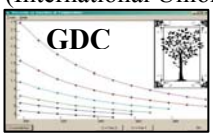
ThermoML – An Emerging IUPAC Standard for Thermodynamic Data Communications

M. Frenkel, R. D. Chirico, Q. Dong, G. R. Hardin (838), and V. V. Diky (Belarussian State University)

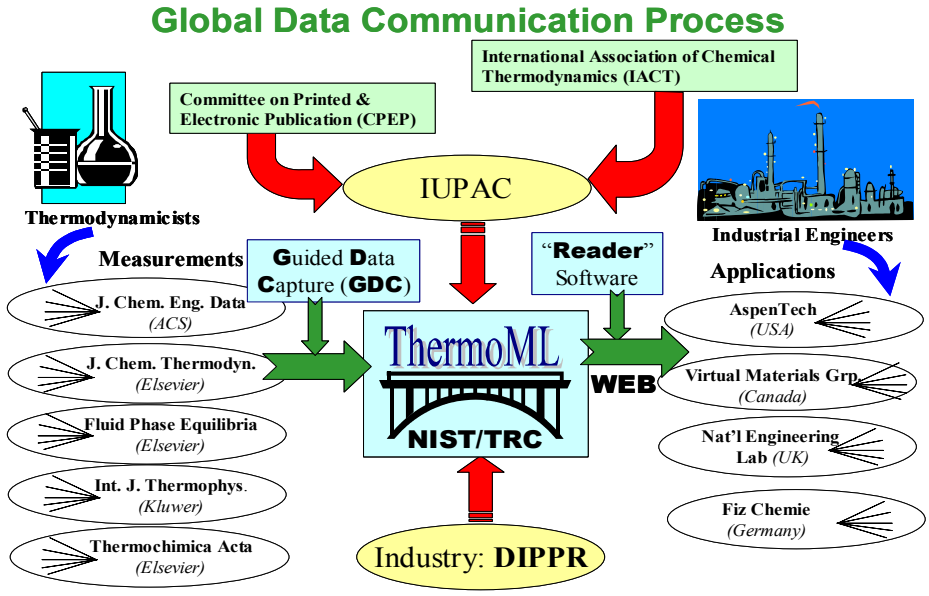
Researchers in the TRC Group have developed ThermoML, an XML (Extensible Markup Language)-based approach for storage and exchange of thermophysical and thermochemical property data.

Establishing efficient means for thermodynamic data communications is absolutely critical for provision of solutions to such technological challenges as elimination of data processing redundancies and data collection process duplication, creation of comprehensive data storage facilities, and rapid data propagation from the measurement to the data management system and from the data management system to engineering applications. Taking into account the diversity of thermodynamic data and numerous methods of 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).

Definitions and descriptions of all quantities related to the expression of uncertainty in ThermoML conform to the Guide to the Expression of Uncertainty in Measurement (ISO, 1993). In order to implement this extension of ThermoML, the TRC Group (formerly the Thermodynamic Research Center) made an interpretation of the US Guide to the Expression of Uncertainty in Measurement for the field of thermodynamics. In 2004, ThermoML was completed with incorporation of extensions for critically evaluated data, predicted data, and equation representations. In early 2004, ThermoML was accepted as the foundation for the development of the IUPAC (International Union of Pure and Applied Chemistry) standard for thermodynamic data communications.



In order to build an infrastructure for the process of global thermodynamic data communication, Guided Data Capture (GDC) software was developed for mass-scale abstraction from the literature of experimental thermophysical and thermochemical property data for organic chemical systems involving one, two, and three components, chemical reactions, and chemical equilibria.



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*. As a result of the implementation of this process, hundreds of

authors worldwide generate ThermoML files of their reported data at the time of publication. The TRC Group has designed and now supports Web distribution of the ThermoML files in the public domain without restriction.

Recommendations to IUPAC for a formal approval of ThermoML as a new IUPAC standard will be finalized in FY 2005.

The TRC Group has also been working with major data-user organizations (Aspentech, US; National Engineering Laboratory, UK; Fiz Chemie, Germany; Virtual Materials Group, Canada; Korean Institute of Science and Technology Information) to develop software 'readers' of the ThermoML files. In the near future ThermoML will be expanded to provide compound identification using the IUPAC-NIST Chemical Identifier (INChI).