. J. M., Harvey, A. H., Holcomb, C. D., Laesecke, A., Magee, J. W., Muzny, C., Eds.; NIST Special Publication 975: Gaithersburg, MD, 2001; pp 83−84.
(34) Frenkel, M. From Molecular Insights and Chemical Technologies to Communications and Expert Systems: a Few Short Thermodynamic Stories. *J. Chem. Thermodyn.* **2007**, *39*, 169–191.
(35) Roy-Faderman, A.; Koletzke, P.; Dorsey, P. *Oracle JDeVeloper 10g Handbook*; McGraw-Hill: New York, 2004; pp 184−186.
(36) http://www.w3.org/Protocols/.
(37) *NIST Ionic Liquids Database, ILThermo. NIST Standard Reference Database 147*; National Institute of Standards and Technology, Standard Reference Data Program: Gaithersburg, MD, 2006.http://ilthermo.boulder.nist.gov/ILThermo/.
(38) Dong, Q.; Muzny, C. D.; Kazakov, A. F.; Diky, V.; Magee, J. W.; Widegren, J. A.; Chirico, R. D.; Marsh, K. N.; Frenkel, M. ILThermo: A Free-Access Web Database for Thermodynamic Properties of Ionic Liquids. *J. Chem. Eng. Data* **2007**, *52*, 1151–1159.
(39) *NIST Web Thermo Tables. NIST Standard Reference Subscription Databases 2 and 3*; National Institute of Standards and Technology, Standard Reference Data Program: Gaithersburg, MD, 2008.http://www.nist.gov/srd/nistwebsub2.htm and http://www.nist.gov/srd/nistwebsub3.htm.
(40) Kazakov, A.; Muzny, C. D.; Chirico, R. D.; Diky, V. V.; Frenkel, M. Web Thermo Tables − an On-line Version of the TRC Thermodynamic Tables. *J. Res. Natl. Inst. Stand. Technol.* **2008**, *113*, 209–220.
(41) http://www.trc.nist.gov/ThermoML.html.
(42) Frenkel, M.; Chirico, R. D.; Diky, V.; Muzny, C.; Dong, Q.; Marsh, K. N.; Dymond, J. H.; Wakeham, W. A.; Stein, S. E.; Königsberger, E.; Goodwin, A. R. H.; Magee, J. W.; Thijssen, M.; Haynes, W. M.; Watanasiri, S.; Satyro, M.; Schmidt, M.; Johns, A. I.; Hardin, G. R. New Global Communication Process in Thermodynamics: Impact on

Quality of Published Experimental Data. *J. Chem. Inf. Model.* **2006**, *46*, 2487–2493.
(43) Marsh, K. N. New Process for data submission and dissemination. *J. Chem. Eng. Data* **2003**, *48*, 1.
(44) Electronic data submission to the NIST Thermodynamics Research Center. *J. Chem. Thermodyn.* **2004**, *36*, IV.
(45) Electronic data submission to NIST Thermodynamics Research Center. *Fluid Phase Equilib*. **2004**, *226*, V.
(46) Electronic data submission to NIST Thermodynamics Research Center. *Thermochim. Acta* **2004**, *421*, 241.
(47) Haynes, W. M. Electronic data submission for the International Journal of Thermophysics. *Int. J. Thermophys.* **2005**, *26*, 307.
(48) http://old.iupac.org/namespaces/ThermoML/ThermoML.xsd.
(49) http://old.iupac.org/standing/cpep.html.
(50) http://www.iactweb.org/.
(51) Drefahl, A. Extraction and Application of Environmentally Relevant Chemical Information from the ThermoML Archive. In *Environmental Informatics and System Research*; Hryniewicz, O., Studzinski, J., Romaniuk, M., Eds.;*Proceedings of the EnviroInfo2007 Conference, Warsaw*, Shaker Verlag: Aachen, 2007; Vol. 1, pp 71−78.
(52) Korichi, M.; Gerbaud, V.; Floquet, P.; Meniai, A.-H.; Nacef, S.; Joulia, X. Computer-Aided Aroma Design. 1-Molecular Knowledge Framework. *Chem. Eng. Proc.: Proc. Intens.* **2008**, *47*, 1902–1911.
(53) Korichi, M.; Gerbaud, V.; Talou, T.; Floquet, P.; Meniai, A.-H.; Nacef, S. Computer-Aided Aroma Design. 2-Quantitative Structure−Odour Relationship. *Chem. Eng. Proc.: Proc. Intens.* **2008**, *47*, 1912–1925.
(54) Homann, J.; Schmidt, M. Infotherm: Thermophysical Data of Mixtures and Pure Compounds. *THERMO International 2006*; Boulder, Colorado, USA, 2006. Book of Abstracts, pp 107−108.
(55) Dong, Q.; Yan, X.; Wilhoit, R. C.; Hong, X.; Chirico, R. D.; Diky, V. V.; Frenkel, M. Data Quality Assurance for Thermophysical Property Databases − Applications to the TRC SOURCE Data System. *J. Chem. Inf. Comp. Sci.* **2002**, *42*, 473–480.
(56) Iglesias-Otero, M. A.; Troncoso, J.; Carballo, E. Density and Refractive Index for Binary Systems of the Ionic Liquid [Bmim][BF_4] with Methanol, 1,3-Dichloropropane, and Dimethyl Carbonate. *J. Solution Chem.* **2007**, *36*, 1219–1230.
(57) Iglesias-Otero, M. A.; Troncoso, J.; Carballo, E. Density and Refractive Index for Binary Systems of the Ionic Liquid [Bmim][BF_4] with Methanol, 1,3-Dichloropropane and Dimethyl Carbonate (Erratum). *J. Solution Chem.* **2008**, *37*, 135–136.
(58) Campos, V.; Marigliano, A. C. G.; Solimo, H. N. Density, Viscosity, Refractive Index, Excess Molar Volume, Viscosity, and Refractive Index Deviations and Their Correlations for the (Formamide + Water) System. Isobaric (Vapor + Liquid) Equilibrium at 2.5 kPa. *J. Chem. Eng. Data* **2008**, *53*, 211–216.
(59) Campos, V.; Marigliano, A. C. G.; Solimo, H. N. Density, Viscosity, Refractive Index, Excess Molar Volume, Viscosity, and Refractive Index Deviations and Their Correlations for the (Formamide + Water) System. Isobaric (Vapor + Liquid) Equilibrium at 2.5 kPa. (Corrections). *J. Chem. Eng. Data* **2008**, *53*, 1046.
(60) Zhang, W.; Kiran, E. (p, V, T) Behaviour and miscibility of (polysulfone+THF+carbon dioxide) at high pressures. *J. Chem. Thermodyn.* **2003**, *35*, 605–624.
(61) Zhang, W.; Kiran, E. Corrigendum to "(*p*,*V*,*T*) Behaviour and miscibility of (polysulfone + THF + carbon dioxide) at high pressures". *J. Chem. Thermodyn.* **2003**, *35* (4), 605–624; *J. Chem. Thermodyn.* **2006**, *38*, 504.
(62) Frenkel, M.; Chirico, R. D.; Diky, V.; Muzny, C. D.; Watanasiri, S.; Chau-Chyun Chen. New Technologies for Product on-Demand Design. *11th International Conference on Properties and Phase Equilibria for Product and Process Design*; Hersonissos, Crete, Greece, 2007; Book of Abstracts, p 99.
(63) Levine, J. R. *Linkers and Loaders*; Acedemic Press, London, 2000.
(64) Wakeham, W. A.; Assael, M. A.; Atkinson, J. K.; Bilek, J.; Fareleira, J. M. N. A.; Fitt, A. D.; Goodwin, A. R. H.; Oliveira, C. M. B. P. Thermopysical Property Measurements: The Journey from Accuracy to Fitness for Purpose. *Int. J. Thermophys.* **2007**, *28*, 372–416.
(65) Yan, X.; Dong, Q.; Hong, X. Reliability Analysis of Group-Contribution Methods in Predicting Critical Temperatures of Organic Compounds. *J. Chem. Eng. Data* **2003**, *48*, 374–380.
(66) Huber, M. L.; Lemmon, E. W.; Diky, V.; Smith, B. L.; Bruno, T. J. Chemically Authentic Surrogate Mixture Model for the Thermophysical Properties of a Coal-Derived Liquid Fuel. *Energy Fuels* **2008**, *22*, 3249–3257.
(67) Paluch, A. S.; Shen, V. K.; Errington, J. R. Comparing the Use of Gibbs Ensemble and Grand-Canonical Transition-Matrix Monte Carlo Methods to Determine Phase Equilibria. *Ind. Eng. Chem. Res.* **2008**, *47*, 4533–4541.
(68) Kazakov, A.; Muzny, C. D.; Chirico, R. D.; Diky, V.; Frenkel, M. Quantitative Structure-Property Relationship (QSPR) Study of Critical Properties for Pure Compounds Based on a Large Dataset Derived from Experimental Data. *18th European Conference on Thermophysical Properties*; Pau, France, 2008; Book of Abstracts, p 220.
(69) Diky, V. V.; Frenkel, M.; Karpushenkava, L. S. Thermodynamics of Sublimation of Cubane: Natural Anomaly or Experimental Error? *Thermochim. Acta* **2003**, *408*, 115–121.
(70) Bashir-Hashemi, A.; Chickos, J. S.; Hanshaw, W.; Hui, Zhao.; Farivar, B. S.; Liebman, J. F. The Enthalpy of Sublimation of Cubane. *Thermochim. Acta* **2004**, *424*, 91–97.
(71) Frenklach, M.; Packard, A.; Djurisic, Z. M.; Golden, D. M.; Bowman, C. T.; Green, W. H.; McRae, G. J.; Allison, T. C.; Rosasco, G. J.; Pilling, M. J. Process Informatics Model (PrIMe): A customer for ThermoML. *231st National Meeting of the American Chemical Society*; Atlanta, GA, March 26−30, 2006.

Received for review November 18, 2008. Accepted December 16, 2008.

JE800877F