

# **Gas Hydrate Research Database and Web Dissemination Channel**

Final Technical Report  
October 1, 2006 to September 30, 2009

Principle Investigator: Michael Frenkel  
Report Prepared By: Kenneth Kroenlein  
National Institute of Standards and Technology

October 2009

DE-AI26-06NT42938

K. Kroenlein, V. Diky, R.D. Chirico, A. Kazakov, C.D. Muzny, and M. Frenkel  
National Institute of Standards and Technology  
Thermophysical Properties Division  
Thermodynamics Research Center (TRC)  
325 Broadway  
Boulder, CO 80305-3328, USA

## **NOTICE:**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## **ABSTRACT**

To facilitate advances in application of technologies pertaining to gas hydrates, a United States database containing experimentally-derived information about those materials was developed. The Clathrate Hydrate Physical Property Database (NIST Standard Reference Database # 156) was developed by the TRC Group at NIST in Boulder, Colorado paralleling a highly-successful database of thermodynamic properties of molecular pure compounds and their mixtures and in association with an international effort on the part of CODATA to aid in international data sharing. Development and population of this database relied on the development of three components of information-processing infrastructure: (1) guided data capture (GDC) software designed to convert data and metadata into a well-organized, electronic format, (2) a relational data storage facility to accommodate all types of numerical and metadata within the scope of the project, and (3) a gas hydrate markup language (GHML) developed to standardize data communications between “data producers” and “data users”. Having developed the appropriate data storage and communication technologies, a web-based interface for both the new Clathrate Hydrate Physical Property Database, as well as Scientific Results from the Mallik 2002 Gas Hydrate Production Research Well Program was developed and deployed at <http://gashydrates.nist.gov>.

## TABLE OF CONTENTS

<b>ABSTRACT .....</b>	2
<b>TABLE OF CONTENTS .....</b>	3
<b>EXECUTIVE SUMMARY .....</b>	4
<b>REPORT DETAILS .....</b>	6
<b>Introduction .....</b>	6
<b>Development of Project-Specific Support Resources .....</b>	7
<b>Literature Archive.....</b>	8
<b>Guided Data Capture.....</b>	9
<b>Database Architecture.....</b>	11
<b>Gas Hydrate Markup Language .....</b>	13
<b>World Wide Web-based Data Dissemination .....</b>	14
<b>Conclusions .....</b>	15
<b>GRAPHICAL MATERIALS LIST .....</b>	16
<b>REFERENCES .....</b>	49
<b>LIST OF ACRONYMS AND ABBREVIATIONS.....</b>	51

## EXECUTIVE SUMMARY

To facilitate advances in application of technologies pertaining to gas hydrates, a United States database containing experimentally-derived information about those materials was developed. The Clathrate Hydrate Physical Property Database (NIST Standard Reference Database # 156) was created by the Thermodynamics Research Center (TRC) Group at the National Institute of Standards and Technology (NIST) in Boulder, Colorado. The database development paralleled that of a highly-successful relational database of thermodynamic properties of pure molecular compounds and their mixtures (SOURCE) that is maintained by TRC at NIST. The project was completed in association with an international effort by the Committee on Data for Science and Technology (CODATA) to aid in international data sharing. Development and population of the new database relied on the development of three components of information-processing infrastructure: (1) Guided Data Capture (GDC) software designed to convert data and metadata into a well-organized, electronic format, (2) a relational data storage facility to accommodate all types of numerical and metadata within the scope of the project, and (3) a gas hydrate markup language (GHML) developed to standardize data communications between “data producers” and “data users”, i.e., the academic and industrial research communities.

Guided Data Capture (GDC) software was developed for this project with specific application to characterization and properties of gas hydrates. This software was and continues to be utilized in the data collection process to aid technically competent individuals, typically student associates, who are not expert in data collection and measurement technology in capturing all relevant information from a literature source. The captured information is used by staff experts to produce optimally accurate properties with reliable uncertainty estimates.

The NIST/TRC Gas Hydrate Data Entry Facility was established in support of this project. The facility is staffed by student associates in relevant technical fields under the supervision of NIST professionals, and is at the center of all data selection, capture, evaluation, and archiving activities for this project. The facility staff used the developed GDC software for capture of gas hydrate data sets from articles, selected by senior staff, that were determined to contain information relevant to this project. A continually-growing literature archive of gas hydrate data publications, currently containing 6000 unique titles, was analyzed for data content. The archive consists of refereed journal publications, theses, technical reports, and conference proceedings.

A new database design was developed based upon the existing SOURCE database design to accommodate complexities associated with gas hydrates. Tables for storage of gas hydrate thermophysical and crystallographic data in SOURCE were designed and populated from the data compiled from the literature archive, and continue to gain content as new information becomes available. A new Sun SPARC Enterprise T5240 server was obtained and configured to run RDBMS system Oracle 10g and MySQL to serve the hardware data needs of this project.

The gas hydrate markup language (GHML) was initially developed by a CODATA taskgroup as the intended medium for hydrate data communication across disparate hydrate disciplines. In order to guarantee that such records would contain appropriate data constraint, GHML was modified in conjunction with the taskgroup to meet the needs of the broadly-based gas hydrate community while maintaining consistency with the IUPAC-standard ThermoML and providing the structure necessary to communicate well constrained datasets. This data format is actively being used in international data-sharing development efforts. By providing a uniform model for data and metadata communication for the gas hydrates community, communication across the

varied disciplines associated with these studies can be improved and scientific progress in the field can be facilitated.

An advanced World Wide Web-based interface for the database was developed and made public at <http://gashydrates.nist.gov>. This interface supports complex queries of a chemical or citation nature and supports dynamically interactive tables, visualization of datasets using customizable charts, and download of data for off-line analysis. This interface also provides ready access to the 2002 scientific results of the Mallik Gas Hydrate Production Research Well Program. All capabilities of the viewing software developed originally for dissemination with physical copies of the Mallik project dataset are maintained in the new on-line tools.

All milestones of this three-year data collection, evaluation, and dissemination project were met or exceeded.

## REPORT DETAILS

### Introduction

The interdisciplinary field of gas hydrate research is undergoing rapid growth. Publication rates in peer-reviewed journals have displayed nearly exponential growth in the century following the discovery of hydrates in the laboratory, culminating in more than 3000 refereed publications in the 1990's [1]. Much of this recent growth is due to the perceived value of methane clathrate as a non-petroleum-derived large-scale energy resource [2]. Recent estimates of the world's naturally-occurring hydrated methane vary widely, ranging from  $2.5 \times 10^{15} \text{ m}^3$  [3] to  $1.2 \times 10^{17} \text{ m}^3$  [4] at STP, but the amount of organic carbon in hydrates can be conservatively estimated as a factor of two greater than the total of all remaining petroleum and natural gas reserves [5]. The remote locations where hydrate exists and the dispersed nature of the deposits has prevented prospecting at present, but the perceived potential has encouraged many nations, including Japan, Germany, India, China, Korea, Taiwan, Canada, and the United States to invest heavily in hydrate recovery programs.

Study of natural hydrate occurrences has led to the understanding that they typically exist close to their thermodynamic stability limit [4], so slight changes in ambient temperature or pressure may result in catastrophic release of methane, a potent greenhouse gas, with implications on global climate change [6] and seafloor slope stability [7]. Massive releases of organic carbon to the atmosphere and mass extinction events during the Permian Triassic [8], Late Jurassic [9], Late Paleocene Thermal Maximum [10], and other eras are often connected to the sudden release of hydrated gas.

Gas hydrate publication rates are now such that a diligent researcher could be easily overwhelmed in attempting to maintain a broad understanding of the state of the art. One solution to this difficulty is the centralization of critically evaluated data sets. Such a database can facilitate understanding of naturally occurring hydrate interactions with geophysical processes, aid in the application of hydrate knowledge to technologies involved in resource recovery and storage, and support the gas hydrate research community in general. The developed database, the scope of which includes thermophysical and structural data, provides to researchers the ability to submit new datasets and retrieve high quality, critically evaluated data. By establishing the hydrate database at the United States National Institute of Standards and Technology (NIST) in Boulder, Colorado, the viability of this project is secured well into the future. A critically evaluated hydrate database is essential for eliminating data redundancies, highlighting key data gaps, and providing an assurance of data quality to aid research efforts within the broader research community.

A hydrate database center has been established at NIST as part of the Thermodynamics Research Center (TRC) [11]. The existing database at the core of previous TRC activities is the NIST SOURCE Data Archival System (SOURCE) [12-13] which is the largest relational archival experimental data system, currently including more than 120 properties (including chemical structural information) for pure compounds, mixtures, and chemical reactions, with data records numbering in the millions. All TRC developments are built upon the algebraic constraints of the Gibbs phase rule, which states the number free parameters that must be specified in a system in order for that system to be thermodynamically well-defined is equal to two more than the difference of number of chemical components and the number of phases. The gas hydrate database represents a critically evaluated dynamic data set, allowing for continuous updating and reliability analysis. TRC has extensive experience in software development for dynamic critical data evaluation, with particular application to thermophysical properties. The ThermoData Engine software [14-15], developed at TRC, is the first full-scale implementation of the dynamic

data evaluation concept [16]. TRC currently has agreements with major publishers in the field of thermophysical properties for implementation of data quality assurance (DQA) procedures at the time of data submission by authors. Authors provide their manuscripts and data summaries that are used at NIST with Guided Data Capture (GDC) software [17-18] for generation of structured data files. This approach assures that submitted data are in an appropriate format [19-21] and include sufficient supporting information regarding methods and materials to allow for accurate reliability estimates. Application of this proven model has been essential to the successful development of this hydrate database. Modifications to the GDC software were implemented and appropriate schema for storage of well-constrained hydrate data were developed specifically for this collection effort.

The data-transfer approaches associated with this data capture and storage effort are being coordinated with CODATA, which has been developing a markup language called Gas Hydrate Markup Language (GHML) [22-25] for communicating gas hydrate data throughout the research community and an international hydrate portal technology for centralized access to a number of database efforts. So that the data collected by this effort will be available to such a portal, all database output is fully consistent with GHML. To aid in database access prior to the availability of such a portal, a state-of-the-art web interface was designed for the data archive. Using a number of technologies [30-33], a native application-like interface rendered using a traditional web browser was developed. This interface provides the capability to search and navigate through the data sets, view them with sortable tables, chart data sets against each other and download each data set to a local machine for additional analysis. This interface also provides access to the 2002 scientific results of the Mallik Gas Hydrate Production Research Well Program, reproduced with permission, and provides all capabilities of the viewing software previously developed for dissemination with physical copies of the Mallik dataset.

### **Development of Project-Specific Support Resources**

In order to identify sources of data for capture and storage, as well as to provide other field-specific advice, an Advisory Committee was established in October 2006. The members of the Advisory Committee are:

Dr. Dendy Sloan (Committee Chair)  
Center for Hydrate Research  
Colorado School of Mines  
Golden, CO 80401  
[esloan@mines.edu](mailto:esloan@mines.edu) (303-273-3723)

Dr. Timothy S. Collett  
USGS Denver Federal Center MS 940  
Denver, Colorado 80225  
[tcollett@usgs.gov](mailto:tcollett@usgs.gov) (303-236-5731)

Dr. George Claypool  
Former Chief Scientist / Chevron JIP Project  
8910 W. 2nd Avenue  
Lakewood, CO 80226  
[geclaypool@aol.com](mailto:geclaypool@aol.com) (303-237-8273)

Tom Smith, Data and Security Architect, Office of Converging Technologies  
Queens College, City University of New York  
Flushing, NY 11367-1597  
tom.smith@cinaplex.com (718-997-5935)

Dr. Michael Frenkel, NIST, Director / Thermodynamics Research Center  
325 Broadway, Mailstop 838.01  
Boulder, CO 80305-3328  
frenkel@boulder.nist.gov (303-497-3952)

As expert members of the gas-hydrate community in the areas of property research (Dr. Sloan), energy resource characterization (Dr. Collett), and field formation and assessment (Dr. Claypool), the members provided guidance in establishing the scope of the database collection to be housed at NIST. As a member of the GHML development team in addition to his expertise in information technologies, Mr. Smith facilitated communication between the development teams and helped guarantee GHML's evolution would be compatible with both the data requirements of well constrained thermophysical data sets, as well as the broader needs of the gas hydrates community at large.

In addition to the thermophysical and crystallographic data collected and disseminated as per the mandate of this project, the Advisory Committee recommended that scientific results from the Mallik Gas Hydrate Production Research Well Program be disseminated via the same channel. Mallik represents a novel exploration of the technical feasibility of natural gas production from a permafrost-based deposit of gas hydrates. The technical results of the 2002 studies have not been available outside of physical materials published by the Geologic Survey of Canada, and so with appropriate permissions those results and all associated visualization capabilities have been replicated through the web interface.

### **Literature Archive**

A primary task of this program was the collection and characterization of a literature archive. This collection had as its basis a large (3,500 unique sources) electronic document archive donated to the collection effort by Dr. Dendy Sloan at the Center for Hydrate Research, Colorado School of Mines. At present, the complete archive, maintained as electronic PDF files, contains approximately 6,000 unique sources. The archive includes peer-reviewed journal articles, technical reports, master's theses and doctoral dissertations in a number of languages and dates as far back as the 18<sup>th</sup> century. Of the total archive, approximately 300 have been determined to contain independent thermophysical or crystallographic data which are sufficiently well-constrained for the goals of this project. This document archive continues to grow as additional new material is published and culled from the academic literature.

Even rudimentary review of an archive of this size for bibliographic information is a non-trivial activity. For the task of reviewing this article set and evaluating data content, the TRC Gas Hydrates Data Entry Facility was established in January 2008, paralleling the previously established TRC Data Entry Facility. This new group included four undergraduates in relevant technical fields from the University of Colorado at Boulder and the Colorado School of Mines. Under the direction of Dr. Kenneth Kroenlein of the TRC group, the students reviewed source materials, assembled an in-house citation database to track documents and collected information from those files following data collection protocols established for the proven TRC Data Entry Facility. The group functioned independently until May 2009, at which time it was folded into

the TRC Data Entry Facility as the data processing operation transitioned into a maintenance phase. The group continues to analyze newly obtained materials and will continue to add them to the database as appropriate.

### **Guided Data Capture**

Information from original data sources is not entered directly into the NIST SOURCE Data Archival System (SOURCE) but is captured or “compiled” in the form of batch data files (coded ASCII text). This allows application of extensive completeness and consistency checks during the capture process before the data are loaded into a central repository. Due to the complexity of the properties and chemical systems involved, extensive expertise has traditionally been required for data compilation. Moreover, expertise in data and measurements is needed to assess uncertainties for each property value. In establishment of the Data Entry Facility at NIST, two major concerns were identified: (1) how to ensure quality of captured information with technically sound but inexperienced data compilers and (2) how to minimize errors before the data are introduced into SOURCE. To meet these goals, interactive Guided Data Capture (GDC) software, written in Microsoft Visual Basic, was developed. The program guides data capture and provides convenient review and editing mechanisms. Undergraduate students involved in in-house data capture played, and continue to play, a key role in the testing of the GDC software.

With the development of collaborations with major peer reviewed journals for the capture of experimental data as they are published, an additional role for GDC evolved. In addition to the creation of batch data files for loading into SOURCE, the GDC software simultaneously creates a separate text document coded in XML [19-21] format for easy access and use by the scientific and data management community. These formatted text documents are available on the internet together with a full description of the XML definitions and schema.

TRC data-quality-assurance (DQA) policies, as they relate to a database effort such as SOURCE, can be subdivided into six steps: (1) literature collection, (2) information extraction, (3) data-entry preparation, (4) data insertion, (5) anomaly detection, and (6) database rectification. The initial steps (1-4) can be very labor intensive and represent key components of the entire data-system operation. GDC serves to provide expert guidance to novice data compilers and minimize errors during steps 1-4. Additional detail of steps 5 and 6 have been discussed previously [27].

GDC functions to guide inexperienced but technically-competent individuals through the process of extracting information from the literature, ensuring the completeness of the information extracted, validating the information through data definition, range checks, etc., and guiding uncertainty assessment to ensure consistency between compilers with diverse levels of experience. A key feature of the GDC software is the capture of information in close accord with customary original-document formats and leaving transformation to formalized data records and XML formats within the scope of the software procedures. Thus, GDC relieves the compiler of the need for knowledge related to the structure of the SOURCE data system or XML formats, thereby eliminating common errors related to data types, length, letter case, and allowable codes. The users of GDC are scientists or students in a science or engineering discipline with varying levels of experience but with competence in the fields of chemistry and chemical engineering.

The GDC program was developed to serve as a powerful and comprehensive tool to be used for both TRC in-house data capture operations as well as a data-collection aid for authors of scientific and engineering publications. The original software, without support for gas hydrate property capture, is available for free downloading via the World Wide Web [18].

Comprehensive documentation for the software is included. The GDC software has features that allow ready detection of 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. Additional information on the development of GDC can be found in the literature [17].

In order to capture experimental data sets pertaining to samples of gas hydrate, the existing GDC software required significant modification. Whereas data normally processed through GDC are either for a pure compound or a mixture of a small number of well-defined compounds in well-defined ratios, a gas hydrate is a non-stoichiometric structure, where determination of crystal compositional distribution may not ever be measured, but can still yield valuable data. Whereas it might be desirable to simply designate such studies as unreliable, the comparative paucity of data precludes such a determination. The solution to this conflict was determined to be the creation of an original data structure within the GDC framework that behaves in many ways like a new compound, defined by the combination of its constituents and known thermodynamic properties. With these modifications, the GDC software supports the capture and organization of data pertaining to bulk properties (e.g. mass specific volume, thermal conductivity, heat capacity at constant pressure per unit mass, speed of sound), phase equilibrium with an arbitrary number of components and phases, crystalline structure and enthalpy of hydrate decomposition for gas hydrates. In particular, the data format for crystalline structure represents an entirely new development with this software. The level of functionality thus attained represents significant progress towards a complete GDC software package for gas-hydrate data; however, experience in this area has shown that continued capture of published data sets may motivate modifications and further extensions.

The basic tree structure of GDC data (Figure 1) is organized around that of the data source document. Following from that are definitions of chemical components in the systems presented within the citation and specific sample information with detailed purity information. A gas hydrate system is then defined by a combination of those chemical components (Figure 2) and a gas hydrate sample is defined through the association of specific samples of those components, as well as the conditions under which the hydrate was formed, if appropriate (Figure 3). It is only once this detailed information regarding purity of constituent compounds is defined that measured properties are entered, allowing for a detailed understanding of the resultant data reliability.

In order to guarantee a well-defined thermodynamic state and to prevent storage of dependent variables as independent, the system is constrained according to the Gibbs Phase Rule. For example, if a three-phase region is being defined in a gas hydrate sample formed from two guest molecules (Figure 4), there exist two degrees of freedom in the system, and hence, two dependent variables are required. The data for the system are then recorded in an internal data table (Figure 5). To prevent transcription errors on the part of the data entry technician, data are copied directly from electronic versions of the source, either obtained via electronic distribution or via text recognition software applied to digitized material. Data consistency can then be verified using native graphing capabilities (Figure 6) within the GDC software.

In order to properly characterize enthalpy of decomposition of a gas hydrate, it is necessary to have well defined ratios of host to guest molecules; for this reason, the system for such a decomposition is characterized as a physical reaction and the enthalpy of decomposition is stored as an enthalpy of reaction (Figure 7). This methodology has additional benefits in that, as the comparatively slow kinetics associated with hydrate formation and the dynamics of hydrate decomposition may yield a condition where the ambient pressure and temperature at which a

study are performed do not necessarily correspond to the equilibrium phase boundary, such data can be stored for future consideration and critical review.

For bulk property measurements, such as specific volume, thermal conductivity, heat capacity at constant pressure, or speed of sound, experimental measurement techniques do not vary significantly from those for pure compounds. For this reason, a significant amount of parallelism was possible between the newly-developed and previously-existing treatments [17]. In order to have a well characterized bulk measurement, we first must define the type of property being measured and the method of measurement (Figure 8) after which the conditions under which the measurements were made must be defined (Figure 9). The numerical data are captured and estimates of the uncertainties for the data are made based on the experimental details and sample purity (Figure 10). For larger data sets, the previously mentioned native graphing capabilities can be utilized.

Characterizing crystal structure is a wholly novel addition to GDC intended for gas hydrate data collection. In order to maintain future extensibility, as well as collect detailed information about the cage structure, information is stored regarding the crystallographic space group, unit cell dimensions and both raw and processed information regarding constituent atom distribution (Figure 11). To provide a reasonable guarantee of generality and compatibility with likely crystalline structure data sets, this new data structure was modeled upon the Crystallographic Information File (CIF) data file format. CIF is an International Union of Crystallography (IUCr) standard used within the crystallographic community for communication of experimental results [28].

### **Database Architecture**

The stated goal of the TRC group at NIST is capture from the world's literature of essentially all experimental data available for thermophysical and thermochemical properties of organic chemical compounds in an automatically-interpretable form. The enormous growth of published thermophysical and thermochemical property data makes simply tracking all newly published materials an intensive task. The data collection serves as the basis for dynamic data evaluation, implemented by TRC in its ThermoData Engine (TDE) software [14-15]. Fundamental to a large-scale data analysis effort, such as this, is a well-structured database complete with appropriate metadata and uncertainties.

The enormous growth of published thermophysical and thermochemical property data (doubling almost every 10 years) makes it practically impossible to use traditional (static) methods of data evaluation. The new concept of dynamic data evaluation requires a large electronic database capable of storing essentially all of the published "raw/observed" experimental data with detailed descriptions of metadata and uncertainties. The combination of this electronic database with artificial intellectual (expert-system) software provides the means to generate recommended property values dynamically or "to order". This concept contrasts sharply with static compilations, which must be initiated far in advance of need. Capture of metadata and uncertainties for the experimental values allows propagation of reliable data-quality limits to the recommended values and, subsequently, to all aspects of chemical process design.

Establishment of a comprehensive data depository is one of the major challenges in implementation of the dynamic data evaluation concept. The NIST SOURCE Data Archival System [12-13] was designed and built to be such a depository for experimental thermophysical and thermochemical properties for organic chemical compounds reported in the world's scientific

literature. The scope of the data system includes more than one hundred defined properties for pure compounds, binary and ternary mixtures, and reacting systems. SOURCE now contains nearly four million numerical values for many thousands of pure compounds, binary and ternary mixtures, and reaction systems. In conjunction with the expansion associated with this project, SOURCE is being migrated to a Sun SPARC Enterprise T5240 server running RDBMS system Oracle 10g and MySQL.

In designing data structures to accommodate the gas hydrate data sets, limitations of the SOURCE architecture associated with definition of complex materials were highlighted. In order to support these relationships as well as those of ionic liquids, stereo-isomeric mixtures, and other complex samples, a new table structure has been designed. The relationships for the total gas hydrate system are shown broadly in Figure 12 and the specific details of defining a complex are shown in Figure 13. All gas-hydrate-specific tables are denoted by the “GH” prefix. To define a chemical complex (table CMPLXID), a series of well-defined compounds (table CMPID) are associated with compositional information, if appropriate, through a pivot table (table CMPLXCMP). Each complex is assigned an identifier which is unique between both the CMPID and CMPLXID tables. This allows property data to be defined equivalently in either case, independent of whether it is associated with a pure compound or a complex system. A gas hydrate complex entry is then associated with the literature source of its data through the unmodified, previously-existing literature reference tables in the GHSYSREF table. As purity information of feedstock is relevant to the ultimate properties of a crystal sample, that information is tied to the system for each component in the system through a gas-hydrate specific GHSAMPLE table.

If the given study was crystallographic in nature, a table entry is made in table GHSTRUCT (Figure 14). This contains basic crystallographic data (space group, lattice parameters) in addition to experimental conditions (system temperature, system pressure, uncertainty methodology). If additional information is provided on either interatomic spacing or Cartesian atomic distribution within the unit cell, such information is stored in tables GHSTRCUTRAW or GHSTRCUTPROC, respectively. This data structure follows that of a Crystallographic Information File (CIF).

Characterization of the complex phase equilibria for gas hydrates, necessary to properly specify the conditions of a thermophysical measurement, required significant extension to the existing SOURCE data storage format. Given that gas hydrate systems may contain from two to an arbitrarily large number of chemical components, the fixed table width approach, previously utilized to guarantee proper system constraint, becomes untenable with a gas-hydrate system. This is easily demonstrated with the application of the Gibbs phase rule to a hydrate-forming natural gas system in equilibrium with sea-water. For an eight-component representation of sea water and an eight component natural gas, this three-phase condition would require fifteen data values for proper constraint. Designing a single table to accommodate this set, in addition to data sets containing two compounds and four phases, would be very inefficient and would be inherently limited if a more complex set were encountered in the future.

The solution adopted in this scenario is shown in Figure 15. Each data series from a given study, defined to be a set of measurements performed by one experimental method on a system with a prescribed set of phases present, is uniquely defined in the GHDATASETS table. Observed phases for this data set are stored in the GHPHASELST table. As nearly all data points have temperature and pressure data associated with them, the primary key for a given data point is specified in the GHTP table. Any additional compositional information for that point is stored in

the GHCOMPOSITION table, which stores not only the data and uncertainty, but what compound was measured and in which phase. Data integrity for that composition data is checked by ensuring that the referenced phases and compounds are present in the system and GHSAMPLE data provided. Property data sets, such as speeds of sound or specific heats, are stored in the GHPROP table with similar constraints. The number of compounds, phases and data values can then be compared and the thermodynamic completeness of a set, as expressed by the Gibbs phase rule, can be determined.

All GDC output files generated by undergraduate data compilers in the Gas Hydrates Data Entry Facility are uploaded into the SOURCE archive after each file is checked for consistency with the original source material by senior staff. When appropriate validating data are available, new materials are also verified against existing stored values, providing integrity checks on historic data as well as providing additional verification for new data.

### **Gas Hydrate Markup Language**

Thermodynamic property data represent a key foundation for development and improvement of all chemical process technologies. However, rapid growth in the number of custom-designed software tools for engineering applications has created an interoperability problem between the formats and structures of thermodynamic data files and required input/output structures for the software applications. Establishment of efficient means for thermodynamic data communications is critical for provision of solutions to such technological challenges as elimination of data processing redundancies, creation of comprehensive data archives, and rapid data propagation from measurement to data management system and from data management system to engineering application. Taking into account the diversity of thermodynamic data and the numerous methods of their reporting and presentation, it is apparent that standardization of thermodynamic data communications is a complex task.

A component of the work performed consisted of reconciling the GHMLv1.0 schema [22-25], an XML format developed prior to this project for communication of gas hydrate data, and ThermoML [19-21], a TRC development and an IUPAC standard for experimental and critically-evaluated thermodynamic property data communication and storage. The structure of ThermoML is based on rational storage of property data with the origin of the data as a major component of the organization construct. Initial efforts attempted to cause a minimum disturbance to the existing data structure and maintained a series of parallel (i.e., non-intersecting) sections that described various types of property data (“field” [23], “laboratory” [24], and “modeling” [25]). Consistency with ThermoML was generated primarily via modification of the laboratory section and the addition of citation information. Given the significant inconsistency in style and nomenclature across the different sections present in the initial schema, a specific effort in reformulating the laboratory section was put toward rectifying this disparity. As developing GHML is an international effort under the auspices of the International Council for Science’s Committee on Data for Science and Technology (CODATA), any proposed schema had to be approved by that body, and this happened at a meeting of the CODATA Hydrate Database Steering Committee on October 27, 2007, with the note that additional unification across the disparate GHML branches was desirable.

Examination of additional data sets across the range of disciplines associated with gas-hydrate studies revealed a range of data that was unsupported by both GHMLv1.0 and the newly accepted version, and it was not clear that these data could be represented by reasonable extensions to either format. In response, a significant modification of GHML was formulated which combined

the FieldData, LabData and ModelData elements into a single DataSet element (Figure 16). Rather than specifying the structure of datasets to be encoded within the XML Schema Definition (XSD), this DataSet element specifies the encoding for metadata common to many different datasets, in the broad categories literature citation (Figure 17), investigation details (Figure 18), chemical compound information (Figure 19), and sample history (Figure 20), and then specifies the data organization of a formatted data-tuple (doubly-delimited list) through the inclusion of data labels that include appropriate data attributes to maintain data relationships (Figure 21); for example, a mole fraction data series includes relational information to specify a compound being measured and the phase in which it was measured (Figure 22). This development was discussed previously [29].

GHML is being used as the basis for the current web-dissemination technology development efforts underway by the CODATA gas hydrate task group. It is expected that by providing a uniform model for data and metadata communication for the gas hydrates community, communication across the varied disciplines associated with these studies can be improved and scientific progress in the field can be facilitated.

### **World Wide Web-based Data Dissemination**

In addition to collaboration with the CODATA data sharing efforts described above, a web-based interface specific to the Clathrate Hydrate Physical Property Database was developed to guarantee free and open access to the data resources upon completion of primary development, independent of CODATA developmental progress. The basis for the web interface is Google Web Toolkit [30], an open source set of tools that generate complex JavaScript front-end applications from Java source code. This set of tools was used to couple server-side database interaction to a client-side user interface built with the Ext JS [31] graphical JavaScript libraries and coupled together with open-source GWT-Ext libraries [32].

This choice of technologies allows for an interactive user experience more akin to a desktop application than with traditional web technologies. Traditional web-based database interfaces support simple search and display capabilities and require additional network transactions to modify data display. In contrast, the advanced technologies underpinning this web interface allow significantly more complex user interactions without an associated increase in required network traffic, usually the slowest step in any World Wide Web interaction. For example, the Clathrate Hydrate Physical Property Database web interface supports complex queries regarding combinations of chemical compounds through dynamically populated and easily searchable lists of chemical compounds (Figure 23). Once a user has specified search criteria, the client interface downloads the associated datasets to the client and no further server interactions are necessary. The retrieved datasets will dynamically populate a tree based structure (Figure 24), based upon chemical composition and literature source. The user can see information on the compounds involved in the systems of interest including two dimensional structural representation (Figure 25) and full citation information (Figure 26). The user can display the datasets of interest in interactive tables (Figure 27), that allow the user to sort data as desired. As well, all contained data can be displayed graphically (Figure 28) where the user can control axis scale, can select logarithmic or inverse scaling (Figure 29) and can relabel datasets or chart objects as desired (Figure 30). Finally, if the user's need requires he interact with the data in some way not supported by the interface, all datasets can be downloaded in a commonly-accessible format (Figure 31).

In addition to providing access to the recently compiled literature archive of clathrate hydrate physical properties, users can also freely access the 2002 scientific results from the Mallik Gas Hydrate Production Research Well Program, a novel exploration of the technical feasibility of natural gas production from a permafrost-based deposit of gas hydrates. This data is reproduced with explicit, written consent from the copyright holder (Natural Resources Canada). The web-based interface offers all interactive capabilities and access to all data available through the data viewing software distributed with the original data materials without the need to install any software or permanently download any materials. This includes access to all included study metadata as well as original tabulated data files (Figure 32).

## **Conclusions**

The TRC group has completed design, population, and publication to the World Wide Web of the Clathrate Hydrate Physical Property Database, NIST Standard Reference Database 156. This data resource is available on a free and open basis at the URI <http://gashydrates.nist.gov>. It contains well defined and critically evaluated experimentally-derived thermophysical and structural data for clathrate hydrates, including the complex systems associated with clathrate hydrates of natural gas. The design of this database is derived from the structure of the SOURCE Data Archive and includes several novel extensions to represent complex gas hydrates data sets in a well-constrained manner. To support of this new data resource, new computing hardware was obtained and configured to serve the associated data needs. This interface also provides access to the 2002 scientific results of the Mallik Gas Hydrate Production Research Well Program, reproduced with permission, and provides all capabilities of the viewing software previously developed for dissemination with physical copies of the dataset.

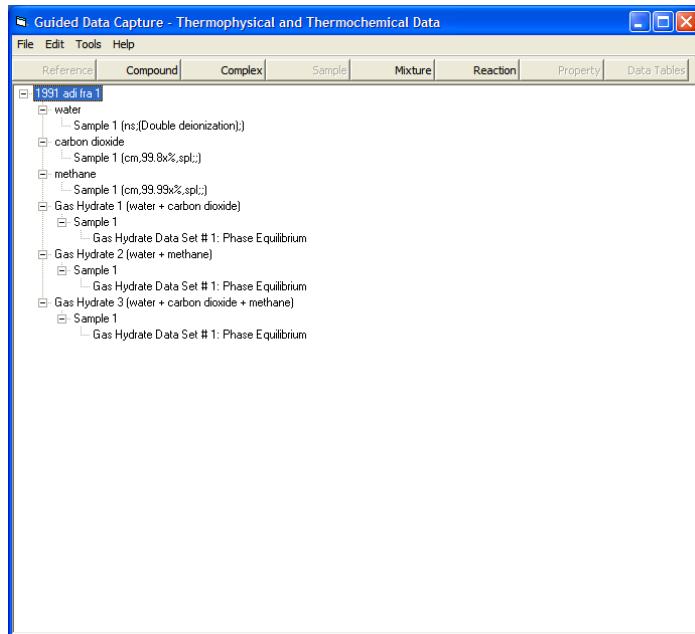
Population of this database was supported by a gas hydrates library collected by TRC staff. Presently at nearly 6,000 documents, this literature archive will continue to grow as new materials become available and will continue to serve as a source of new datasets to be added to the Clathrate Hydrate Physical Property Database. This literature archive was cataloged using an internal database and experimental datasets were extracted using specially designed Guided Data Capture software, facilitating data collection by individuals who are technically competent but lack the broad experience usually required to dissect complex thermophysical systems.

The gas hydrate markup language (GHMLv1.0) was modified significantly to meet the needs of the broadly-based gas hydrate community. The range of data sets it can represent was increased, support to represent citation information was added, and consistency with the IUPAC-standard ThermoML was included. This new GHML data format is actively being used in international data-sharing development efforts. It is expected that in the future, GHML will serve as a major tool for data exchange across the boundaries of traditional academic disciplines within the gas hydrate community.

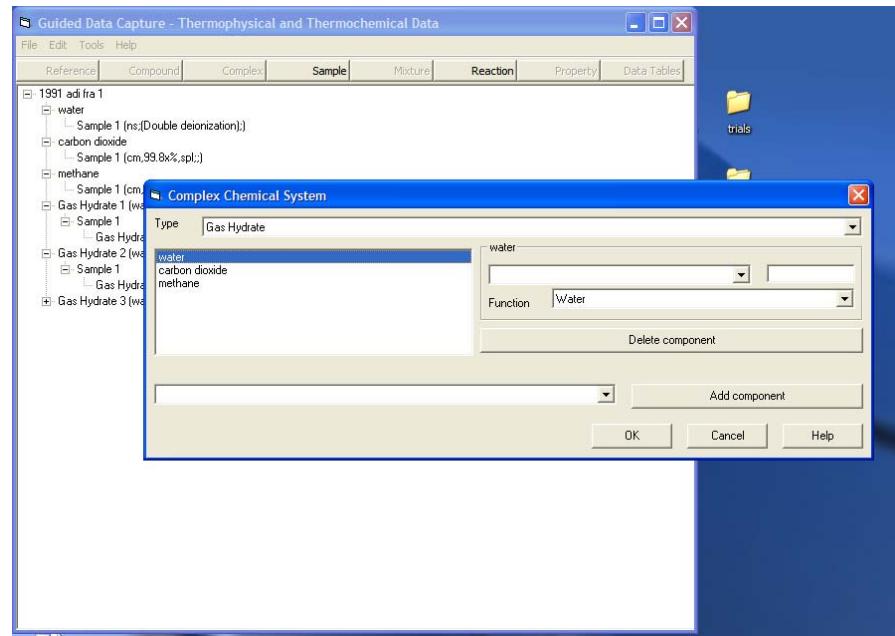
In support of this data collection, evaluation and dissemination project, all milestones have been met or exceeded.

**GRAPHICAL MATERIALS LIST**

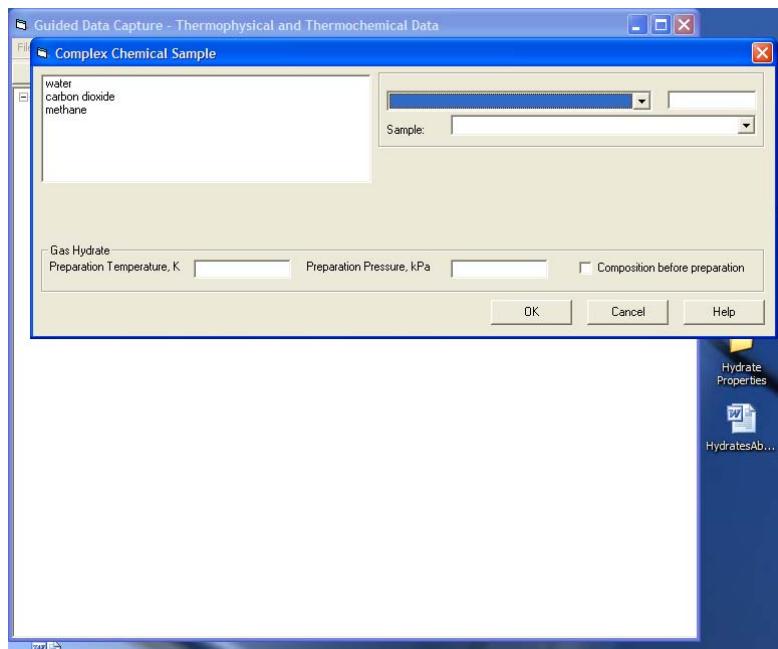
<b>Figure 1.</b> Screen capture of tree structure for a gas hydrate sample characterization within GDC...	17
<b>Figure 2.</b> Screen capture of GDC dialog for definition of a gas hydrate system.....	18
<b>Figure 3.</b> Screen capture of GDC dialog for definition of a gas hydrate sample .....	19
<b>Figure 4.</b> Screen capture of GDC dialog for defining phase equilibrium constraints and variables on a given set of phase equilibrium data .....	20
<b>Figure 5.</b> Screen capture of GDC dialog for entering tabulated data associated with a given set of phase equilibrium data .....	21
<b>Figure 6.</b> Screen capture of natively-generated graph of data entered into GDC tabulated data dialog .....	22
<b>Figure 7.</b> Screen capture of GDC dialog for defining the physical reaction associated with gas hydrate decomposition.....	23
<b>Figure 8.</b> Screen capture of GDC dialog for defining a type of bulk measurement and the associated measurement methodology.....	24
<b>Figure 9.</b> Screen capture of GDC dialog for defining the thermodynamic conditions under which a bulk measurement was performed .....	25
<b>Figure 10.</b> Screen capture of GDC dialog for entering tabulated data associated with a given set of bulk property data with automated reliability estimate .....	26
<b>Figure 11.</b> Screen capture of GDC dialog for storing crystallographic data, including space group, unit cell parameters and atom distribution.....	27
<b>Figure 12.</b> Schematic representation of new SOURCE table structure and gas-hydrate-relevant table substructure .....	28
<b>Figure 13.</b> SOURCE tables relevant to defining a specific gas hydrate sample, dependant upon the literature source of the data, the chemical compounds present and the compositional purity of the feed materials .....	29
<b>Figure 14.</b> SOURCE tables relevant to defining data from crystallographic studies, including atomic distribution if reported .....	30
<b>Figure 15.</b> SOURCE tables relevant for defining thermodynamic state and property data, including temperature, pressure and compositional information.....	31
<b>Figure 16.</b> Root element of GHML .....	32
<b>Figure 17.</b> GHML citation element, consistent with ThermoML.....	33
<b>Figure 18.</b> GHML investigation element .....	34
<b>Figure 19.</b> GHML compound element .....	35
<b>Figure 20.</b> GHML history element .....	36
<b>Figure 21.</b> GHML data element .....	37
<b>Figure 22.</b> Exemplar data category from GHML, specifically the ChemicalData subtype .....	38
<b>Figure 23.</b> Screenshot from web interface, demonstrating a search for a chemical system, including compound filtering.....	39
<b>Figure 24.</b> Screenshot from web interface, demonstrating search results for chemical system including dynamically populated tree .....	40
<b>Figure 25.</b> Screenshot from web interface, demonstrating display of compound information, associated with a given chemical system .....	41
<b>Figure 26.</b> Screenshot from web interface, demonstrating display of full citation information for a given data set .....	42
<b>Figure 27.</b> Screenshot from web interface, demonstrating tabular display of data, including sorting capability .....	43
<b>Figure 28.</b> Screenshot from web interface, demonstrating Arrhenius plot of propane-water phase equilibria .....	44
<b>Figure 29.</b> Screenshot from web interface, demonstrating chart options menu for native charting capability .....	45
<b>Figure 30.</b> Screenshot from web interface, demonstrating dataset options menu for native charting capability .....	46
<b>Figure 31.</b> Screenshot from web interface, demonstrating downloaded table file with original table within web viewer context .....	47
<b>Figure 32.</b> Screenshot from web interface, demonstrating table from 2002 Mallik scientific results with data methodology information displayed .....	48



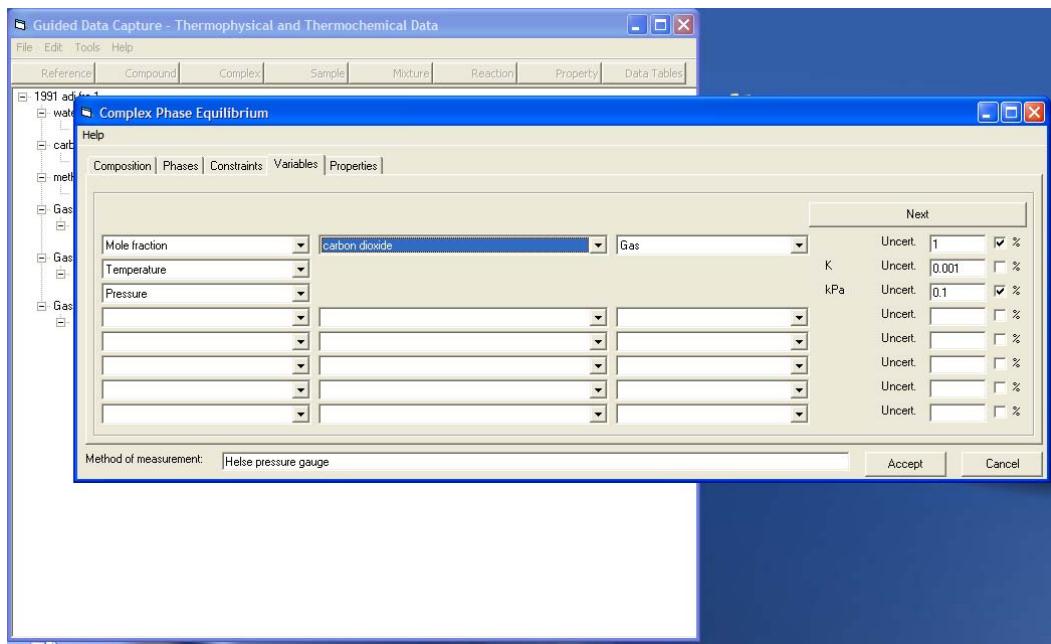
**Figure 1.** Screen capture of tree structure for a gas hydrate sample characterization within GDC



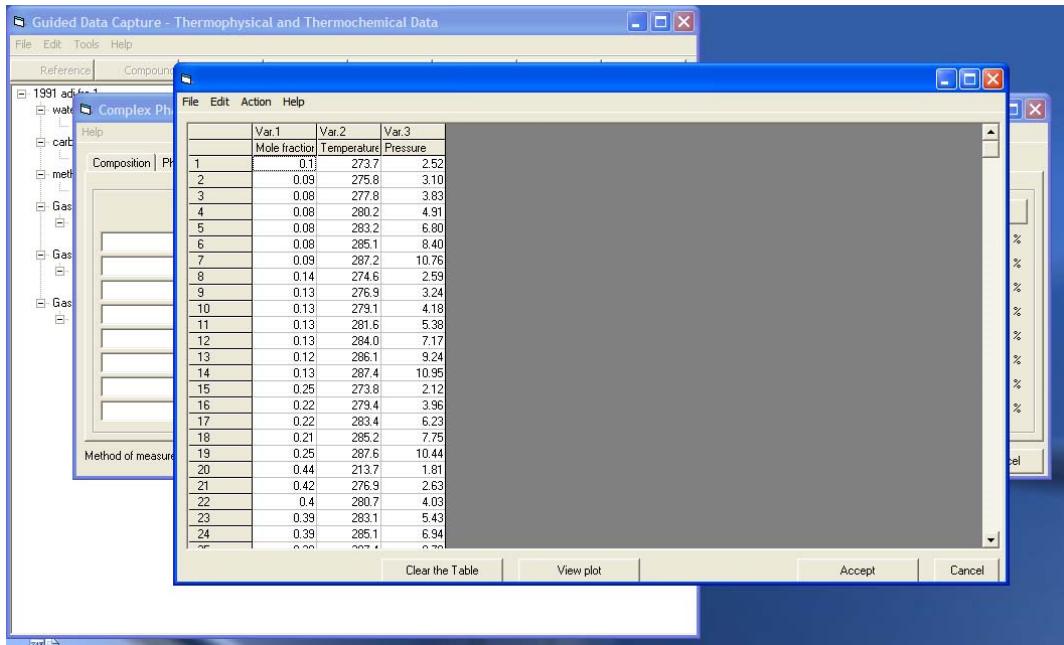
**Figure 2. Screen capture of GDC dialog for definition of a gas hydrate system**



**Figure 3. Screen capture of GDC dialog for definition of a gas hydrate sample**



**Figure 4. Screen capture of GDC dialog for defining phase equilibrium constraints and variables on a given set of phase equilibrium data**



**Figure 5. Screen capture of GDC dialog for entering tabulated data associated with a given set of phase equilibrium data**

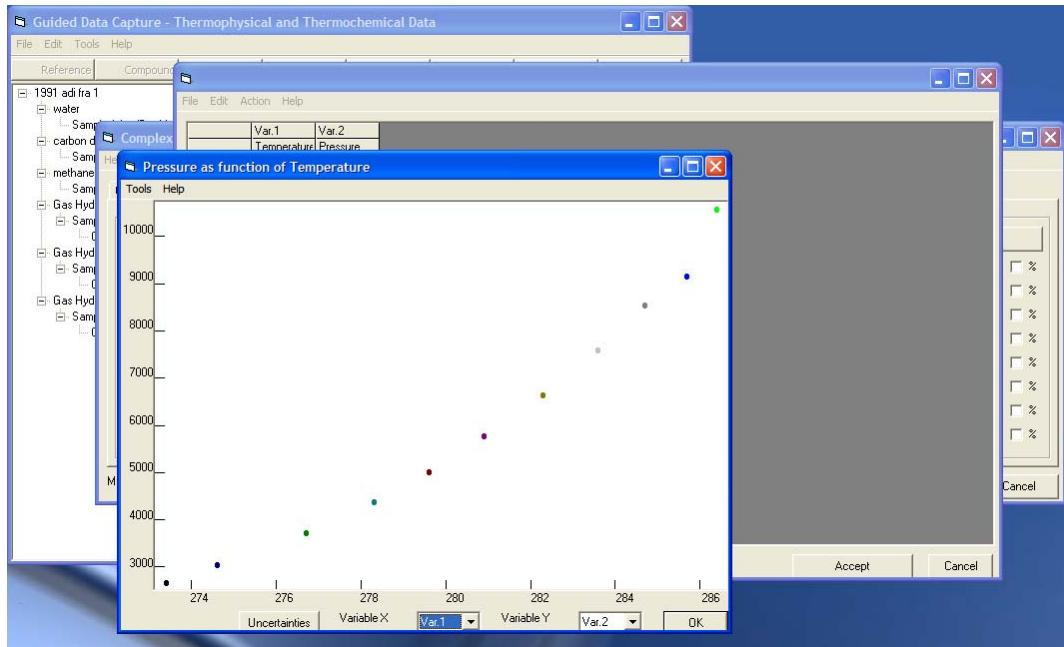
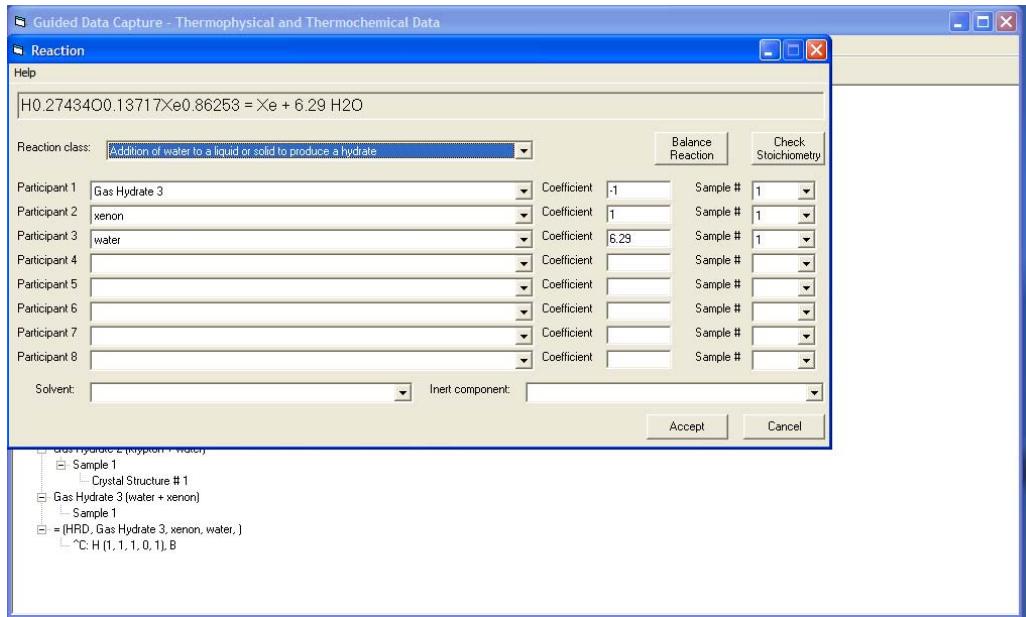
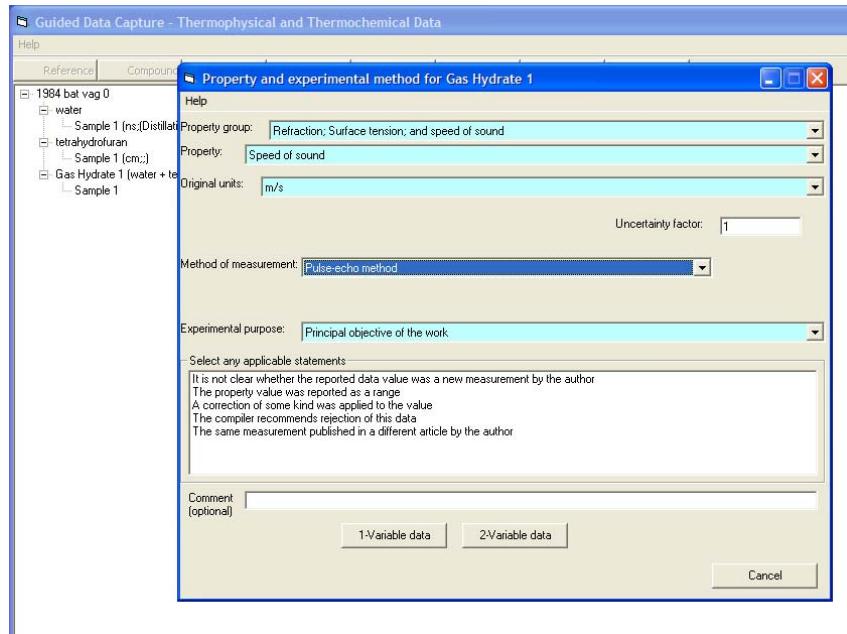


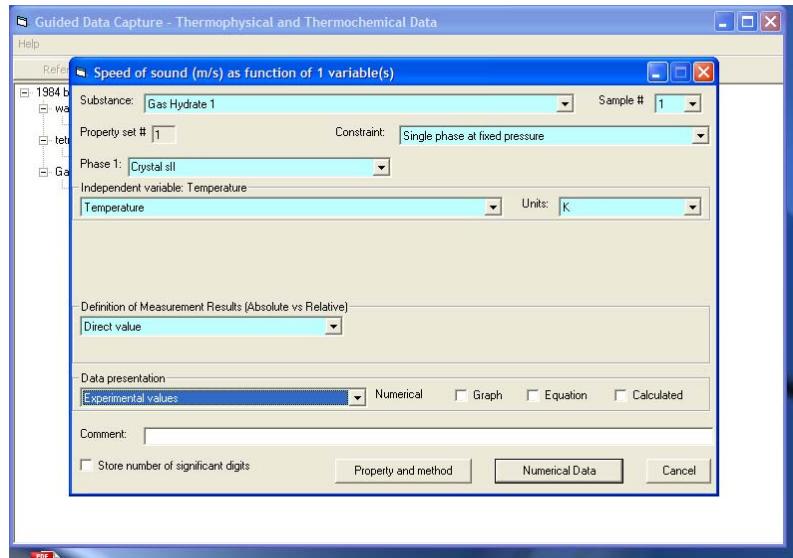
Figure 6. Screen capture of natively-generated graph of data entered into GDC tabulated data dialog



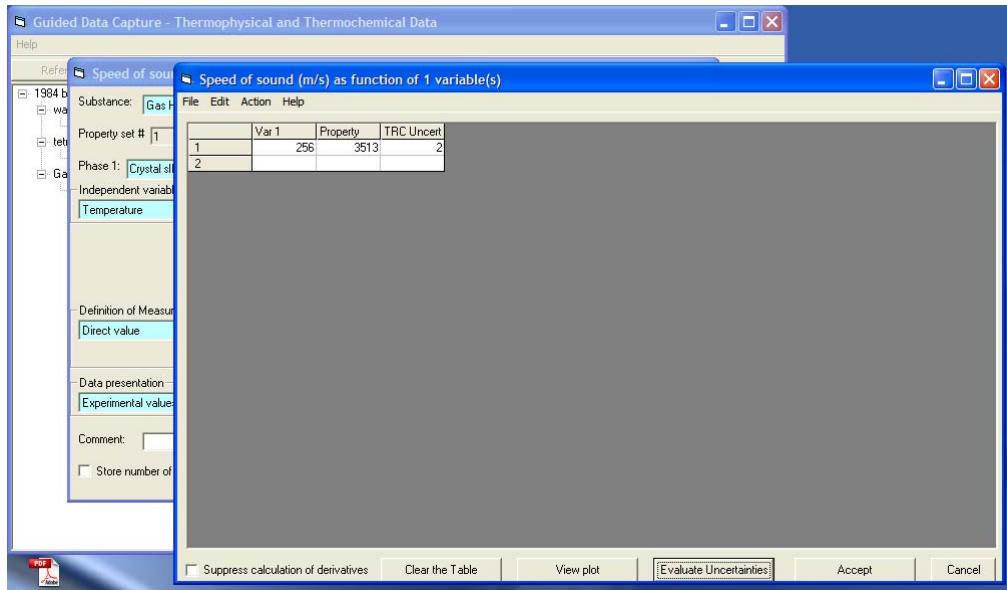
**Figure 7. Screen capture of GDC dialog for defining the physical reaction associated with gas hydrate decomposition**



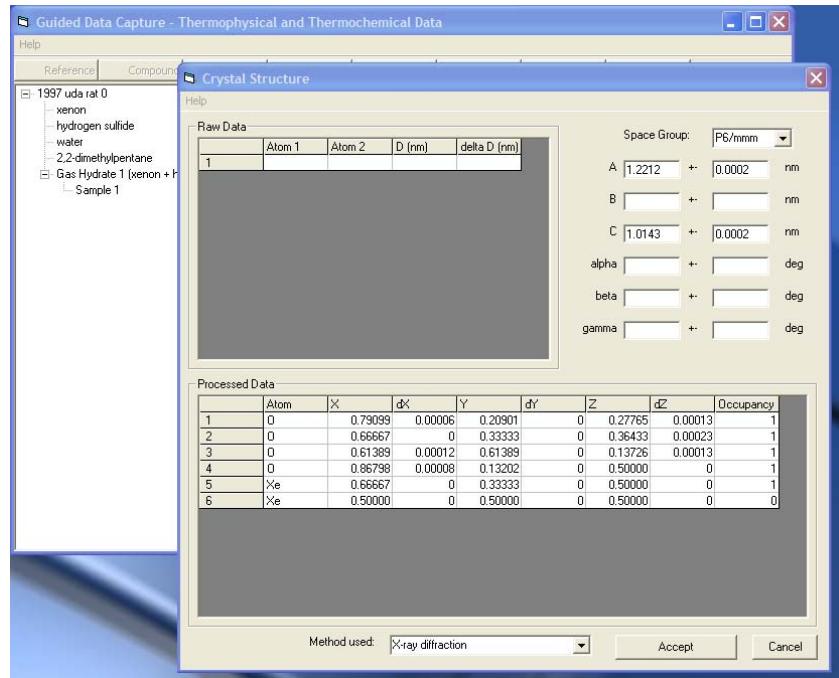
**Figure 8. Screen capture of GDC dialog for defining a type of bulk measurement and the associated measurement methodology**



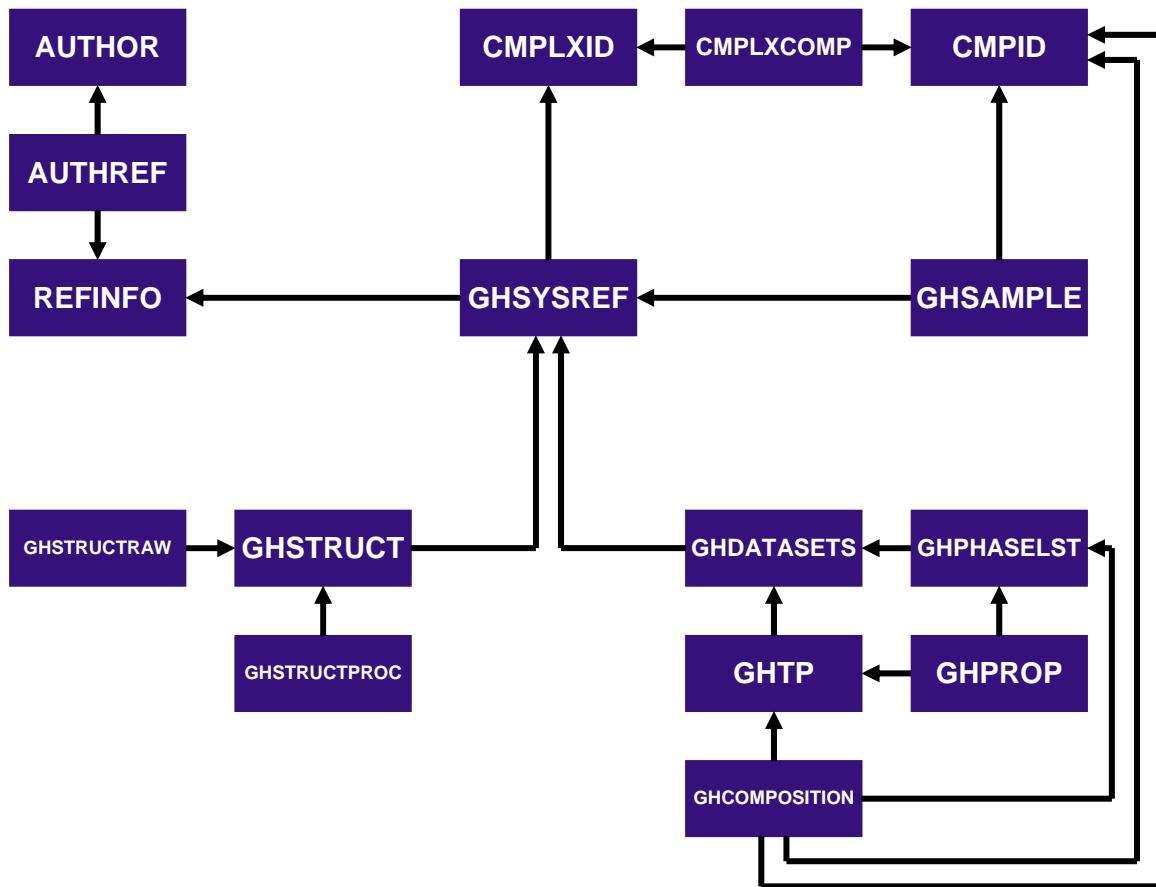
**Figure 9.** Screen capture of GDC dialog for defining the thermodynamic conditions under which a bulk measurement was performed



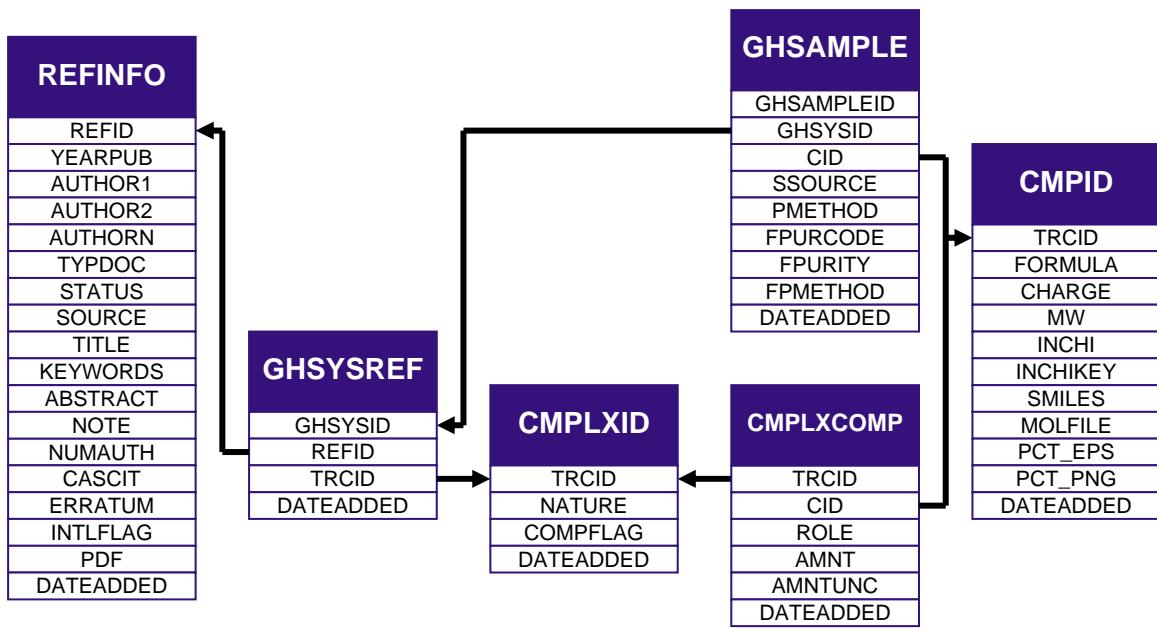
**Figure 10. Screen capture of GDC dialog for entering tabulated data associated with a given set of bulk property data with automated reliability estimate**



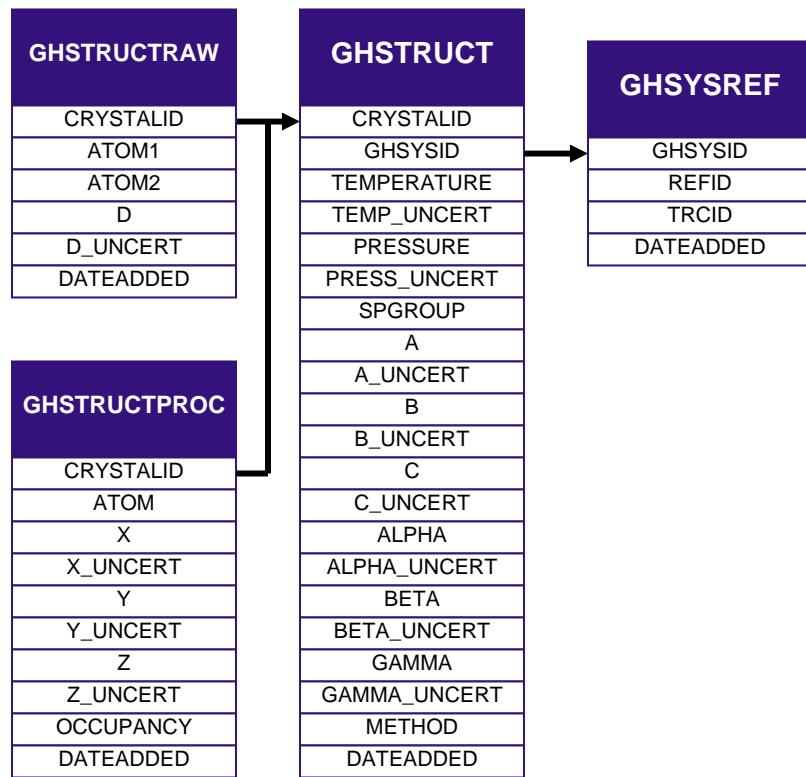
**Figure 11. Screen capture of GDC dialog for storing crystallographic data, including space group, unit cell parameters and atom distribution**



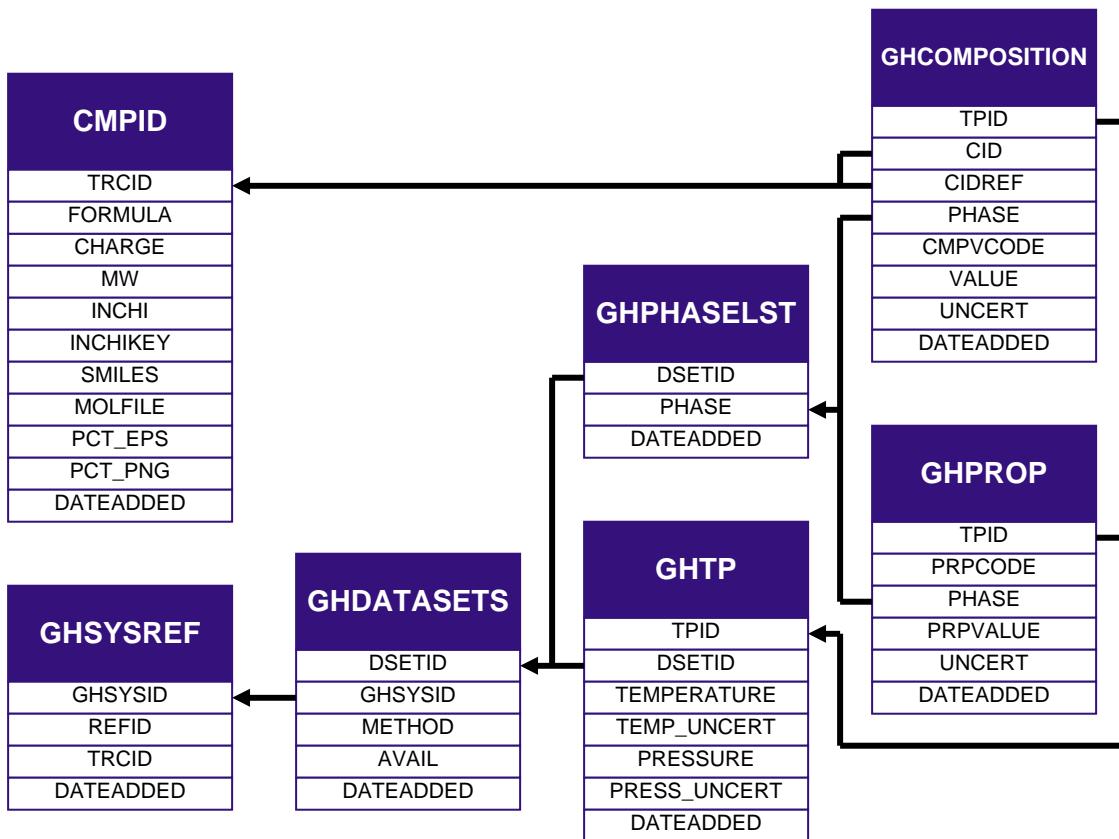
**Figure 12. Schematic representation of new SOURCE table structure and gas-hydrate-relevant table substructure**



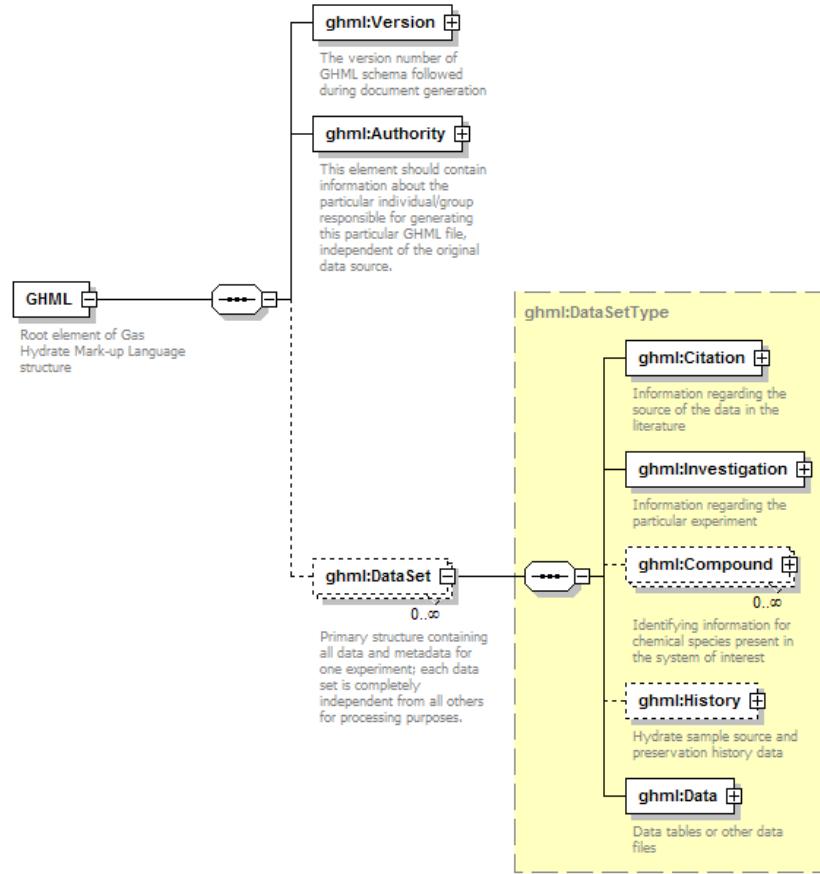
**Figure 13. SOURCE tables relevant to defining a specific gas hydrate sample, dependant upon the literature source of the data, the chemical compounds present and the compositional purity of the feed materials**



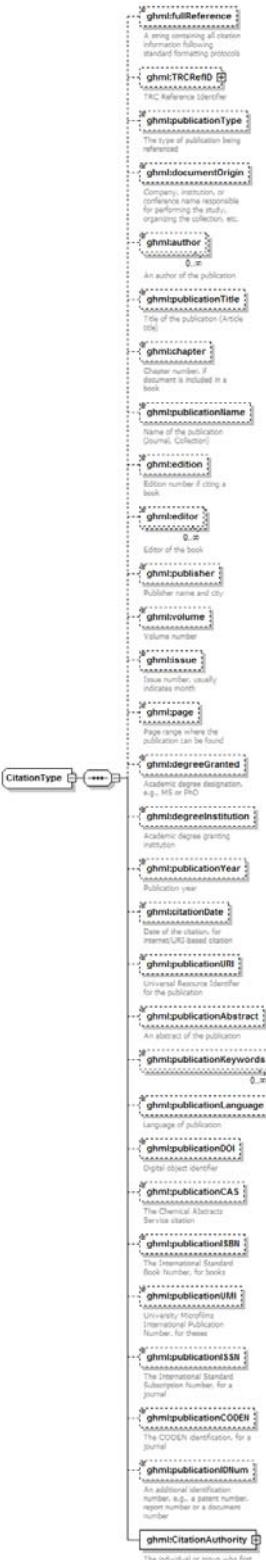
**Figure 14.** SOURCE tables relevant to defining data from crystallographic studies, including atomic distribution if reported



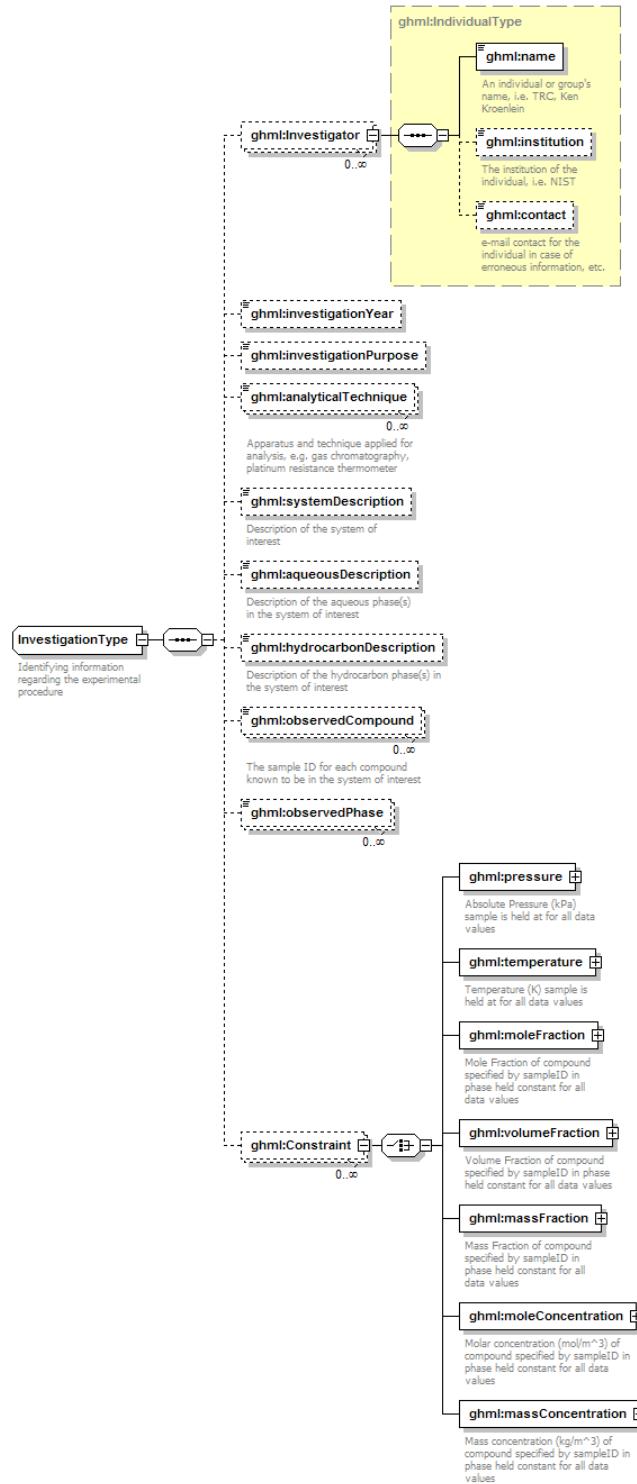
**Figure 15. SOURCE tables relevant for defining thermodynamic state and property data, including temperature, pressure and compositional information**



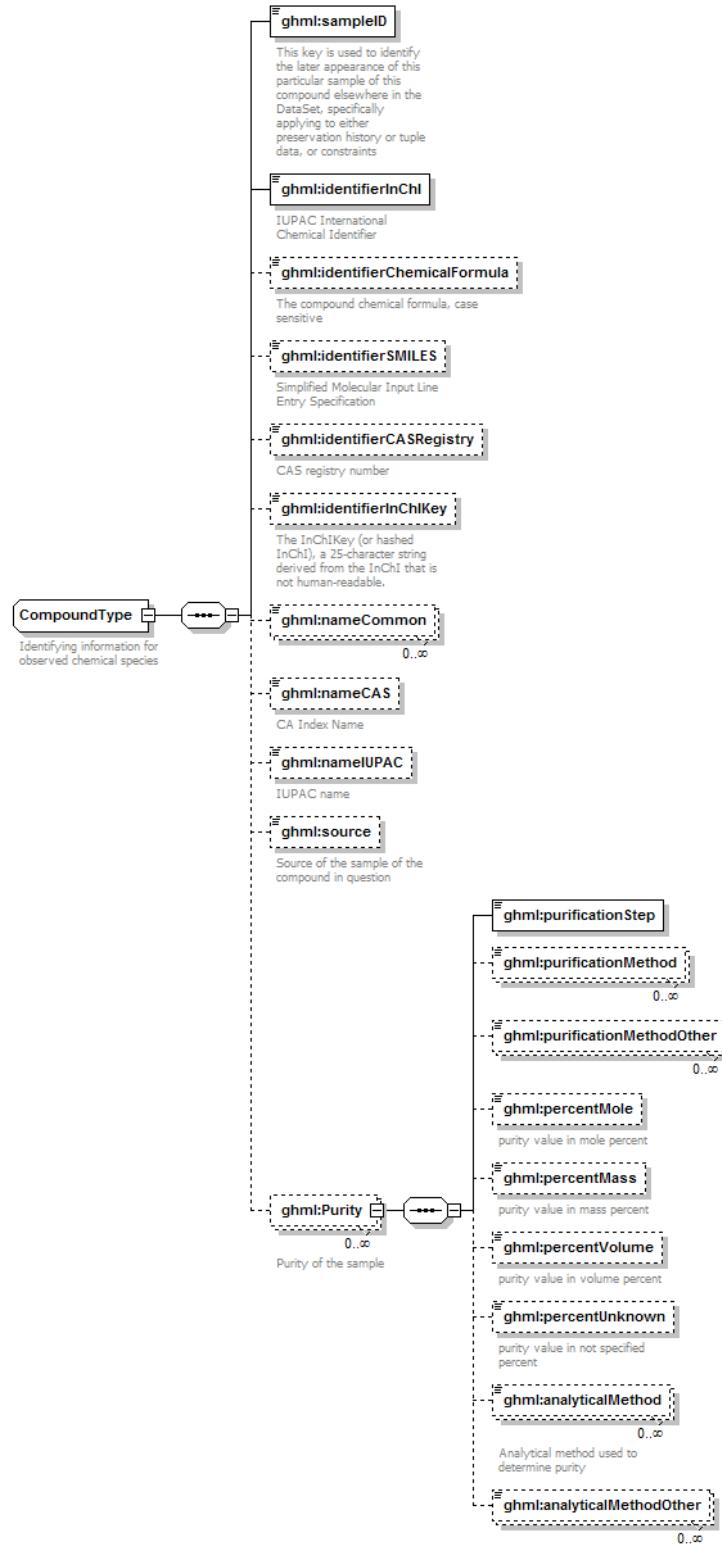
**Figure 16. Root element of GHML**

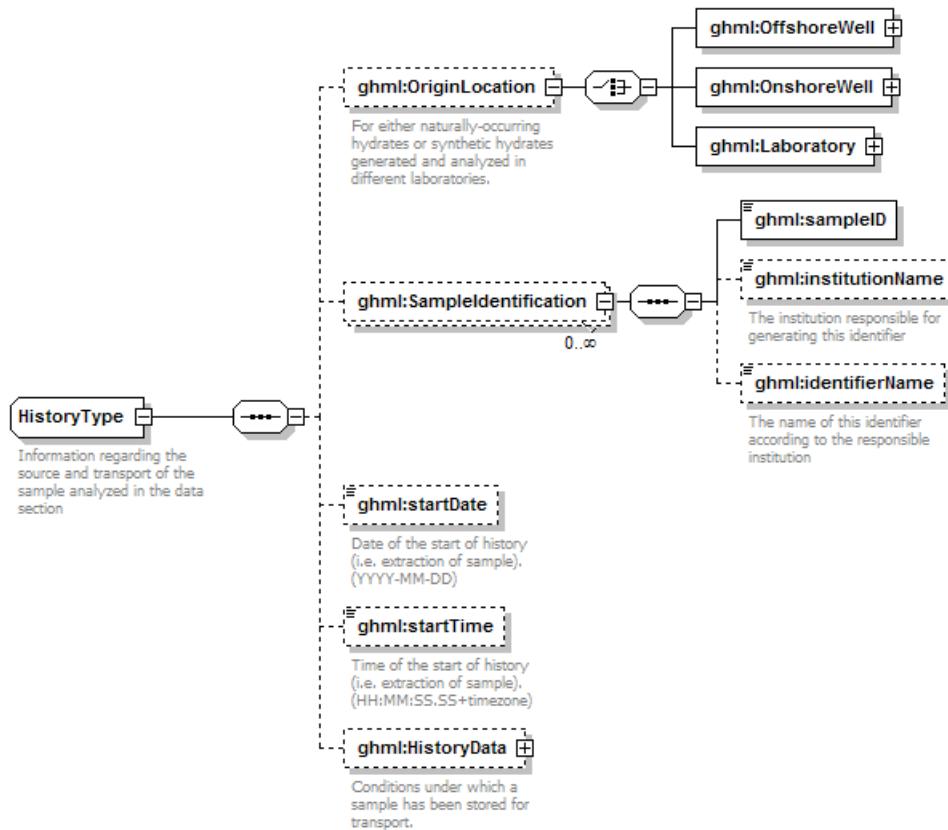


**Figure 17. GHML citation element, consistent with ThermoML**

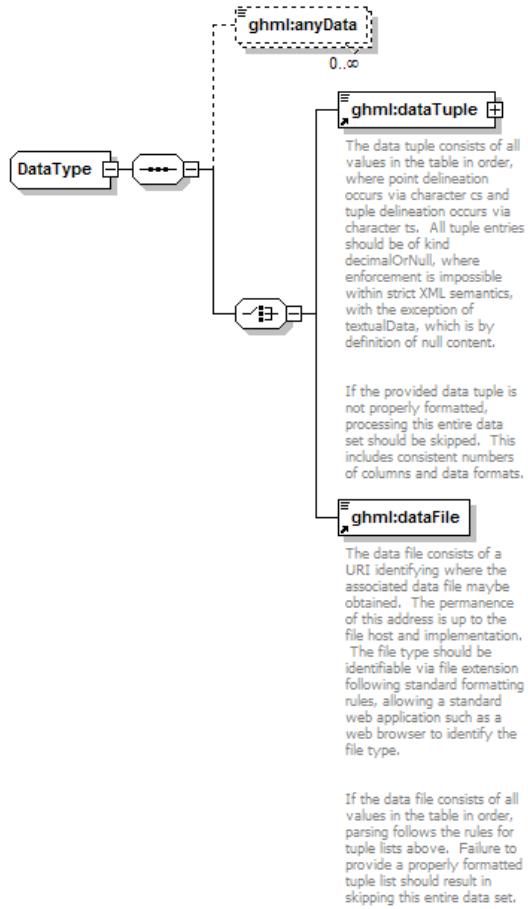


**Figure 18. GHML investigation element**

**Figure 19. GHML compound element**



**Figure 20. GHML history element**



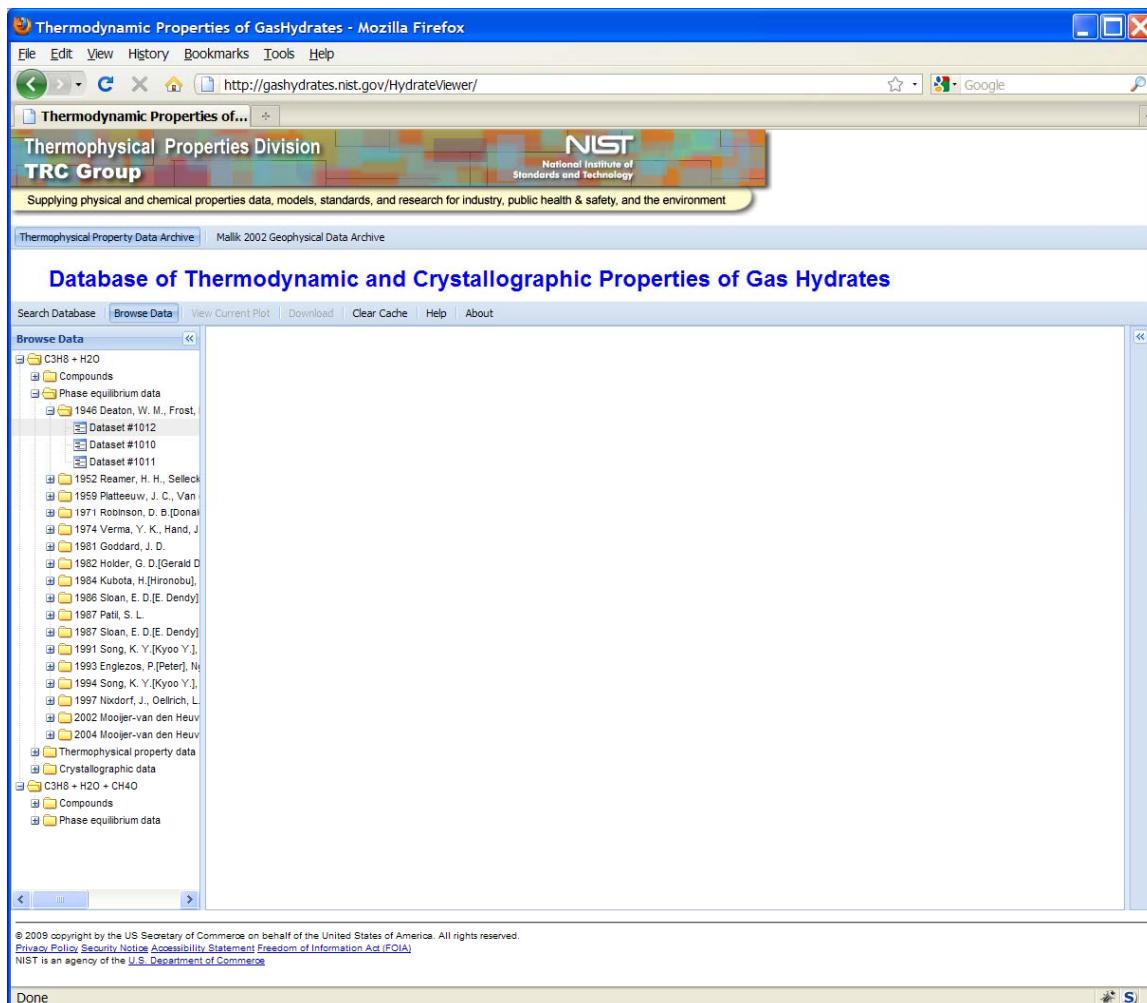
**Figure 21. GHML data element**

```
<element name="chemicalData">
  <annotation>
    <documentation>Broad Categories describing chemical properties of a sample. See:</documentation>
  </annotation>
  <complexType>
    <simpleContent>
      <extension base="ghml:chemicalDataType">
        <attribute name="columnNumber" type="integer" use="required"/>
        <attribute name="descriptor" type="string" use="optional"/>
        <attribute name="sampleID" type="integer" use="optional"/>
        <attribute name="phaseID" type="ghml:phaseEnumeration" use="required"/>
      </extension>
    </simpleContent>
  </complexType>
</element>
<simpleType name="chemicalDataType">
  <restriction base="string">
    <enumeration value="Mole Fraction"/>
    <enumeration value="Mass Fraction"/>
    <enumeration value="Volume Fraction"/>
    <enumeration value="Mass Concentration (kg/m^3)"/>
    <enumeration value="Molar Concentration (mol/m^3)"/>
    <enumeration value="Salinity (ppt)"/>
    <enumeration value="Salinity Class"/>
  </restriction>
</simpleType>
```

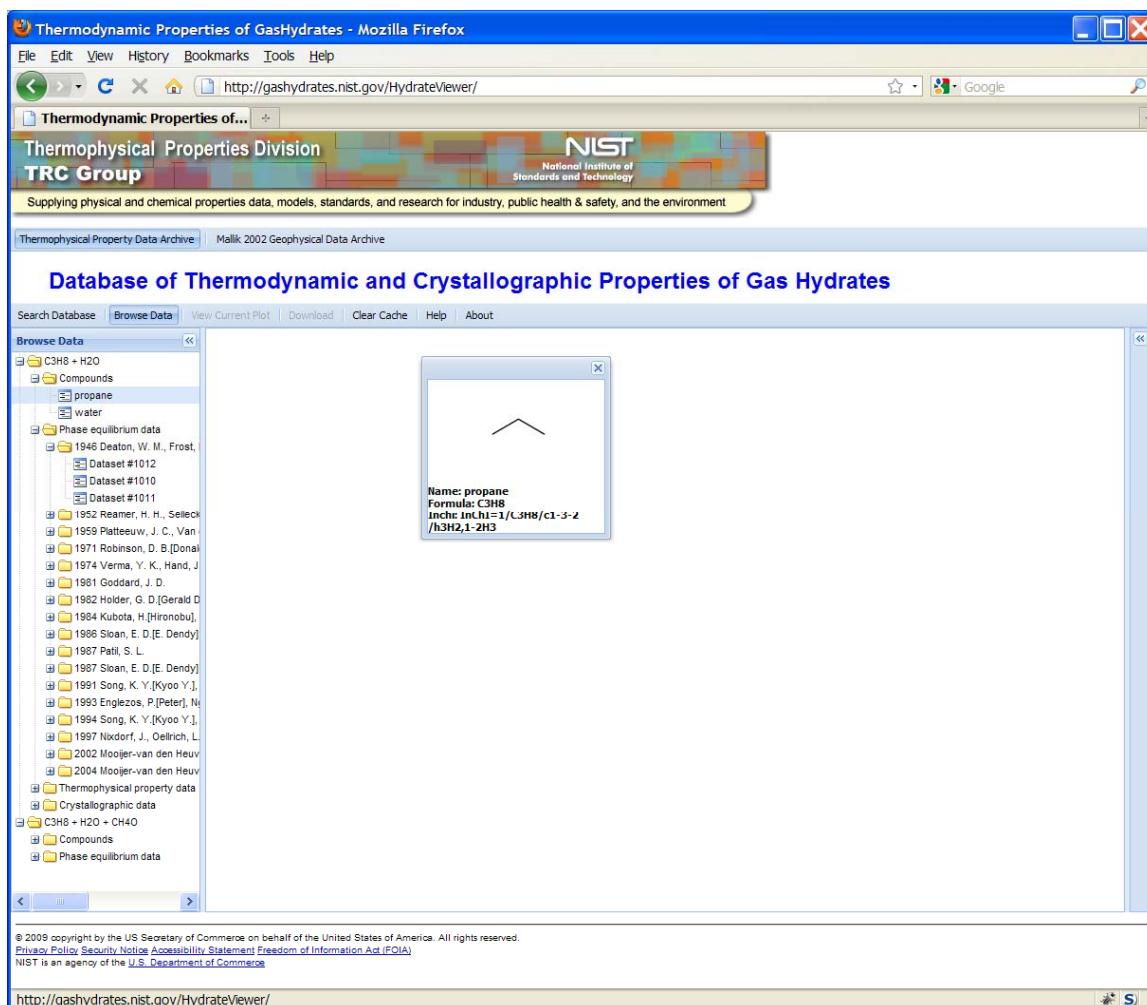
Figure 22. Exemplar data category from GHML, specifically the ChemicalData subtype

The screenshot shows a Firefox browser window displaying the NIST Gas Hydrate Research Database. The title bar reads "Thermodynamic Properties of GasHydrates - Mozilla Firefox". The address bar shows the URL "http://gashydrates.nist.gov/HydrateViewer/". The main content area features the NIST logo and the text "Supplying physical and chemical properties data, models, standards, and research for industry, public health & safety, and the environment". Below this is the heading "Database of Thermodynamic and Crystallographic Properties of Gas Hydrates". A "Search Database" sidebar on the left lists options: Search by compound, Search by citation, and Search by property. The main search interface contains two tables: "Available Compounds" and "Compounds (Guest, Host, Inhibitor...)". The "Available Compounds" table includes entries like C3H8 (propane), C3H6 (cyclopropane), C3H6 (propylene), C3H8O (2-propanol), C3H6O (1,3-propylene oxide), C3H6O (acetone), C3H6O (propylene oxide), C3H6O2 (1,3-dioxolane), C3H3NO (1,2-oxazole), C3H8O3 (1,2,3-propanetriol), C3H7NO (N,N-dimethylformamide), and C3H6O3 ((2S)-2-hydroxypropanoic acid). The "Compounds (Guest, Host, Inhibitor...)" table includes H2O (water), CH4O (methanol), and C3H8 (propane). Between the tables are buttons for "">>>" and "<<". Below the tables are filtering options: "Filter:" set to "C3", "Filter by formula", "Return systems containing exactly the selected compounds", "Save previous results", and "Submit search". At the bottom of the page, there is a copyright notice: "© 2009 copyright by the US Secretary of Commerce on behalf of the United States of America. All rights reserved. Privacy Policy Security Notice Accessibility Statement Freedom of Information Act (FOIA) NIST is an agency of the U.S. Department of Commerce". The status bar at the bottom of the browser window says "Transferring data from gashydrates.nist.gov...".

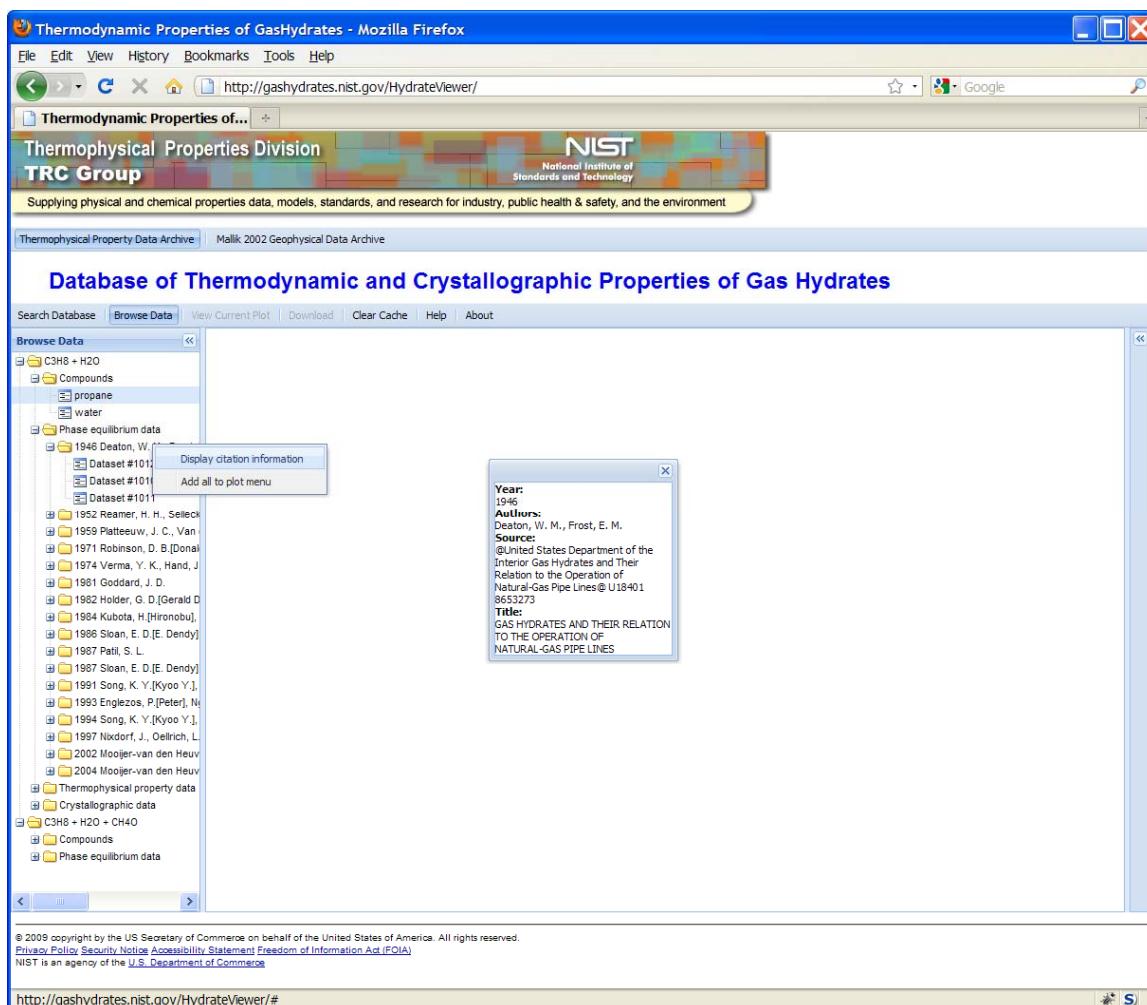
**Figure 23. Screenshot from web interface, demonstrating a search for a chemical system, including compound filtering**



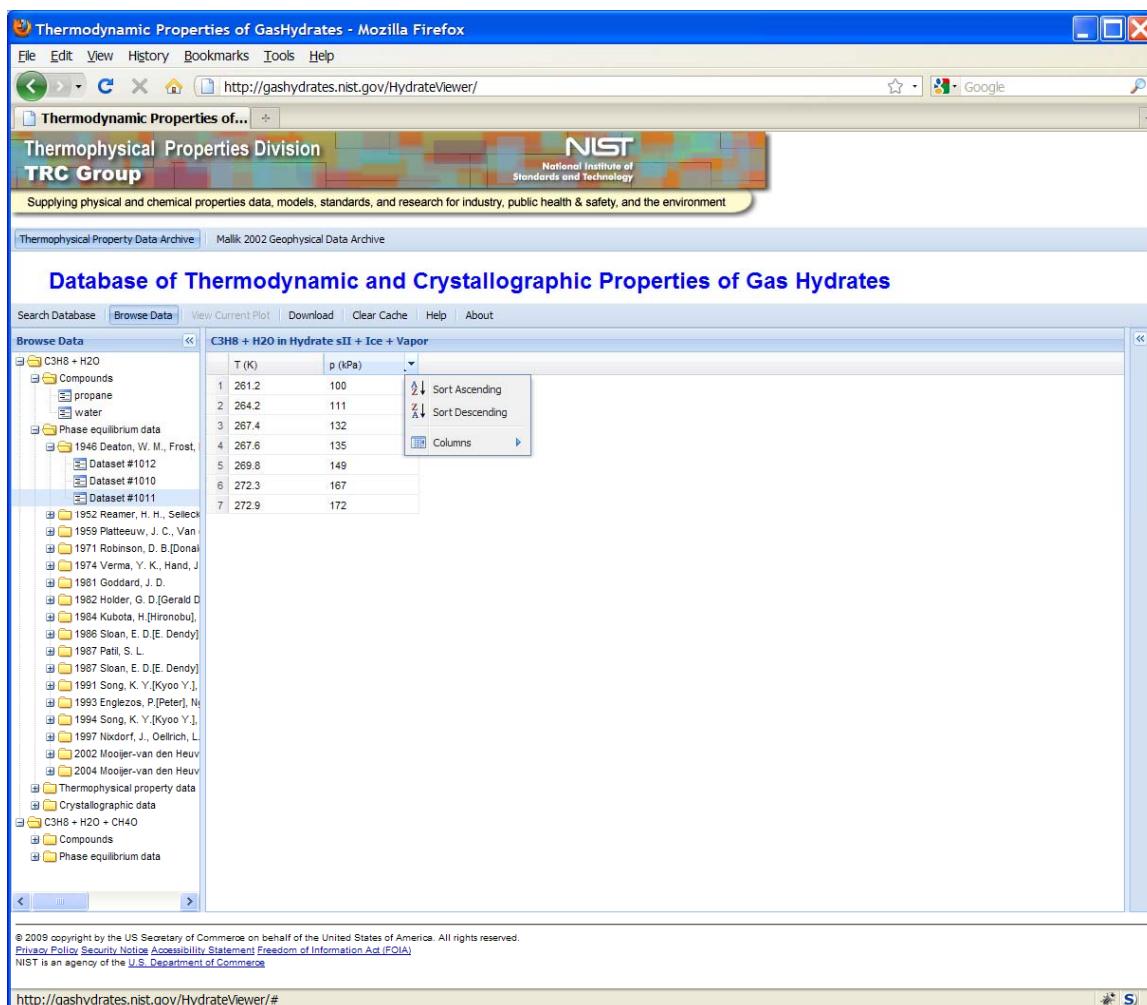
**Figure 24. Screenshot from web interface, demonstrating search results for chemical system including dynamically populated tree**



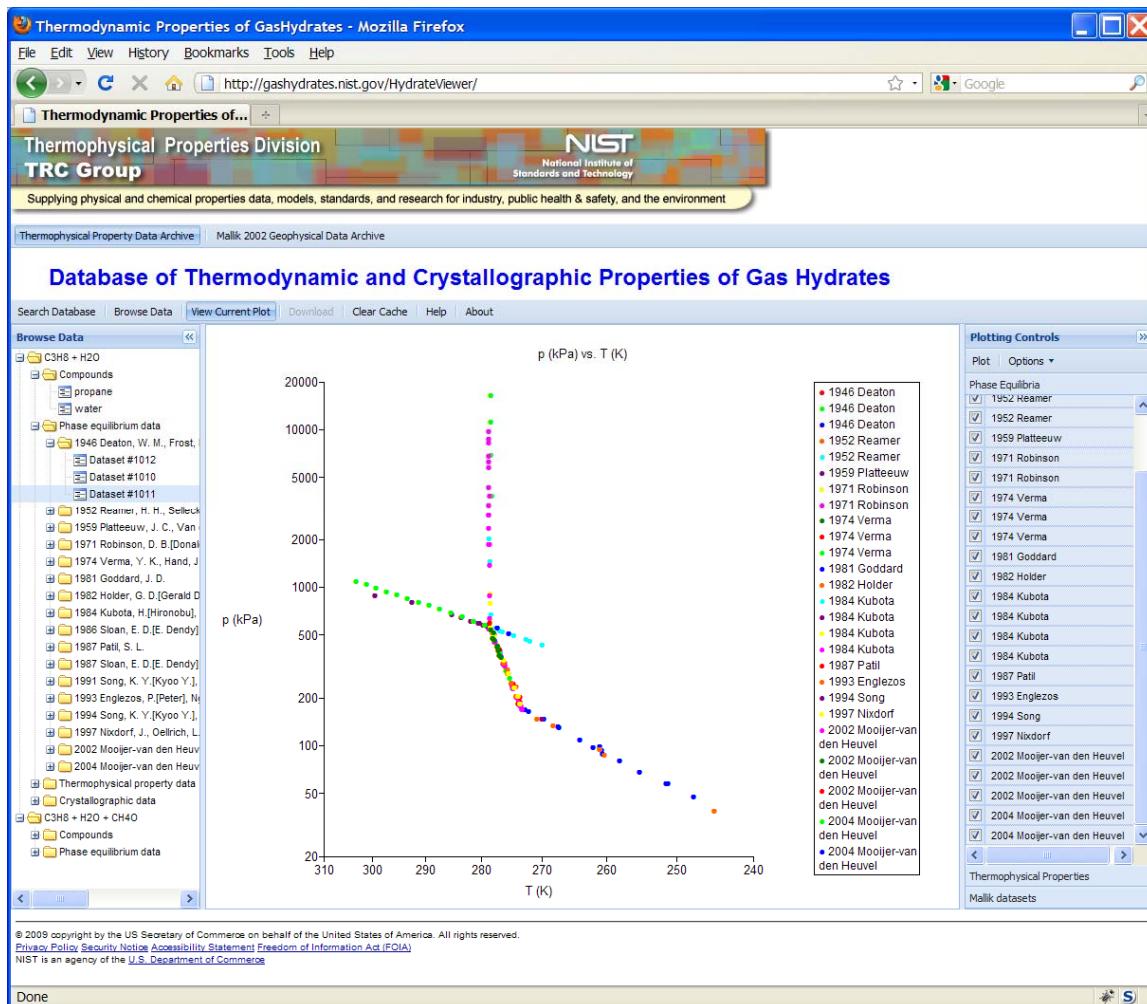
**Figure 25. Screenshot from web interface, demonstrating display of compound information, associated with a given chemical system**



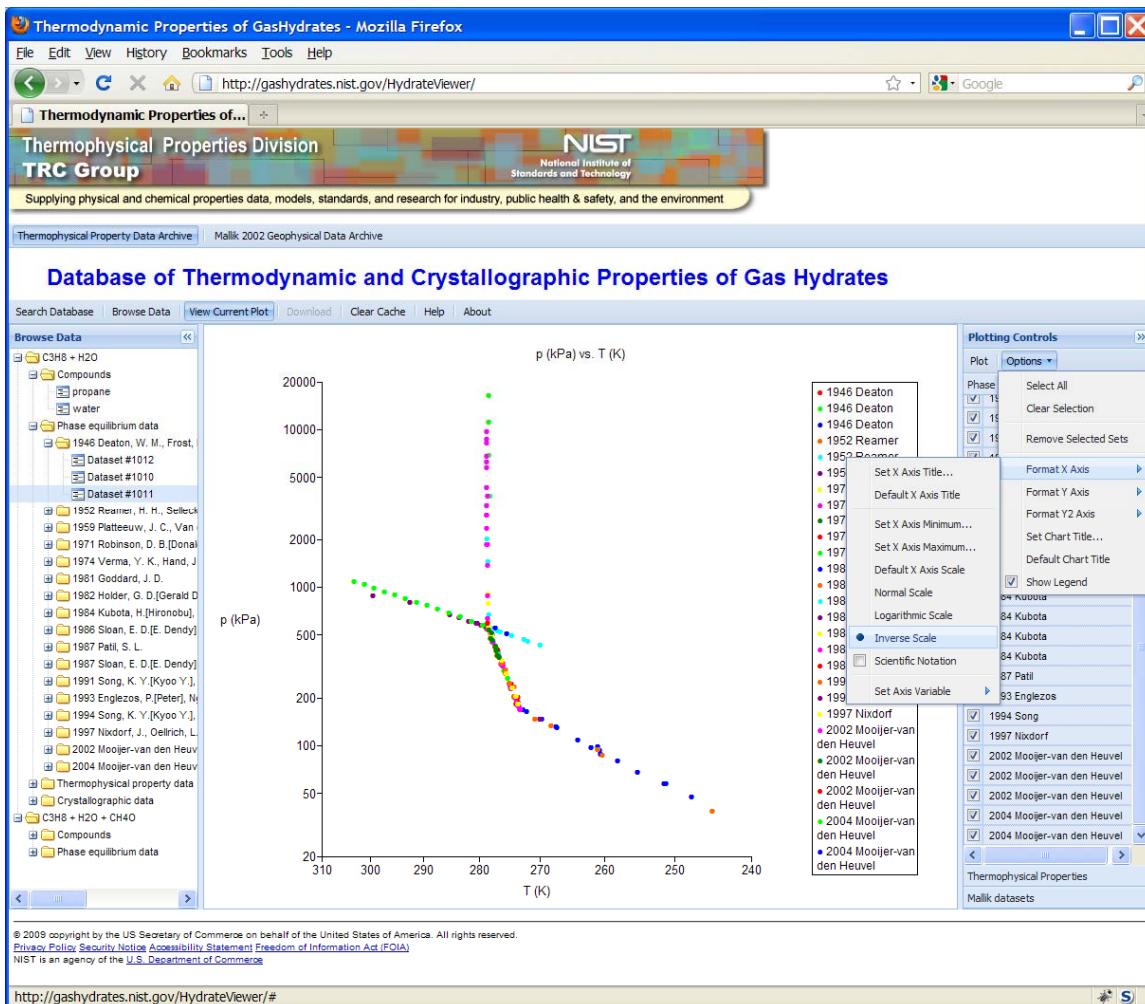
**Figure 26. Screenshot from web interface, demonstrating display of full citation information for a given data set**



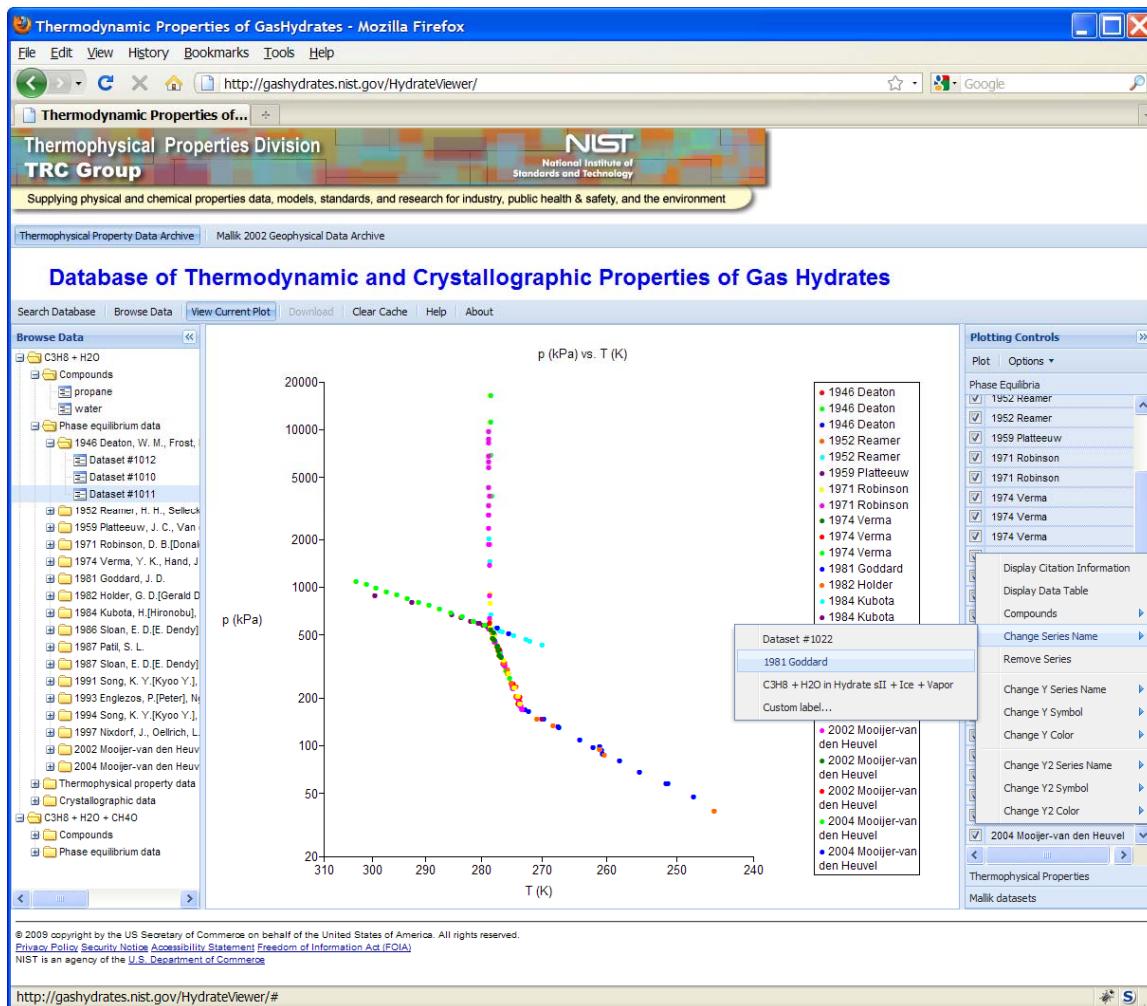
**Figure 27. Screenshot from web interface, demonstrating tabular display of data, including sorting capability**



**Figure 28. Screenshot from web interface, demonstrating Arrhenius plot of propane-water phase equilibria**



**Figure 29. Screenshot from web interface, demonstrating chart options menu for native charting capability**



**Figure 30. Screenshot from web interface, demonstrating dataset options menu for native charting capability**

T (K)	p (kPa)
273.4	181
273.6	192
273.8	193
273.9	205
275.5	291
276.9	402

**Figure 31.** Screenshot from web interface, demonstrating downloaded table file with original table within web viewer context

The screenshot shows a Mozilla Firefox browser window displaying the NIST Gas Hydrate Research Database. The title bar reads "Thermodynamic Properties of GasHydrates - Mozilla Firefox". The address bar shows the URL <http://gashydrates.nist.gov/HydrateViewer/>. The main content area features the "Thermophysical Properties Division TRC Group" logo and the NIST Standards and Technology logo. Below the logo, a banner states: "Supplying physical and chemical properties data, models, standards, and research for industry, public health & safety, and the environment". The header menu includes File, Edit, View, History, Bookmarks, Tools, and Help. Sub-menus for Thermophysical Property Data Archive and Malik 2002 Geophysical Data Archive are visible.

The central part of the interface is titled "Scientific Results from the Mallik 2002 Gas Hydrate Production Research Well Program". A "Tabular Data" tab is selected, showing a table titled "DTS Temperature (0.25 m) (Hennings)". The table contains data for various depths (e.g., 6.45, 6.7, 8.47, 8.72, 8.98, 9.23, 9.48, 9.73, 9.99, 10.24, 10.49, 10.74, 11, 11.25) with corresponding temperatures in degrees Celsius. A context menu is open over the table, with options like "Display Table", "Add to plot menu", "Display citation information", "Display methodology", and "Display contact information".

A "Plotting Controls" sidebar on the right includes options for Plot, Options, Phase Equilibria, Thermophysical Properties, and Malik datasets. A "Methods" pop-up window is displayed, providing technical details about the DTS methodology used in the study. The bottom of the page includes copyright information for the Geological Survey of Canada and NIST, and a footer with links to Privacy Policy, Security Notice, Accessibility Statement, Freedom of Information Act (FOIA), and NIST as an agency of the U.S. Department of Commerce.

**Figure 32. Screenshot from web interface, demonstrating table from 2002 Mallik scientific results with data methodology information displayed**

**REFERENCES**

- [1] Sloan ED. *Introductory overview: Hydrate knowledge development*. American Mineralogist 2004;89:1155-1161.
- [2] *Methane Hydrate Research and Development Act of 2000* 30 United States Code 1902 note; Public Law 106-193, 2000.
- [3] Milkov AV. *Global estimates of hydrate-bound gas in marine sediments: how much is really out there?* Earth-Science Reviews 2004;66(3-4):183-197.
- [4] Klauda JB and Sandler SI. *Global Distribution of Methane Hydrate in Ocean Sediment*. Energy & Fuels 2005; 19:459-470.
- [5] Radler M. *World crude and natural gas reserves rebound in 2000*. Oil & Gas Journal 2000;98(51):121-123.
- [6] Reagan MT and Moridis GJ. *Oceanic gas hydrate instability and dissociation under climate change scenarios*. Geophysical Research Letters 2007;34:L22709.
- [7] Kvenvolden KA. *Potential Effects of Gas Hydrate on Human Welfare*. Proceedings of the National Academy of Sciences of the United States of America 1999;96(7):3420-3426.
- [8] Berner RA. *Examination of Hypotheses for the Permo-Triassic Boundary Extinction by Carbon Cycle Modeling*. Proceedings of the National Academy of Sciences of the United States of America 2002;99(7):4172-4177.
- [9] Padden M, Weissert H and de Rafeis M. *Evidence for Late Jurassic release of methane from gas hydrate*. Geology 2001;29(3):223-226.
- [10] Dickens GR. *Carbon addition and removal during the Late Palaeocene Thermal Maximum: basic theory with a preliminary treatment of the isotope record at ODP Site 1051, Blake Nose*. Geological Society, London, Special Publications 2001;183:293-305.
- [11] TRC Group – NIST, Thermophysical Properties Division. 2007. [online]. [Accessed 30th September 2008]. <http://www.trc.nist.gov/>.
- [12] Frenkel M, Dong Q, Wilhoit RC and Hall KR. *TRC SOURCE Database: A Unique Tool for Automatic Production of Data Compilations*. International Journal of Thermophysics 2001;22(1):215-226.
- [13] Yan X, Dong Q, Frenkel M and Hall KR. *Window-Based Applications of TRC Databases: Structure and Internet Distribution*. International Journal of Thermophysics 2001;22(1):227-241.
- [14] Frenkel M, Chirico RD, Diky V, Yan X, Dong Q and Muzny C. *ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept*. Journal of Chemical Information and Modeling 2005;45(4):816-838.
- [15] *ThermoData Engine*. 2008. [online]. [Accessed 30th September 2008]. <http://www.trc.nist.gov/tde.html>.
- [16] Wilhoit RC and Marsh KN. *Future Directions for Data Compilations*. International Journal of Thermophysics 1999;20(1):247-255.
- [17] Diky VV, Chirico RD, Wilhoit RC, Dong Q and Frenkel M. *Windows-Based Guided Data Capture Software for Mass-Scale Thermophysical and Thermochemical Property Data Collection*. Journal of Chemical Information and Computer Science 2003;43:15-24.
- [18] *Guided Data Capture*. 2007. [online]. [Accessed 30th September 2008]. <http://www.trc.nist.gov/GDC.html>.
- [19] Frenkel M, Chirico RD, Diky VV, Dong Q, Marsh KN, Dymond JH, Wakeham WA, Stein SE, Königsberger E and Goodwin ARH. *XML-Based IUPAC Standard for Experimental, Predicted, and Critically Evaluated Thermodynamic Property Data Storage and Capture (ThermoML)*. Pure and Applied Chemistry 2006;78(3):541-612.
- [20] International Union of Pure and Applied Chemistry. 2006. *International Union of Pure and Applied Chemistry* [online]. [Accessed 30th September 2008]. <http://www.iupac.org/projects/2002/2002-055-3-024.html>.
- [21] *ThermoML*. [online]. [Accessed 30th September 2008].

- http://www.trc.nist.gov/ThermoML.html.
- [22] Sloan D, Kuznetsov F, Lal K, Loewner R, Makogon Y, Moridis G, Ripmeester J, Royer J, Smith T, Tohidi B, Uchida T, Wang J, Wang W and Xiao Y. *A Hydrate Database: Vital to the Technical Community*. Data Science Journal 2007;6(Gas Hydrate Issue):GH1-GH5.
  - [23] Löwner R, Cherkashov G, Pecher I and Makogon YF. *Field Data and the Gas Hydrate Markup Language*. Data Science Journal 2007;6(Gas Hydrate Issue):GH6-GH17.
  - [24] Smith T, Ripmeester J, Sloan D and Uchida T. *Gas Hydrate Markup Language: Laboratory Data*. Data Science Journal 2007;6(Gas Hydrate Issue):GH18-GH24.
  - [25] Wang W, Moridis G, Wang R, Xiao Y and Li J. *Modeling Hydrates and the Gas Hydrate Markup Language*. Data Science Journal 2007;6(Gas Hydrate Issue):GH25-GH36.
  - [26] International Union of Pure and Applied Chemistry. 2007. *International Union of Pure and Applied Chemistry*. [online]. [Accessed 30th September]. <http://old.iupac.org/inchi/>.
  - [27] Dong Q, Yan X, Wilhoit RC, Hong X, Chirico RD, Diky VV and Frenkel MJ. *Data Quality Assurance for Thermophysical Property Databases - Applications to the TRC SOURCE Data System*. Journal of Chemical Information and Computer Science 2002;42(3):473-480.
  - [28] Hall SR, Allen FH and Brown ID. *The Crystallographic Information File (CIF) - A New Standard Archive File for Crystallography*. Acta Crystallographica Section A 1991;47(6):655-685.
  - [29] Kroenlein K, Löwner R, Wang W, Diky V, Smith T, Muzny CD, Chirico RD, Kazakov A, Sloan ED and Frenkel M. *Standardization and Software Infrastructure for Gas Hydrate Data Communications*. Proceedings of the 6<sup>th</sup> International Conference on Gas Hydrates (IGH 2008), Vancouver, British Columbia, CANADA, July 6-10, 2008.
  - [30] Google Web Toolkit - Google Code [online.] [Accessed 7th October 2009]. <http://code.google.com/webtoolkit/>.
  - [31] Ext - A foundation you can build on [online.] [Accessed 7th October 2009]. <http://www.extjs.com/>.
  - [32] GWT-Ext Widget Library [online.] [Accessed 7th October 2009]. <http://gwt-ext.com/>.
  - [33] GChart - Google Code [online.] [Accessed 7th October 2009]. <http://code.google.com/p/gchart/>.

## LIST OF ACRONYMS AND ABBREVIATIONS

ASCII	American Standard Code for Information Interchange, an American National Standards Institute and de facto international standard for digital file character mapping
CIF	Crystallographic Information File, an IUCr standard
CODATA	Committee on Data for Science and Technology, International Council for Science
DQA	Data quality assurance
GDC	Guided data capture, a process for accurate collection of data and associated metadata from literature sources
GHML	Gas Hydrate Markup Language
IUPAC	International Union of Pure and Applied Chemistry
IUCr	International Union of Crystallography, International Council for Science
NIST	National Institute of Standards and Technology, an Agency of the United States Department of Commerce
SOURCE	NIST SOURCE Data Archival System
TDE	ThermoData Engine, NIST Standard Reference Database 103
ThermoML	An XML-based approach for storage and exchange of experimental and critically evaluated thermophysical and thermochemical property data and an IUPAC standard
TRC	Thermodynamics Research Center, Thermophysical Properties Division (838) at NIST
XML	Extensible Markup Language
XSD	XML Schema Definition, a file used to specify XML file structure