ThermoML – a New IUPAC Standard for Thermodynamic Data Communications

Industry relies heavily on thermodynamic data as the basis for product and process development and optimization. For effective and efficient use, these data, the vast majority of which are presented in peer-reviewed journals, must be made readily available and assessed for uncertainty. Providing the infrastructure that fills this need is an important program at NIST.

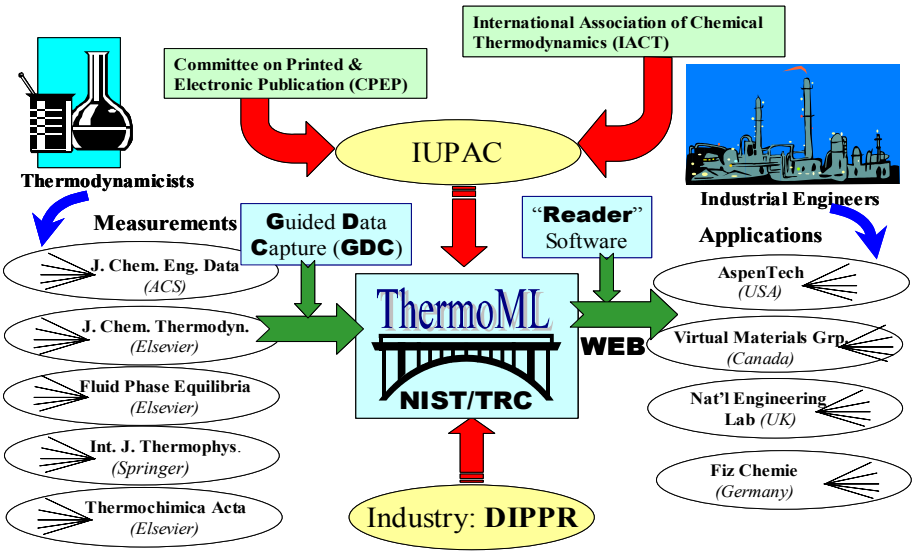
M. Frenkel, R. D. Chirico, V. V. Diky, Q. Dong, C. Muzny, G. R. Hardin (Div. 838)

In developing the data communication standard the following criteria were primary. The standard should:

- take advantage of existing data exchange tools and standards to the extent possible;
- be readily adapted and extended to the many disciplines that require thermodynamic data;
- be structured to support all the information necessary to unambiguously identify the chemical system, the conditions of measurement, the properties reported, and the uncertainty that should be associated with the numerical values;
- be developed with the research and engineering community and become the basis for an internationally sanctioned standard.

Working with colleagues, the Thermodynamics Research Center (now TRC Group) developed ThermoML, an XML (Extensible Markup Language)-based approach for storage and exchange of thermophysical and thermochemical property data. Taking into account the diversity of thermodynamic data and numerous methods for their reporting and presentation, standardization of thermodynamic data communications is very complex. The ThermoML structure represents a balanced combination of hierarchical and relational elements. The ThermoML schema structure explicitly incorporates structural elements related to basic principles of phenomenological thermodynamics: thermochemical and thermophysical properties (equilibrium and transport), state variables, system constraints, phases, and units. Meta- and numerical data records are grouped into "nested blocks"of information corresponding to data sets. The structural features of the ThermoML metadata records ensure unambiguous interpretation of numerical data and allow data-quality control based on the Gibbs Phase Rule. 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).

Global Data Communication Process



The NIST approach to creating the overall infrastructure for addressing industry's thermodynamic data needs is illustrated in the Figure. As can be seen, a standard for data communications, ThermoML, is a central element of this infrastructure.

The standard includes definitions and descriptions of all quantities related to the expression of uncertainty in a manner that conforms to the Guide to the Expression of Uncertainty in Measurement (ISO, 1993). This very important characteristic of the standard required an extensive scientific interpretation of the "Guide" for the field of thermodynamics. ThermoML currently covers all major data types: experimental, critically evaluated, and predicted. Communication of the parameters of predictive equations has been accomplished by linking the *Ther-*

moML-Equation Definition schema to MathML (The Mathematical Markup Language). Schema elements in ThermoML for equation representation provide for storage of the various equation components required for the specific equation definition. The nature or scope of the equations is not restricted in any way. ThermoML has recently been extended to cover thermodynamic properties of polymers and electrolytes. Also, the compound identification of the ThermoML schema has been expanded to include the IUPAC-NIST Chemical Identifier (INChI).

Final recommendations to IUPAC on ThermoML as the IUPAC standard for thermodynamic data capture and exchange were accepted by the IUPAC Committee on Printed and Electronic Publications in August, 2005 and by the IUPAC Interdivisional Committee on Terminology, Nomenclature and Symbols in December, 2005. IUPAC Recommendations on ThermoML as a new IUPAC standard were published in IUPAC's *Pure and Applied Chemistry* in 2006.

Guided Data Capture (GDC) software has been developed for mass-scale collection from the archival literature and about-to-be-published experimental thermophysical and thermochemical property data for organic chemical systems involving one, two, and three components, chemical reactions, and chemical equilibria. GDC allows for robust, user-friendly collection of all the data and metadata elements specified in the ThermoML standard.

Combination of the software tools incorporating GDC and ThermoML allowed establishment of a new data communication process, which now includes major journals in the field of thermodynamics, such as the *Journal of Chemical and Engineering Data*, *The Journal of Chemical Thermodynamics*, *Thermochimica Acta*, and *Fluid Phase Equilibria*. In FY05 this process has been extended to *International Journal of Thermophysics*. As a result of the implementation of this process, hundreds of authors worldwide generated ThermoML files of their reported data at the time of publication which resulted in significant improvement of the quality of the published experimental data. The TRC Group has designed and now supports Web distribution of the ThermoML files in the public domain without restriction. It has also been working with major data-user organizations (Aspentech, U.S.A.; National Engineering Laboratory, U.K.; Fiz Chemie, Germany; Virtual Materials Group, Canada; Korean Institute of Science and Technology Information) to develop software "readers" of the ThermoML files.

Future Plans

Future plans include an expansion of the ThermoML as well as its software supporting infrastructure to biochemicals and their reactions.

References

M. Frenkel, R. D. Chirico, V. V. Diky, Q. Dong, S. Frenkel, P. R. Franchois, D. L. Embry, T. L. Teague, K. N. Marsh, R. C. Wilhoit, ***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.

R. D. Chirico, M. Frenkel, V. V. Diky, K. N. Marsh, R. C. Wilhoit, ***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.

M. Frenkel, R. D. Chirico, V. V. Diky, K. N. Marsh, J. H. Dymond, W. A. Wakeham, ***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.

V. V. Diky, R. D. Chirico, R. C. Wilhoit, Q. Dong, M. Frenkel ***Windows-Based Guided Data Capture Software for Mass-Scale Thermophysical and Thermochemical Property Data Collection***, *J. Chem. Inform. and Comp. Sci*., **2003,** 43, 15-24.

M. Frenkel, ***Global Communications and Expert Systems in Thermodynamics: Connecting Property Measurement and Chemical Process Design***, *Pure and Applied Chemistry*, **2005**, 77, 1349-1367.

M. Frenkel, R. D. Chirico, V. V. Diky, K. N. Marsh, J. H. Dymond, W. A. Wakeham, S. E. Stein, E. Koenigsberger, A. R. H. Goodwin, ***XML-based IUPAC Standard for Experimental, Predicted, and Critically Evaluated Thermodynamic Property Data Storage and Capture (ThermoML): IUPAC Recommendations 2005***, *Pure and Applied Chemistry*, **2006**, 78.