Development of a Gas Hydrate Markup Language

Final Report

Prepared By:

E. Dendy Sloan, Director and Weaver Chair

Center for Hydrate Research
Colorado School of Mines
Golden, Colorado 80401
esloan@mines.edu

For:

U.S. Department of Energy

National Energy Technology Laboratory

Under Contract DE-AF26-067Nt00440

Table of Contents

Preface

This report has been compiled from four papers, the titles and authors of which appear below.

- **A Hydrate Database: Vital To the Technical Community,** by D. Sloan[1], F. Kuznetsov[2], K. Lal[3], R. Loewner[4], Y. Makogon[5], G. Moridis[6], J. Ripmeester[7], Jean-Jacques Royer[8], T. Smith[9], B. Tohidi[10], T. Uchida[11], J. Wang[12], W. Wang[13], Y. Xiao[14]
- **Gas Hydrate Markup Language: Laboratory Data,** by Tom Smith[15], John Ripmeester[16], Dendy Sloan[17], Tsutomu Uchida[18]
- **Field Data and the Gas Hydrate Markup Language,** by Ralf Löwner[19], Georgy Cherkashov[20], Ingo Pecher[21], and Y. F. Makogon[22]
- **Modeling Hydrates and the Gas Hydrate Markup Language,** by Weihua Wang[23], George Moridis[24], Runqiang Wang, Yun Xiao and Jianhui Li[25]

Please direct questions and comments to Dendy Sloan (esloan@mines.edu)

---

[1] Center for Hydrate Research, Colorado School of Mines, Golden, Colorado 80401 USA Email: esloan@mines.edu

[2] Institute of Inorganic Chemistry, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 3 Lavrentiev Pr. 630090, Russia Email: fk@che.nsk.su

[3] National Physical Laboratory, Dr. K. S. Krishnan Road, New Delhi, 110012, India Email: klal@mail.nplindia.ernet.in

[4] GeoForschungsZentrum, Telegrafenberg, Potsdam, D-14473, Germany Email: loewner@gfz-potsdam.de

[5] Department of Petroleum Engineering, Texas A&M University, College Station, Texas 77843 USA Email: makogon@tamu.edu

[6] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, ESD 90-1116 University of California, Berkeley, California, 92720, USA Email: gjmoridis@lbl.gov

[7] Steacie Institute, National Research Council, Ottawa, Ontario K1A OR6 Canada Email: John.Ripmeester@nrc.ca

[8] Ecole Nationale Supérieure de Géologie, I.N.P.L./C.R.P.G./E.N.S.G., Computer Science Department, rue du Doyen Marcel Roubault, B.P. 40, 54501 Vandoeuvre-les-Nancy, France Email: royer@ensg.inpl-nancy.fr

[9] MIT Systems, 565 Plandome Road #294 , Flushing, NY 11367-1597, USA. Email: tom.smith@cinaplex.com

[10] Institute of Petroleum Engineering, Heriot-Watt University, Edinburgh EH14 4AS, Scotland, Email: bahman.tohidi@pet.hw.ac.uk

[11] Department of Applied Physics, Hokkaido University, N13W8 Kita-ku, Sapporo 060-8628, Japan, Email: t-uchida@eng.hokudai.ac.jp

[12] Computer Network Information Center, Chinese Academy of Science, 4, 4th South Street Zhong guan cun, PO Box 349, Beijing 10080 China Emails: wrqiang@sdb.cnic.cn, whwang@sdb.cnic.cn, yunxiao@sdb.cnic.cn

[13] Computer Network Information Center, Chinese Academy of Science, 4, 4th South Street Zhong guan cun, PO Box 349, Beijing 10080 China Emails: wrqiang@sdb.cnic.cn, whwang@sdb.cnic.cn, yunxiao@sdb.cnic.cn

[14] Computer Network Information Center, Chinese Academy of Science, 4, 4th South Street Zhong guan cun, PO Box 349, Beijing 10080 China Emails: wrqiang@sdb.cnic.cn, whwang@sdb.cnic.cn, yunxiao@sdb.cnic.cn

[15] M.I.T. Systems Inc, USA, Email: tom.smith@cinaplex.com

[16] The Steacie Institute for Molecular Sciences, National Research Council of Canada, Canada, Email: John.Ripmeester@nrc.ca

[17] Center for Hydrate Research, Colorado School of Mines, USA, Email: esloan@mines.edu

[18] Hokkaido University, Japan, Email: t-uchida@eng.hokudai.ac.jp

[19] GeoForschungsZentrum Potsdam, 14473 Potsdam, Germany (loewner@gfz-potsdam.de)

[20] VNIIOkeangeologia, St. Petersburg, Russia

[21] Herriot-Watt University, Edingburgh, Scotland

[22] Texas A&M University, College Station, Texas, USA

[23] Scientific Database Center, Computer Network Information Center, CAS.
Email: whwang@sdb.cnic.cn

[24] Lawrence Berkeley National Laboratory, University of California, USA. Email: gjmoridis@lbl.gov

[25] Runqiang Wang, Yun Xiao and Jianhui Li, Scientific Database Center, Computer Network Information Center, CAS.
Email: whwang@sdb.cnic.cn

1.0 Introduction

A Beta version of a Gas Hydrate Markup Language (GHML) has been developed and published. The language structure has three parts that relate to: (a) laboratory data, (b) field data, and (c) simulation data. GHML is a method to catalogue and tag hydrate data in such a way that they can be easily found by a database user. This report explains the structure of this language, as well as the motivation and history behind its development. The detailed schema (GHML_BetaVersion1_0.xsd) is located at http://www.codata.org/ghml/. Publication of the GHML will be followed in 2007 by the development of a Portal to connect the hydrate databases which are growing in various parts of the world. Individual database developments are proceeding in parallel to this connection effort, with encouragement to use GHML as a common language for communication among these databases.

1.1 *Objectives of the Database and GHML Development Effort*

The objective of this work is to provide free access to non-proprietary international gas hydrate databases, so that parties interested in gas hydrate research work can quickly and easily access multiple data sets around the world and query them for information on gas hydrates. The objective is being accomplished through: (1) the development and publication of GHML during 2006 and (2) the establishment of a Gas Hydrate Portal during 2007.

A conceptual illustration of the databases and connections is given in Figure 1.

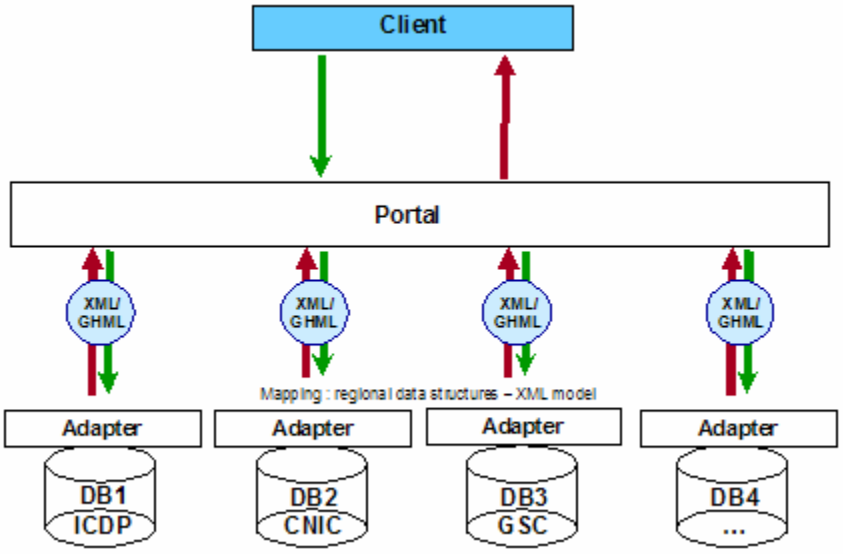

Figure 1. A conceptual diagram illustrating the application of Gas Hydrate Markup Language.

The concept rests on the establishment of multiple databases (DB1 through DB4 at the bottom of the diagram). Without the foundation of these underlying databases, there would be no need for a language to communicate with them. The database examples shown here are labelled ICDP (for InterContinental Drilling Program), CNIC (for Computer Network Information Center), GSC for

the Geological Survey of Canada, etc., after some of the well known databases that contain hydrate information.

When a client requests some information about hydrates the client query is sent to a Portal, which communicates with the various databases through a common language – the Gas Hydrate Markup Language. The client's query and response is communicated with each participating database, so that a number of data responses are obtained.

The option of utilizing an existing Markup Language was considered, but the field of hydrates is so general that while segments of several markup languages could be utilized for portions of the data, no single entire language would suffice. For example, when the IUPAC standard Markup Language for thermodynamic data (ThermoML, as described by Frenkel, et al., 2006) was considered, there was no capability for inserting geological and geophysical data such as well logs, a capability that would seem to be vital to hydrates database communications. Similarly, other Markup Languages were missing components for thermodynamic and kinetic data. In the end, a new language was justified. However, the study of the existing Markup Languages provided valuable guidelines for the construction of GHML.

This work is being funded by the Department of Energy (DOE) and by CODATA, the Committee on Data for Science and Technology, an interdisciplinary Scientific Committee of the International Council for Science (ICSU). The work will also be authenticated under the CODATA aegis. CODATA was established in 1966 by ICSU to promote and encourage, on a world-wide basis, the compilation, evaluation and dissemination of reliable numerical data of importance to science.

1.2 *Motivation for a Hydrate Database and GHML*

Recent analyses by Bernstein (2004) and Sachs (2005) suggest that technology is a major driver of societal economies. In turn, energy is a foundation of technology, and efficient access to data enables effective use of energy. The thesis of this work is that access to and analysis of the growing volume of data related to methane hydrates will enable development of an important new energy source, and that the development of GHML is the first step in connecting multiple international databases into a single source of information for researchers.

Natural gas clathrate hydrates are ice-like compounds which form when water encages small (<0.9nm) guest hydrocarbon molecules in crystal cages, as reviewed by Makogon (1997) and by Sloan (1998). Recent estimates of the methane content of hydrates in nature range from 2.5×10^{15} m^3 STP (Milkov, 2004) to 120×10^{15} m^3 at standard temperature and pressure (STP) (Klauda and Sandler, 2005). These estimates, while wide-ranging, are extremely large relative to the estimated conventional gas reserve of 0.15×10^{15} m^3 of methane (STP) (Radler, 2000).

Hydrate knowledge is expanding exponentially, as shown in the semi-logarithmic plot of Figure 2 (Sloan, 2004). This plot gives the number of refereed hydrate publications in the last century, by decade; there were two publications from 1900–1910, and 3010 publications from 1990–2000. A semi-logarithmic extrapolation into the first decade of the 21st century leads to a staggering 7,500 publications — slightly more than two refereed publications per day, seven days per week.

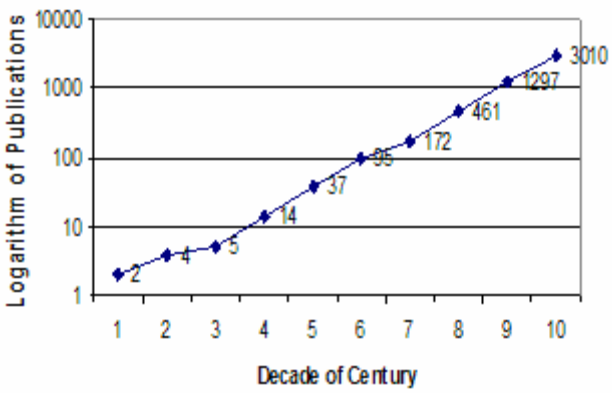

Figure 2. Number of refereed hydrate publications in the 20th century, by decade (Sloan, 2004).

Historically it has been sufficient to merely store such data in isolated and disparate databases for later retrieval and reporting. Data was then provided to third party requesting entities in a variety of ad-hoc formats leaving the requesting researchers with the unfortunate and daunting task of having to process these various non-standardized extracts, rather than being able to focus on the actual research at hand.

With the expanding knowledge base, it is apparent that we need an efficient means of managing hydrate data. CODATA provides a convenient, internationally-sanctioned means of dealing with such information.

The GHML described here is an Extensible Markup Language (XML) based implementation and standard which is designed to readily allow the modeling and subsequent exchange of data pertaining to the more common Gas Hydrate constructs encountered in the research environment. By careful coordination with the Gas Hydrate research community, a workable and viable GHML schema has effectively been realized.

Concurrent with the architecture of the GHML, existing related markup languages and standards were investigated and researched to help ensure compliance with industry accepted standards and practices wherever applicable. Careful concern was given to such issues as the integration of preexisting markup language constructs, enumeration, naming conventions, attributes and abstraction.

1.3 *History of the GHML and Associated Databases*

The development of the GHML is in the final phase of a four phase effort begun in 2000. A brief history of this effort is provided below.

Phase I (2001 – 2002). In 2000 CODATA authorized the first hydrate database Task Group, chaired by F. Kuznetsov. During this phase the Task Group met twice to organize the project and publicized the first local database, GASHYDAT (2001) developed by Drs. Klerkx (Belgium) and Dimitrov (Bulgaria), with European Union funding. While a landmark in hydrate databases, the database was initially small and limited by the funding level, and work stopped in 2001. The GASHYDAT Web site (www.gashydat.org) domain expired recently.

Phase II (2003 – 2004). In 2002 a second CODATA Hydrate Task Group was authorized, also chaired by F. Kuznetsov. During the ensuing two years the Task Group held six world regional meetings (Table 1) to educate the hydrate community regarding the need for a database and to gain worldwide acceptance of the idea of an internationally-distributed database. The project was supported by several National and International bodies such as ICSU.[26] Also during the second phase, the Task Group distributed a hydrate bibliography database including, authors, titles, page numbers, dates, and article abstracts for approximately 5,000 articles.

Table 1. Hydrate database meetings held during 2002-2004.

Region	Russia	North America	Japan	India	China	Europe
Meeting Place	Novosibirsk	Salt Lake City	Chiba	New Delhi	Beijing	Potsdam
Meeting Date	27/01/03	13/05/03	11/12/03	20/02/04	15/03/04	5/11/04

Phase III (2005 – 2006). In 2004, CODATA gave the Hydrate Task Group its third charter. It was during the third phase of the effort that a need was recognized for a "common language" for communication among hydrate databases in the US, Canada, Germany, China, Bulgaria, and elsewhere (e.g., the Mallik 5L database, which was established simultaneously by the Canadian Geological Survey and GeoForschungsZentrum (GFZ) in Germany, and the extensive Chinese gas hydrate database found at www.gashydrate.csdb.cn).

Information technology experts (e.g., Jason Wang, and Ralf Löwner) suggested that the data surrounding hydrates were sufficiently unique to warrant a unique markup language. Löwner's perspective was influenced by his having just completed the database for the Mallik 2002 hydrate well described by Dallimore and Collett (2005).

The first meeting of a revised hydrate database committee (Table 2) composed of both hydrate and information technology experts, was held on June 11-12, 2005 in Trondheim, Norway, just prior to the 5th International Conference on Gas Hydrates. A second meeting was held just following the International Hydrate Conference in Kauaii, Hawaii March 16-18, 2006. At this meeting the committee constructed the markup language structure of three basic elements: (1) laboratory data, (2) field data, and (3) data in support of modeling.

[26] From 2002 to 2004, ICSU supported the CODATA gas hydrate initiative by providing a US$ 100,000 grant over two years. It helped in organizing regional meetings in Russia, US, France, China and India (see Table 1).

Table 2. Revised hydrate database committee.

Country	Member	E-mail
Canada	John Ripmeester	John.Ripmeester@nrc.ca
China	Xiao Yun Wen Hao Zhang Hui	yunxiao@sdb.cnic.cn hwen@home.ipe.ac.cn huizhang@codata.cn
France	Jean-Jacques Royer	Jean-Jacques.Royer@ensg.inpl-nancy.fr
Germany	R. Löwner (corresponding member)	loewner@gfz-potsdam.de
India	K. Lal	klal@csnpl.ren.nic.in
Japan	T. Uchida	t-uchida@eng.hokudai.ac.jp
Russia	F. Kuznetsov	fk@che.nsk.su
UK	B. Tohidi (corresponding member)	bahman.tohidi@pet.hw.ac.uk
USA	D. Sloan	esloan@mines.edu

The initial outline of the GHML was generated at the Computer Network Information Center at the Chinese Academy of Science (CNIC), by Willa Wang and Jason Wang in 2005. That initial effort was revised and set as database schemas in 2005 and 2006 with the entire Task Group, with intensified effort by a small steering committee composed of Ralf Löwner, George Moridis, Tom Smith, Willa Wang, Jason Wang, and Dendy Sloan. The Beta-version of the GHML was published at the Beijing CODATA meeting on October 23 with three components: (1) a laboratory data schema by Smith *et al.,* (2) a field data schema, by Löwner *et al*., and a modelling/simulation schema by Wang and Moridis, *et al.* The Beta version of the GHML was also published in a special volume of the CODATA electronic journal.

The developers of GHML considered international standards, particularly the standards defined by the W3C (World Wide Web Consortium) and the OGC (Open Geospatial Consortium). Various related standards were analyzed and compared to requirements, primarily, the Geographic Markup Language (ISO19136, GML) and the whole ISO19000 series. However, the requirements demanded an XML application schema readable for any scientist without background in information technology. Therefore, ideas, concepts and definitions were used to build the GHML without importing any one markup language. This approach enabled a comprehensive schema with a simple maintenance.

Extensive documentation consisting of a detailed explication integrated in the application schema, an HTML-based document and detailed documentation, ensures the usability of GHML. Because of the close collaboration of gas hydrate experts and specialists in geoinformatics, the application schema of GHML is very user-oriented. This usability is considered an important factor in assessing the value of the GHML to the scientific community.

The Beta version of the GHML, including a glossary and documentation, is available on the CODATA website for user comment during a six month period beginning February 1, 2007. Revisions will be made to address the comments received. The authors anticipate an electronic, ever-green publication of GHML, in which each revision will update, but not outdate, the earlier versions.

Phase IV (2007). During the final phase of the project a portal will be generated that will enable the connection of all the extant hydrate databases. This work is expected to be completed at the end of the 2007 calendar year. At the same time, national databases are continuing to be

developed in parallel to this effort. The development of the GHML and portal by CODATA should act to both enable and encourage the further development of these databases.

1.4 *How Will the Portal and GHML Work?*

The Figure 3 illustrates one method in which the GHML can be applied. To begin, a user (utilizing a web browser) could connect to the 'portal' and request whatever information they so desire. The system would then create an XML document (encapsulated via the Simple Object Access Protocol or SOAP, a protocol for exchanging XML-based messages over computer networks, normally using HTTP), containing the request for said data. This document would then be communicated to the adapter of each of the participating organizations. Each of the adapters would then parse the XML request, search its associated database, place the response back into an XML document (conforming to the GHML) and then send that response on to the portal for the user's viewing.

The central role of the portal is to communicate the user requests to all of the participating organizations, collect the responses and then return those responses to the user. In this way the user is presented with a single standardized and central "location" from which to carry out their research rather than multiple locations with varying and questionable interfaces.

In addition to direct user requests via a web browser, it is envisaged that the portal will contain one or more web services capable of automatically communicating gas hydrate data amongst researchers within the community. By adhering to certain standards, organizations can create web services of their own which would allow automated data sharing and exchange without the necessity of direct user intervention.

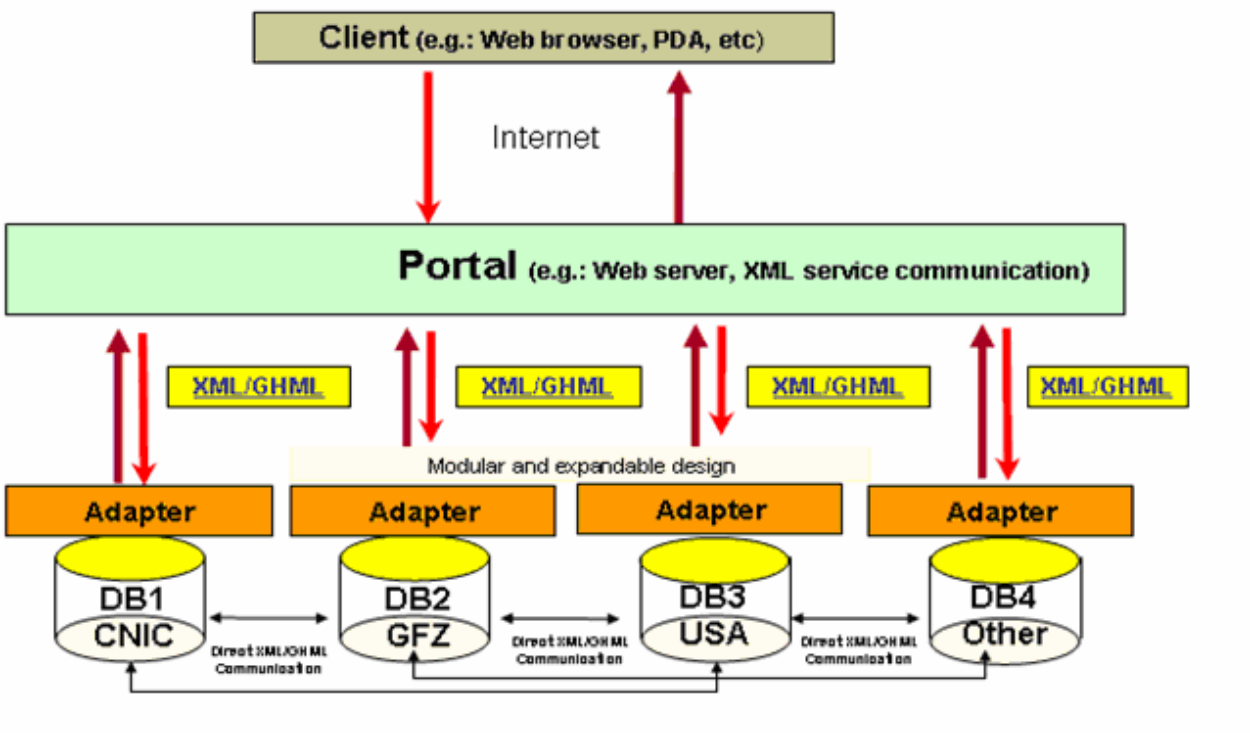

Figure 3. Illustration of how the portal will work.

Furthermore one can also envisage communication occurring directly between the various participating organizations (via XML), bypassing the portal entirely. This ability is inherent in the architectural design of the system and can be implemented on an "as needed" and/or "as desired" basis. Reasons for doing this might include data replication between databases or consistency/validation checks with other datasets.

An XML schema is the definition of the data model. Based on this data model, XML files will be created that are instance documents of the XML schema, the concrete data objects. So, these XML files use the XML schema as building instructions (Figure 4) and contain the scientific data itself.

On the one hand, the XML schema (GHML) is used for the validation of the XML instance documents. On the other hand, these concrete objects could also be created based on the XML schema itself. In the present case, the XML file will contain both data and Meta data of any gas hydrate investigation.

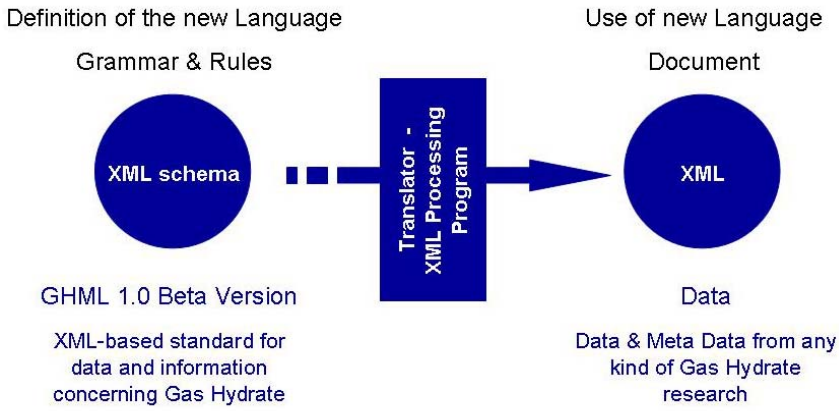

Figure 4. Diagram illustrating the relation between the XML schema definition and the XML instance document.

The resulting XML files are standardized in a unified format that enables data exchange, data storage, data visualization and data mining. The scientist is generally not obliged to handle the XML file itself. Tools and interfaces, which permit all desired data transactions, perform this instead. These functionalities will be provided by services accessible via the Internet based portal.

The distributed heterogeneous international databases are connected by specific adapters to the data infrastructure. These adapters create the XML files and translate the standardized files in proprietary formats used by the local data provider and vice versa. Therefore, the data providers themselves don't need to make any changes to their internal data structure or model.

An early implementation of such an adapter was realized at the GeoForschungsZentrum Potsdam (www-app1.gfz-potsdam.de/ghml/). It is a demonstration of a concrete use case. Instead of a database, a data file that originated from a sensor in a borehole registering mass values of different gas types at different depths could be connected to the portal. An XML file based on the central new virtual data model would be created. With the help of an internet based interface (Figure 5), the user is led through a three step process, permitting the adapter to create the XML file and to store the data from the data file into the exchange format. The resulting data charged XML file could be saved to the hard disk of the user with the XML extension (.xml). This early implementation has not proceeded beyond this point. Neither portal function nor service has yet been realized.

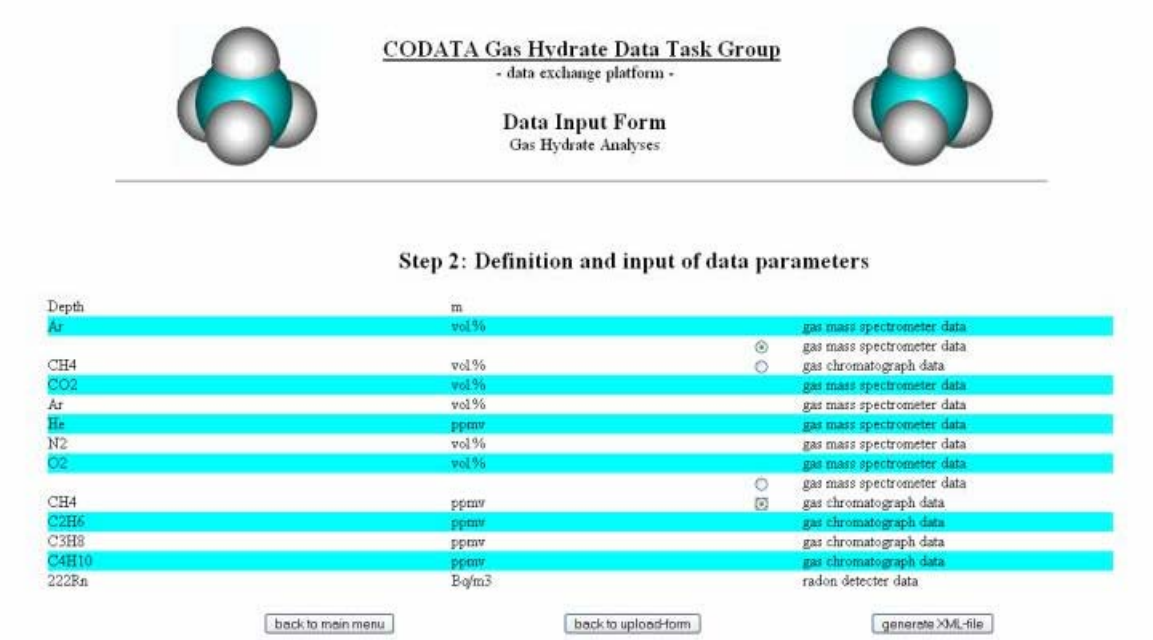

Figure 5. Internet based user interface for the creation of XML files based on the field data portion of GHM: Step two of the three step process.

2.0 Laboratory Data

Laboratory Hydrate Data is one of the three constituent modules comprising the XML based Gas Hydrate Markup Language (GHML) schema (the others being Field Hydrate data and Hydrate Modeling). This module describes the characteristics of natural and synthetic gas hydrates as they pertain to data acquired via analysis within a laboratory environment. Such data includes preservation history (i.e., technique, pressurization gas and pressure, etc.) and macroscopic data (i.e., water-sediment ratio, appearance, P-T behavior, etc.), as well as microscopic data.

The Laboratory Hydrate Data Module of the GHML is currently in Beta revision and as such is a work in progress. By careful integration with the other two portions of the GHML (i.e., Field and Modelling), a fairly complete and robust markup language capable of communicating and sharing gas hydrate data across networks and the Internet should result.

The laboratory data module specifically focuses on data gathered within the laboratory setting as well as the preservation history and basic origin information of the sample. This data includes certain relevant metadata for the sample, the source and conditions involving for example the extraction of a natural hydrate sample, its preservation history, and macroscopic qualities. A section for microscopic and mesoscopic details has been incorporated for anticipated future expansion.

The primary purpose of the schema is to provide a standardized method in which to communicate gas hydrate data amongst potentially disparate and unrelated organizations across the internet. Since it is meant to communicate gas hydrate data, the scope of what can and can not be exchanged has necessarily been limited and thus the schema itself provides a form of validation check on the XML documents that are to be exchanged.

2.1 *Structure of Laboratory Data Schema*

The laboratory data portion of the GHML is constructed in a clear modular format consisting of five (5) primary blocks, which makes it readily easy to understand and implement, as well as to add on additional schema components as needed. Research was done to ensure that the schema was carefully modeled to reflect the way in which data is currently gathered and recorded in the laboratory environment, as opposed to a somewhat more idealized or abstract method. This approach was chosen in order to help facilitate the integration of the GHML with existing databases and data efforts across the research community. Since it is modeled in this fashion, the GHML may necessarily diverge from certain industry standards and/or recommendations in order to meet current data sharing needs. The following figure illustrates the top level tags for the laboratory data GHML.

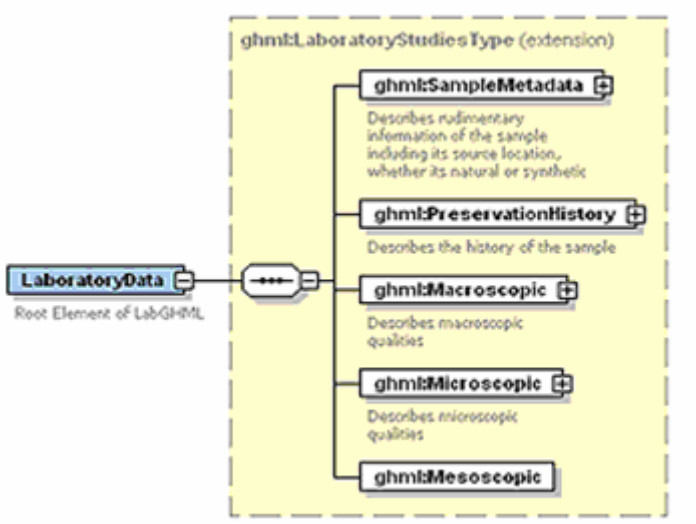

Figure 6. Basic elements of the Laboratory Hydrate Data Module.

In the Laboratory Hydrate Data GHML, the schema elements/types are directly analogous to fundamental scientific constructs within the Gas Hydrate community. This method (as opposed to the equally valid choice of implementing a layer of abstraction), was chosen in order to help reinforce the understanding and meaning of the data being exchanged thereby leaving little to no room for ambiguity. This is considered vital for the success of the GHML, especially in an international venue.

2.1.1 Namespaces

In designing the GHML there was much discussion on the topic of 'namespaces' and in particular which methodology was best and should be adopted. Essentially there are three possibilities when one considers this:

a) Do not use a default namespace

b) Set the default namespace to XMLSchema

c) Set the default namespace to the target namespace (GHML chooses this)

Each of the above have their associated advantages and disadvantages and thus the decision needed to be based not on what was the 'best' overall practice (because arguably there really is not one), but rather on what was most advantageous for the future of GHML. Since incorporation of other markup langauges may indeed occur at some point in the future, namespace qualifying the various schema components was deemed a requirement and thus the consensus was for option 'c'. The target namespace is: **ghml**

2.1.2 Naming Conventions

Careful consideration and thought went into the decision regarding the naming of the elements and data types used throughout. As a result, when one views the schema, it is relatively obvious

as to what data are actually being 'marked' up. Not only has considerable thought been given to the descriptive names of the various 'tags', but equal concern was given to the 'case' of said tags. With that in mind, the GHML adopts a combination of Pascal case, Camel case and Uppercase for the various tags used throughout as follows:

- Pascal Case: First letter of each concatenated word being capitalized – e.g.: PascalCase
- Camel case : First letter is lowercase and first letter of each subsequent concatenated word is capitalized – e.g.: camelCase
- Upper case : Every letter of the tag is capitalized – e.g.: UPPERCASE

The bullets below outline the standard to which GHML adheres to, however depending on the meaning, context and/or source (i.e., another applicable ML) of the tag, GHML may diverge from the standard.

- Complex Types: PascalCase + the word 'Type'
- Simple Types: camelCase + the word 'Type'
- Elements: PascalCase
- Attributes: camelCase, PascalCase, Uppercase

2.1.3 Enumeration, Units and Attributes

Where feasible, the realm of possible data values has been restricted to either an enumerated list or some range of permitted values. Enumerated lists follow their proper case and thus, for example, the chemical representation of methane would be represented as CH_4.

Another example of this would be the enumerated element entitled *Appearance*, which is limited to having the following possible values:

- Massive
- Nodular
- Pore Hydrate

Attributes are the exception rather than the norm in the GHML and thus they have been implemented sparingly and only where it was decided that their use constituted a clear advantage. Otherwise, elements have been opted for in their stead. An example of a clear advantage is in the specification of units and as such, where applicable, the attribute 'uom' has been incorporated to clearly denote the units of the respective data in question.

In general, a standard of SI-type units or some multiple thereof has been chosen throughout the GHML, although some exceptions to this choice do exist to ease and facilitate the integration of data that may exist in other units. Concern was specifically given to circumstances where conversion of such data might prove infeasible or difficult given available resources.

One such example is that of *pressure*. Pressure typically should be recorded in the unit of *Pascal* but data often exists in pounds per square inch and thus *psi* has been incorporated to address this current reality in the research sector.

2.2 *Synopsis of Laboratory Data GHML*

Consisting of five (5) primary blocks, the Laboratory data GHML is outlined as follows:

- *SampleMetadata* – Describes details of the source of the sample.
- *PreservationHistory* – Describes the history of how the sample has been preserved including its Pressure/Temperature behaviour as a function of time.
- *Macroscopic* – Describes various macroscopic qualities of the sample such as its appearance, colour, water/sediment ratio and so on.
- *Microscopic* – Though not fully implemented in this Beta version, the anticipated purpose of this block will be to describe any desired microscopic qualities.
- *Mesoscopic* – That which does not clearly fall into the realm of Macroscopic or Microscopic, will be placed within this block. Currently this is a stub block which does not as of yet have any elements/types within the current Beta version.

In the following subsections, each of the five (5) heretofore mentioned blocks will be overviewed in summary. For a detailed explanation please see the GHML documentation.

2.2.1 Sample Metadata Block

This block describes metadata information related to the origin of the sample. The following list outlines the associated top level tags for this block:

Source	Indicates whether the sample was sourced from Onshore, Offshore or if it is Synthetic.
SampleID	Indicates the unique identifier for the particular sample at hand. The identifier could be an alphanumeric string.
OriginDate	Indicates the date in which the sample was either extracted or created in the laboratory environment.
OriginLocation	Indicates the location from which the sample was sourced. If it is a natural sample the location may include latitude, longitude, water depth, permafrost depth, and meters below sea floor. If synthetic, it includes the research center name and description.
OriginConditions	Indicates the conditions that existed at the source including the in-situ pressure and temperature.
RecoveryMethod	Indicates the method of recovery such as: piston drop core, autoclave, pressure core, pressure temperature core, ROV.
InvestigationData	Information related to the source dataset which includes the analysis date, dataset name, data file, owner, contact information, responsible parties, researcher, comments.

2.2.2 Preservation History Block

This block contains information related to the history of the preservation of the sample. The following outlines the top level tags for this block:

Technique	Indicates the technique utilized for the preservation of the sample such as internal pressurization, external pressurization, liquid nitrogen.
PressurizationGas	Indicates the gas used for pressurization such as He, N2, CH4.
ContainerConditions	Indicates the conditions of the container such as the pressure and temperature.
PTBehaviour	Indicates the pressure and temperature behaviour over time. Data considered include the phases (e.g., LHc, V, H1, H2, Hh, I), pressure, temperature, component and mole fraction.

2.2.3 Macroscopic Block

This block contains information related to the macroscopic qualities of the sample. The following outlines the top level tags for this block:

Appearance	A physical description of the sample.
Color	A description of the color of the sample described by the researcher.
WaterSedimentRatio	Indicates the ratio of water mass to sediment mass.
GasWaterRatioSTP	Indicates the ratio of gas volume to water volume.
PTBehaviour	Indicates the pressure and temperature behaviour over time. Data considered include the phases (e.g.: LHc, V, H1, H2, Hh, I), pressure, temperature, component and mole fraction.
Gas	Description of the gas component(s) of the sample as well as isotope analysis data. The gas component portion contains both the gas (such as Ethane, CH4 etc) along with its related mole percent. The isotope analysis portion contains the gas component, the isotope as well as delta value information.
Water	Contains information regarding ion concentration as well as isotope analysis. Ion concentration contains the ion, mass percent, mole percent and parts per million. The isotope analysis portion contains the gas component, the isotope as well as delta value information.

2.2.4 Microscopic Block

This block contains information related to the microscopic qualities of the sample. Though currently under development, the following outlines the top level tags for this block thus far:

Morphology	Currently under R&D
XRayDiffraction	Currently under R&D

2.2.5 Mesoscopic Block

This entire block, which is currently under consideration and development, is intended to contain information related to that which does not clearly fall within the realm of the microscopic or macroscopic categories.

2.3 *Uncertainty Data*

It was realized during the development of the GHML that there needed to be provisions to exchange uncertainty data for the various measured values used throughout. As such, each measured value has the provision to carry with it any related uncertainty data.

By carefully review of NIST technical note 1297 entitled “Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results,” a comprehensive schema for the uncertainty data was architected. This document was chosen as a basis for two reasons: a) It is based on another document entitled “Guide to the Expression of Uncertainty in Measurement” (GUM) by the International Organization for Standardization (ISO), and b) By adopting these standards, possible integration with ThermoML might be more readily facilitated since ThermoML also adopts similar standards.

The uncertainty information that is carried along with each measured value includes the following:

- Evaluator
- Evaluation Method
- Standard Uncertainty (Uc)
- Expanded Uncertainty (U)
- Coverage Factor (k)
- Confidence Level

For a detailed explanation of the above items please refer to NIST document TN 1297.

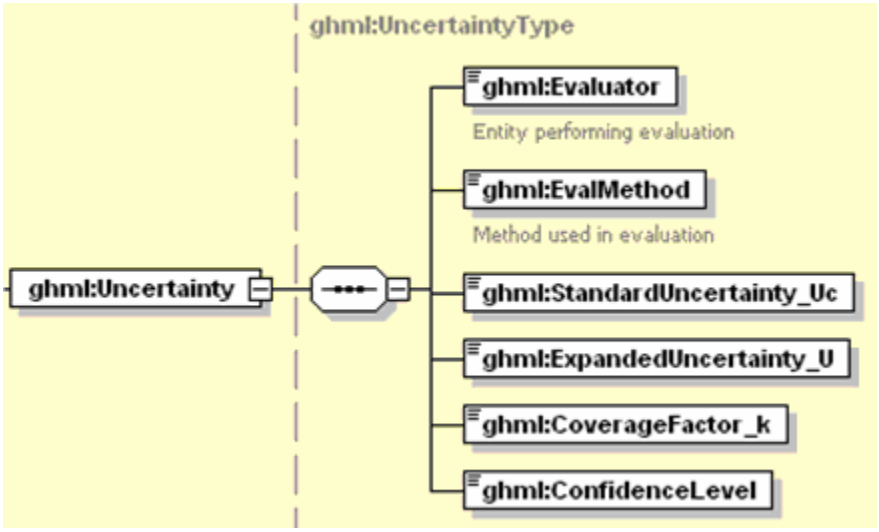

Figure 7. Basic elements of the Uncertainty Data Schema.

3.0 Field Data

Within the GHML standard, the "Field Data" part is a module which is used for all data and information coming from the field. Therefore, the application schema includes all aspects of natural gas hydrate investigations in contrast to analysis of samples in laboratory or research on synthetic gas hydrates. The "Field Data" part is differentiated between spatial related investigations and borehole investigations.

The construction of the field data model as a portion of GHML was carried out within the context of deadline constraints and the requirement for acceptance in the scientific community. However, the result was a "quick win solution," which is self-contained and yet capable of integration into GHML. Due to the self-explanatory way of the schema model, the field data part is clearly comprehensible for all researchers without any background in Information Technology.

The relatively simple schema model dispenses with substitution groups, abstract types, imports or includes. No other namespaces than default (xsd:) and target namespace (ghml:) are defined. Most of the elements are optional, which makes the schema very flexible. The names of the elements reflect their denotation and they are determined together with the scientific community.

The naming convention corresponds to the other parts of GHML. The definition of complex types use UpperCamelCase notations followed by the word "Type" (e.g., FieldDataType), simple types definition uses lowerCamelCase notations followed by the word "Type" (e.g., doubleOrNullType), and elements are described by an UpperCamelCase notation (e.g., FieldData).

Nevertheless, this schema uses attributes which are mostly locally defined. In addition to unique identification numbers, specific information about elements (e.g., wellHoleName) could be assigned to each data set.

3.1 *Structure of Field Data Module of GHML*

The structure of GHML is strictly hierarchical and the field data portion is set as an element (FieldData) on the second level under the main GHML element (Figure 8), beside the other two Gas Hydrate data modules corresponding to research areas.

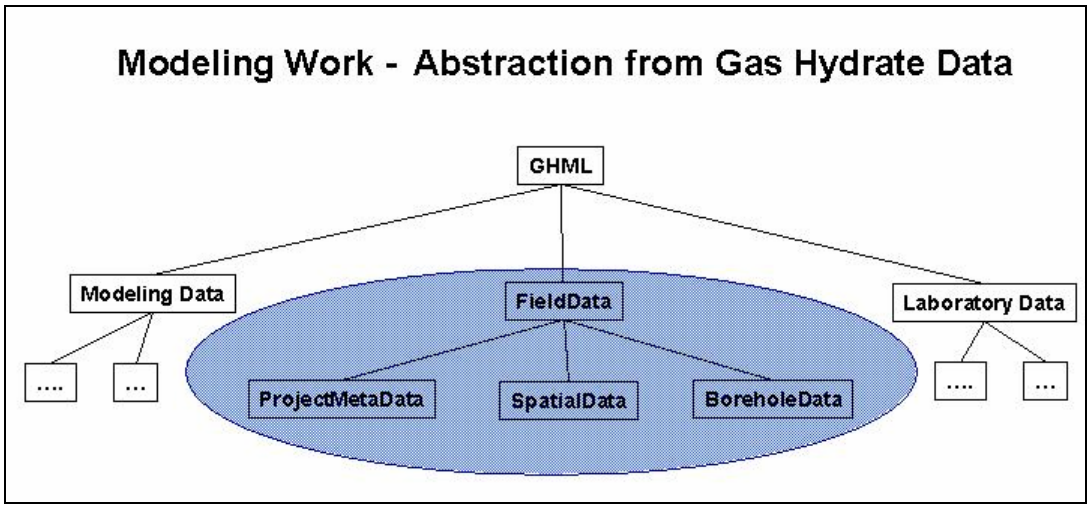

Figure 8. Field data portion within the GHML.

The "FieldData" element is optional and could be unboundedly repeated within an XML instance document. An identification number could be assigned to each instance of this element. It consists of three other elements on a third level (Figure 9).

Firstly, the "SpatialData" element describes all spatial oriented data and Meta data content coming from the field. The data is first of all related to a surface (e.g., seismic investigations or outcrop analyses). Secondly, the "BoreholeData" element describes all data and Meta data content coming from borehole investigations. Here, data is mainly vertically related (e.g. to depth or time). Lastly, the "ProjectMetaData" element describes all principal additional Meta data of the project from which the field data results. This could be information about the investigators or the objectives of the project. The two first elements define both data and Meta data, whereas the "ProjectMetaData" element only describes the Meta data. All elements are available with optional identification numbers as attributes which enables a better adaptation to databases.

Although GHML does not import any other international standards, the field data portion integrates some structures and ideas from other already defined standards to secure exchangeability and extensibility. Various related standards were analyzed and compared to the requirements. Primarily, basic elements of the Geography Markup Language (GML) and other parts of the ISO19000 series are used for spatial related features. Also, other standards were analyzed (e.g., parts of WellLogML) for elements related to drilling activities.

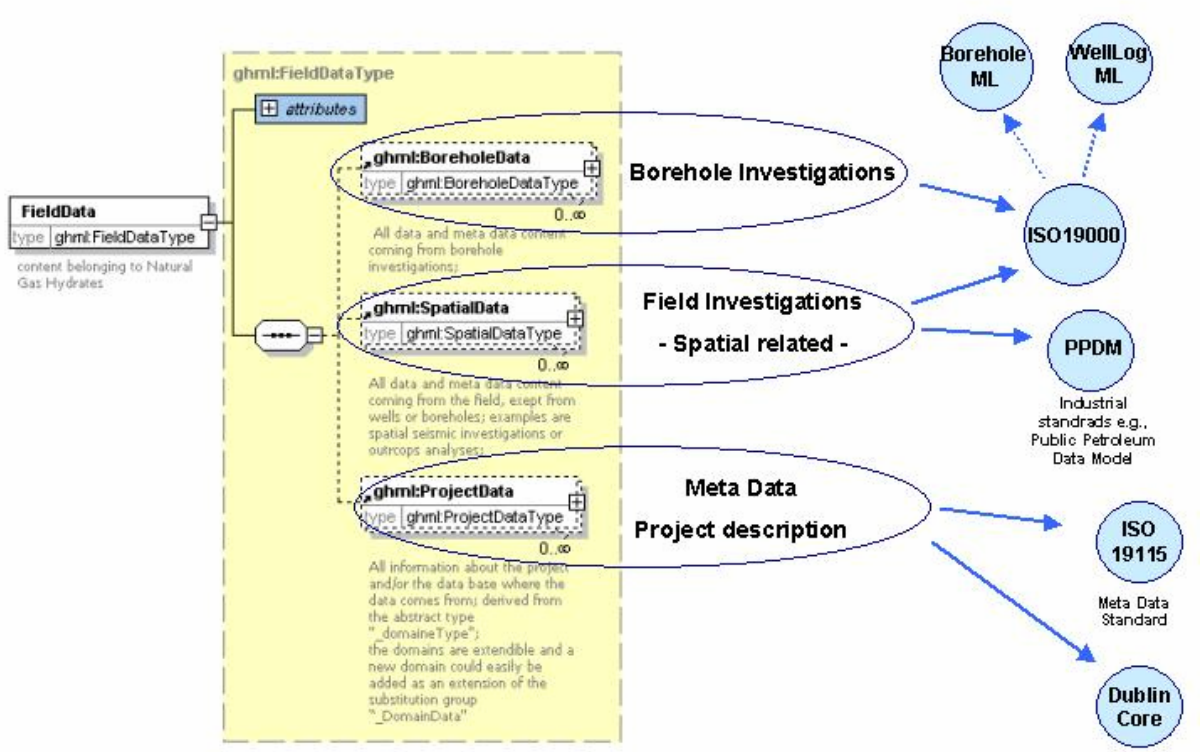

Figure 9. Field data portion with the three different elements on the third level, showing also the relations to already existing standards.

According to most accepted international standards, the use of global elements and types permits the exchangeability and extensibility of the field data portion. The structure, in all of the three elements, will always be defined by a relatively small number of global types. The result is an easy understandable data model, which doesn't require knowledge of each element or parameter. Because of the similarity of each element, the following notes describe only the "BoreholeData" element as an example.

The optional "BoreholeData" element consists of a number of attributes and other elements (Figure 10). The attributes describe direct information about the specific borehole, such as position and identification number. Only the name of the well hole is set as mandatory. Besides the attributes, one can find a list of other optional unbounded elements represented as a research field or domain. Geological, Geophysical, Geochemical and Biological investigations are provided, but this list could be enlarged in the future. These elements are listed under the fourth level of XML hierarchy.

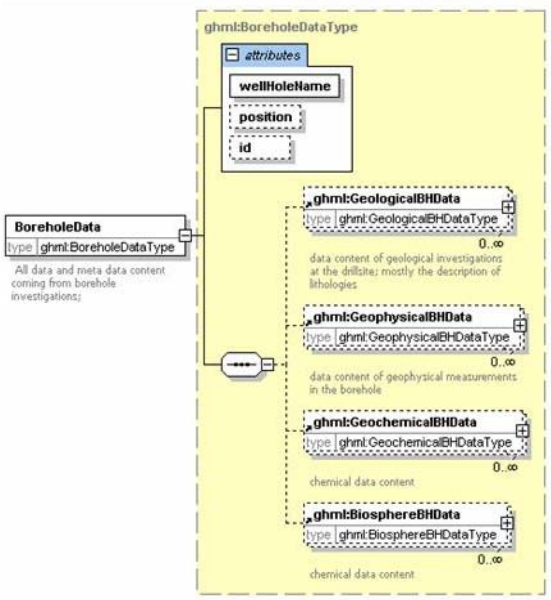

Figure 10. Borehole data element and its different underlying optional elements representing the different Gas Hydrate research investigation in a drilling well.

The different investigation fields or domains contain activities under a fifth level (Figure 11), for example, the Geochemical Investigation contains analyses of gas data (GasData) and analyses of the water column (WaterColumnData). These lists could be enlarged in the future and adapted to other, newer research fields and activities.

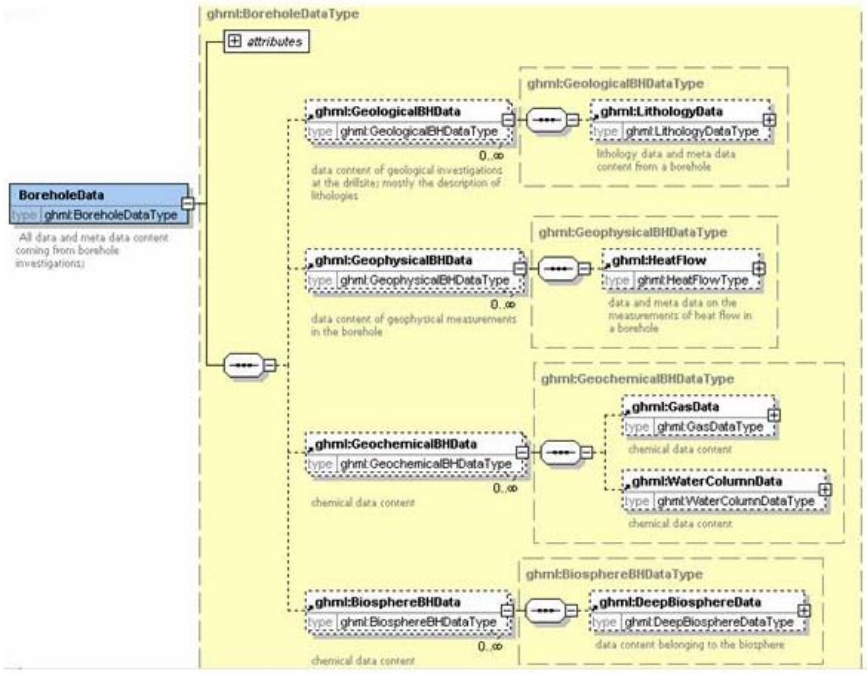

Figure 11. Borehole data element and its different Gas Hydrate research investigation elements and the underlying research activities.

All the activities are built up the same way. They can be added to or deleted from according to requirements. This results in a flexible model, which can be adapted to each new scientific research field. New types and elements can be created by the extension of two types: the "InvestigationDataType" and the "MeasurementDataType" (Figure 12). The first describes the Meta data of the data set, while the second contains the data set itself. Only a few elements are included in each type, so the extension will be adapted with each data set.

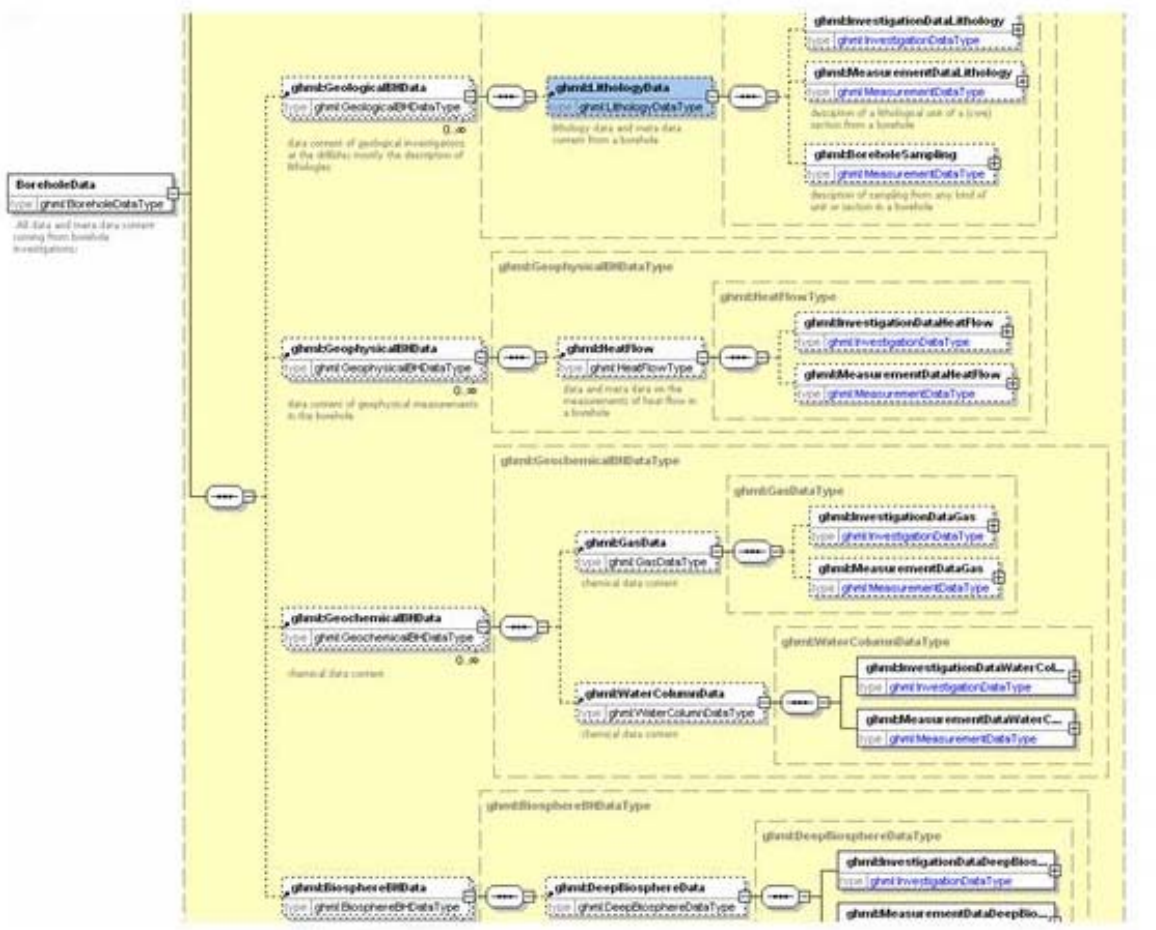

Figure 12. Illustration of the entire expansion of the "BoreholeData" element with the "InvestigationDataType" and the "MeasurementDataType" of each activity.

Regarding the Meta data description element (InvestigationData), an optional attribute for the identification number and several elements of simple types are included, all optional (Figure 13). These elements provide the information of the data set as well as any copyright item. By extension, the specific fields of the different activities are included (Figure 14).

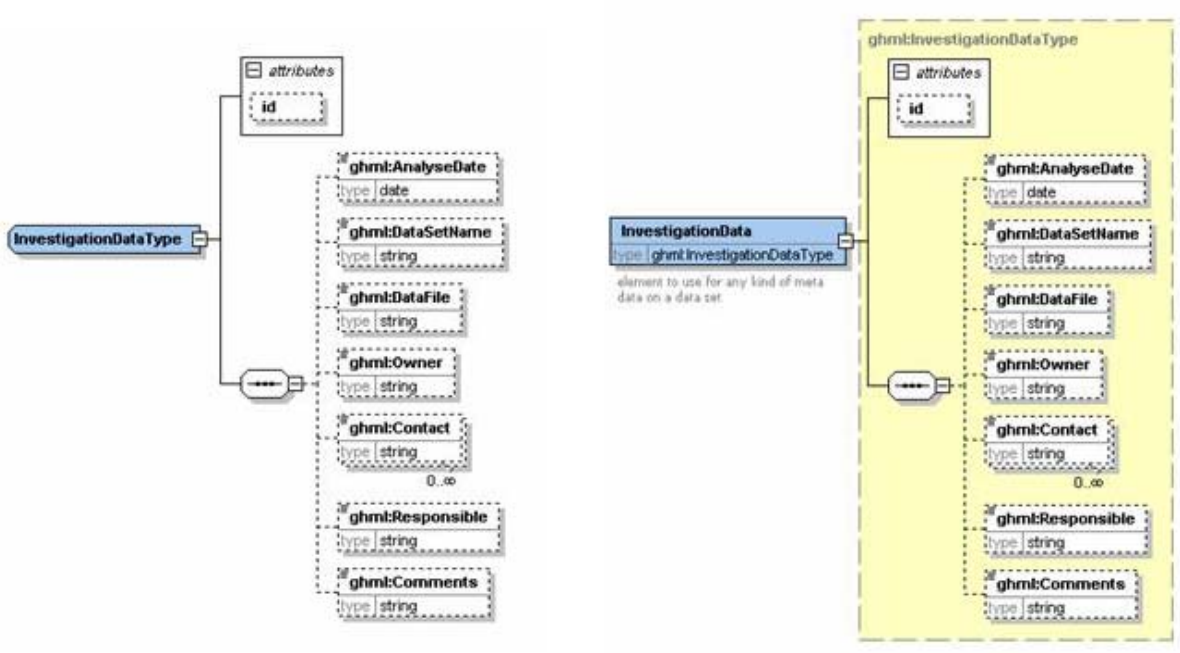

Figure 13. The global "InvestigationDataType" (left) and element (right).

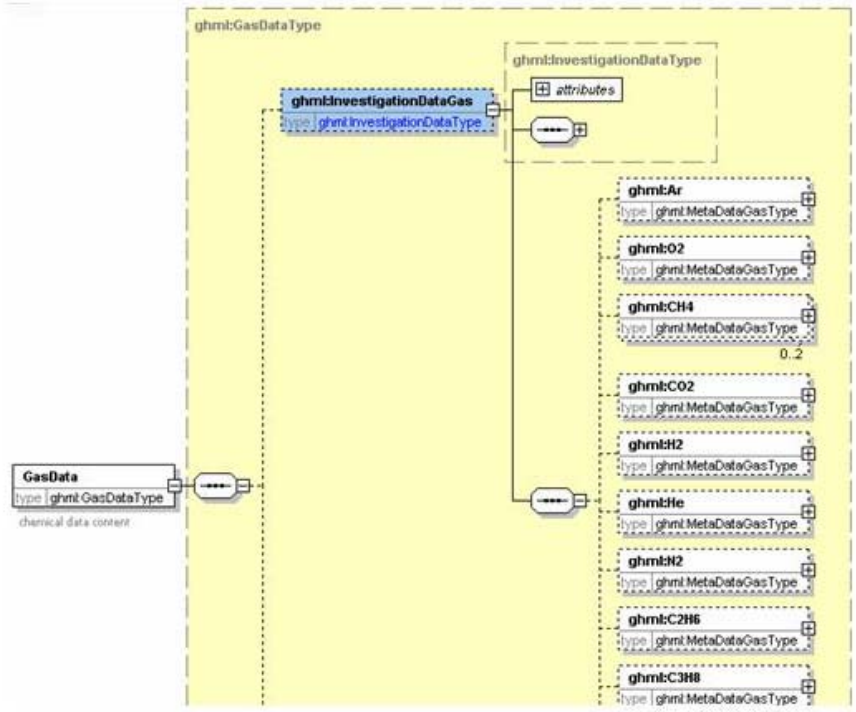

Figure 14. An example of an extension of the "InvestigationDataType": analysis of gas coming from the borehole while drilling.

The "real data" could be described and stored in the XML instance document, which is built according to the "MeasurementDataType" (Figure 15). This type consists of two optional

attributes: an identification number and a digital object identifier (DOI). A persistent DOI could be assigned to each data set. The DOI attribute enables data publication and offers authors an incentive to publish data through long-term repositories.

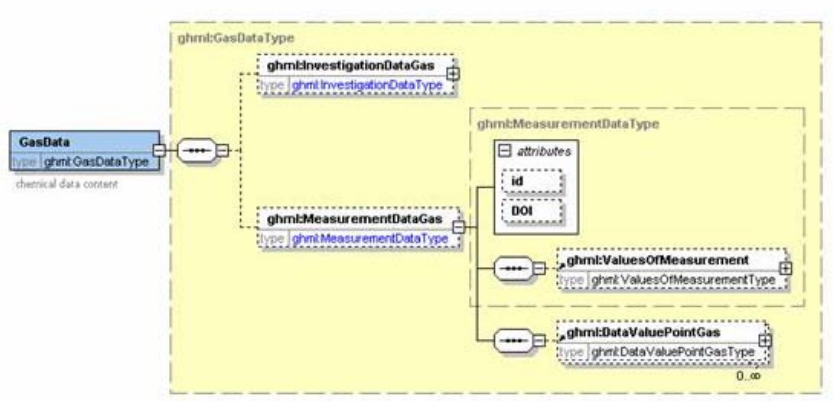

Figure 15. The "MeasuremntDataType" extended for analyze of gas in a borehole.

Additionally, this type includes an optional element, which could describe data in all existing formats: the "ValuesOfMasurement" element, accorded to GML. It includes solutions for the storage of data tables, lists and any kind of data files (Figure 16). A detailed description is attached in the field data model.

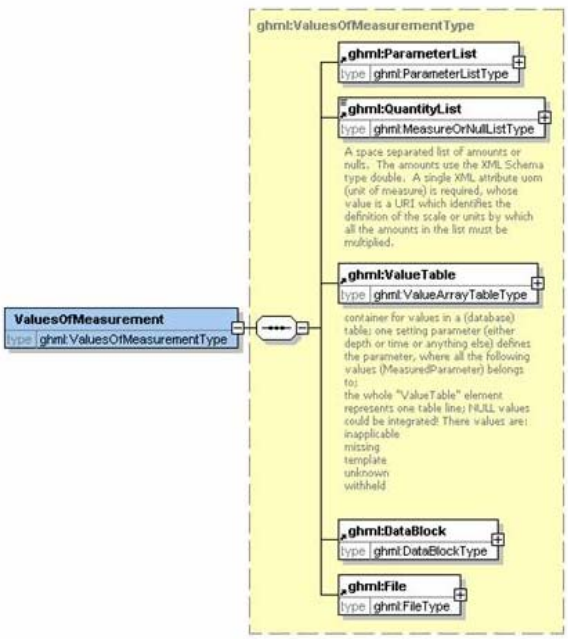

Figure 16. The global "ValuesOfMeasurement" element.

If another description is desired, the extension of this type could be realized by the inclusion of "Data Value Points" (see Figures 15 and 17). The global "DataValuePointType" consists of one mandatory setting parameter (either time or depth) and a number of unbounded related measured parameters. The name of the parameter element describes the name of the parameter (e.g., Ar for Argon) and the attribute the unit of measurement (uom).

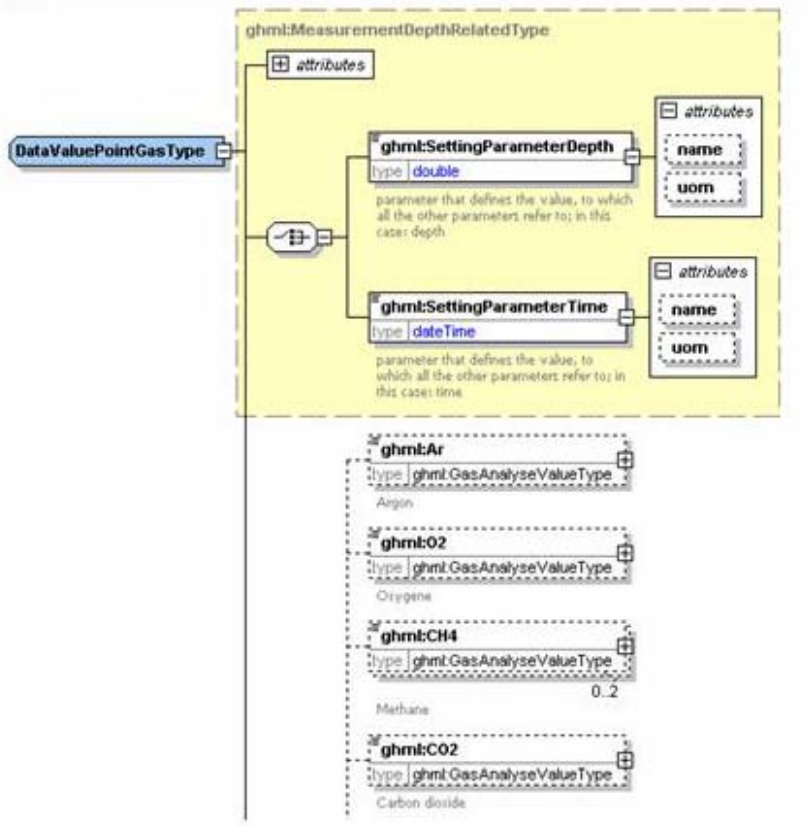

Figure 17. The "DataValuePointType" used and extended for gas analysis.

In general, it is recommended to use the global "MeasureType" for any kind of data values. This type has a simple content (double) and the name and unit of the parameter are attributes. The setting parameter element, as with most of other data value elements, is built with this type. Restriction of possible data inputs is set by enumeration lists, where it is required that the input match an item on the list (e.g., a list of about 50 lithology terms for geological description).

4.0 Modeling Data

The Modeling Data Module of GHML has assimilated information about gas hydrate modeling, consisting of such elements as name, purpose, main theories, and input/output parameters; information generally used in hydrate behavior modeling.

The generation of this part of the GHML was based on software named TOUGH-Fx/Hydrate which is widely used for modeling gas hydrate resources. TOUGH-Fx/Hydrate (now known as TOUGH+/Hydrate) is representative of a state-of-the-art model in this field of research. So, although the GHML used a single software program as a reference, elements abstracted from it are considered to be suitable for most applications.

During the development of this module of the GHML, the authors consulted many international metadata standards (Markup Languages) that provided good references for structure design, naming conventions, annotation formats, etc. Consistency with these existing international standards ensured good communication and transferability between GHML and other international standard markup languages.

This standard specifies a conceptual data schema that defines the structure of a metadata/data instance for gas hydrate modeling. This conceptual data schema doesn't specify distribution, encoding, or means for use of this standard that can be determined from other endorsed standards. In this standard, those characteristics of gas hydrate modeling may be divided into metadata category and data category. The standard can be used for gas hydrate data description, integration, management, and exchange, not only by metadata producers but also by data managers.

4.1 *Basic Structure of Modeling Data Module of GHML*

The base schema defines the elements:

Name - The name by which the element is referenced.

Definition - The definition of the data element.

Datatype - Indicates the value's type (string, decimal, id, etc.).

MaxOccurs - The max permitted occurrence times of an element.

MinOccurs - The minimum permitted occurrence times of an element.

The naming convention in this standard is as follows:

(1) Element

First letter of each concatenated word is capitalized, e.g.: InputParameter.

(2) Complex Type

Each complex type ends with "Type" and the first letter of each concatenated word is capitalized, e.g.: InputType.

(3) Simple Type & Attribute

First letter is lowercase and first letter of each subsequent concatenated word is capitalized, e.g.: valueList.

This standard uses XML Schema to define GHML. XML Schema is a recommended metalanguage by W3C that can define the data element, datatype, value space and element relations conveniently and clearly.

"Namespaces in XML," which will let browsers interpret more than one XML-based language in a single document without confusing different elements with the same tag names. This standard defines the following namespaces:

targetNamespace = http://www.codata.org/ghml/

xmlns:ghml = http://www.codata.org/ghml

4.2 *Detailed Structure of Modeling Data Schema*

The basic structure of this portion of GHML is shown in Figure 18.

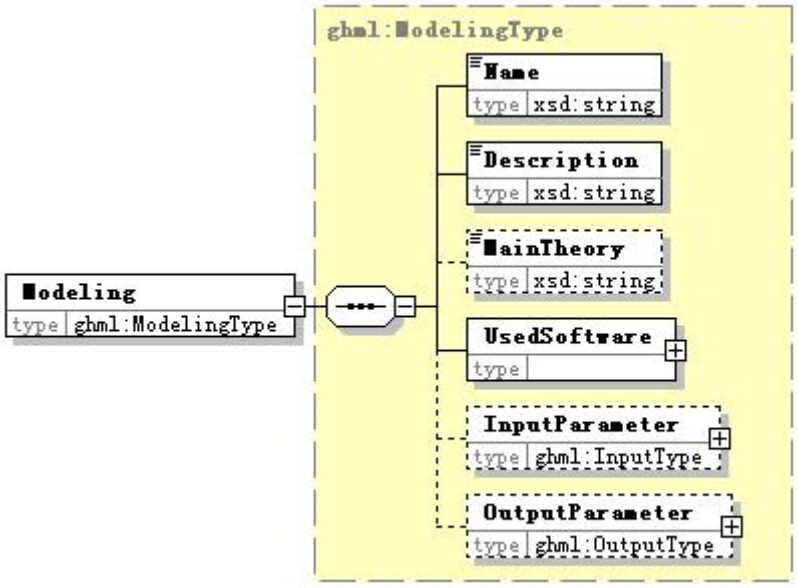

Figure 18. Basic Structure of Modeling Markup Language.

This part of GHML consists of Descriptive elements and data elements:

Descriptive elements:

- **Name[string]** - the name of this model.
- **Description[string]** - Some descriptive information (e.g. equation of state –EOS) and if an inhibitor is included, it should be mentioned by this element.
- **MainTheory[string]** - The main theory used in the model.
- **UsedSoftWare** - this element includes two sub-elements: **Name [string]** -The name of the software with which the model was built. **Version [string]** -Version information of this software.

Data elements:

- **InputParameter[ghml:InputType]**: This is one of the most important elements in this portion which includes five complex elements, each of which answers one important modeling question (Figure 19).

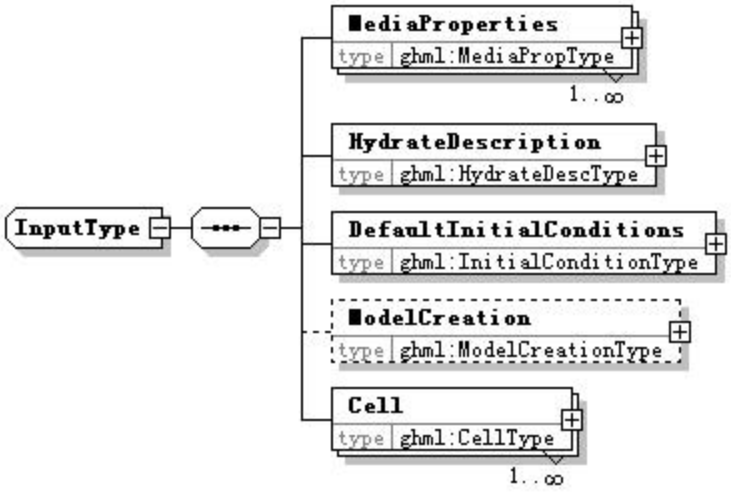

Figure 19. InputType.

- **MediaProperties[MeidaPropType]** -A data block that lists all the parameters that describe the hydraulic, thermal and wettability properties of the geologic medium. (Figure 20).

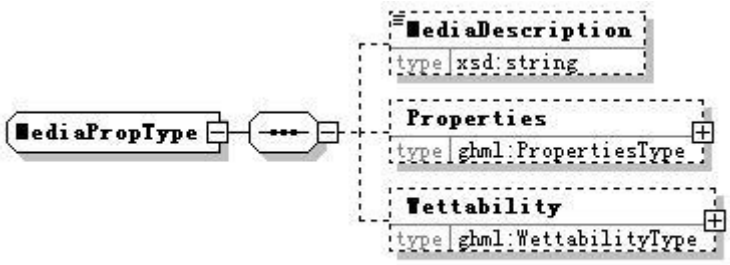

Figure 20. MediaPropType.

  - **MeidaDescription [string]** - Some descriptive information about the media.
  - **Properties [ghml:PropertiesType]** - Physical and chemical properties of the media. This block includes such media properties as: **Density[double]**, **Porosity[double]**, **Permeability[double]**, **WetHearConductivity[double]**, **SpecificHeat[double]** and **ThermalExpensivity[double]**.

- **Wettability [ghml:WettabilityType]** - Describes the porous media response as it is being wetted by fluids.(Figure 21).

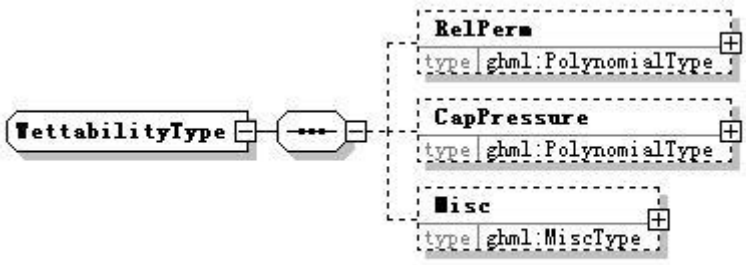

Figure 21. WettabilityType.

- **RelPerm [ghml:PolynomialType]** - 3-phase relative permeability as a function of saturation. This element includes tree sub-element describes the characteristics of the polymonial: **Polynomial [string]** -, **CoefficientNumber[integer]**- The number of coefficients needed by the polynomial and **Coefficient[double]**- Coefficient value.
- **CapPressure [ghml:PolynomialType]**- Capillary pressure. This element uses the same datatype as **RelPerm** we described above.
- **Misc[ghml:MiscType]** – miscellaneous material data.(Figure 22).

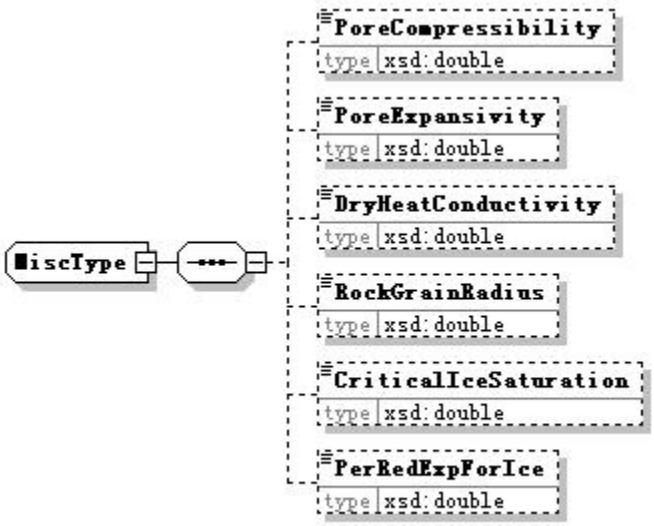

Figure 22. MiscType.

This element consists of **PoreCompressibility[double]**- This defines how the pore volume changes as a function of pressure, **PoreExpansivity[double]**- how the pore volume changes with temperature, **DryHeatConductivity[double]**- Used with the wet heat conductivity to change the thermal conductivity of the rock, **RockGrainRadius[double]**- the radius of the rock grain,

CriticalIceSaturation[double], **PerRedExpForIce[double]**- Permeability reduction exponent for ice.

- **HydrateDescription [HydrateDescType] –** Hydrate properties and behavior (Figure 23).

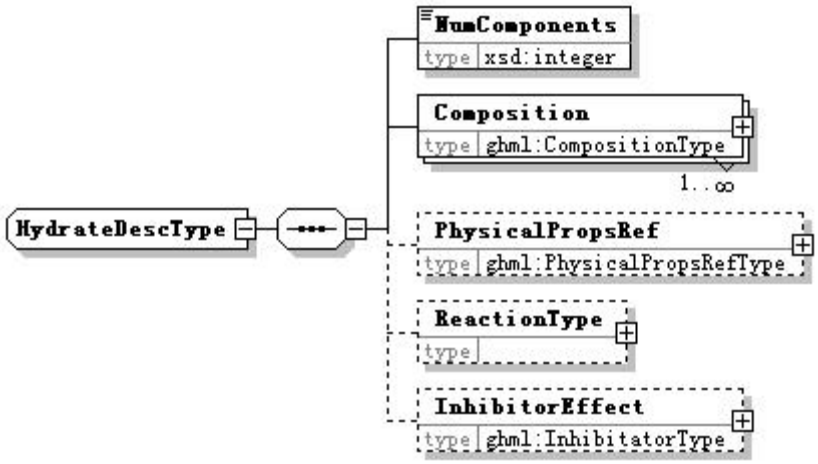

Figure 23. HydrateDescType.

This element includes the following sub-elements:

- **NumComponents[xsd:integer] -** The number of components in a complex hydrate.
- **Composition[ghml:CompositionType]** - Compositon of the complex hydrate (Figure 24).

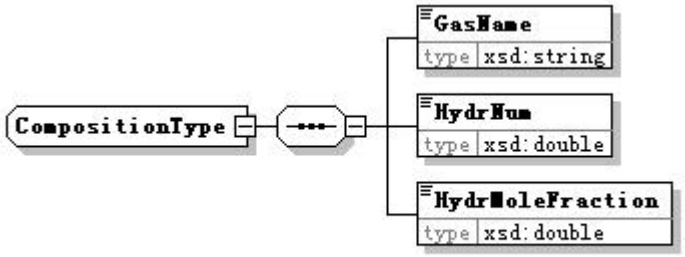

Figure 24. CompositionType.

GasName[String] is the name of the hydrate-forming gas. **HydrNum[double]** is the corresponding hydration number. And **HydrMoleFraction[double]** is Hydrate mole fraction in the complex hydrate.

- **PhysicalPropsRef[ghml:PysicalPropsRefType]** (Figure 25).

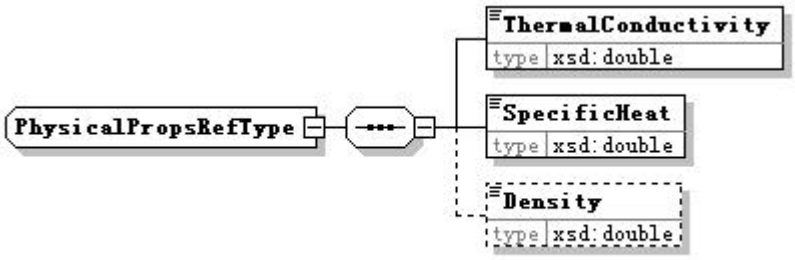

Figure 25. PhysicalPropsRefType.

This block consists of three properties of hydrate:

ThermalConductiveity [double], **SpecificHeat[double]** and **Density[double]**.

- **ReactionType -** This element describes the type of hydrate reaction, and can take one of two values. For simulations under equilibrium conditions, ReactionType is 'Equilibrium'. Kinetic hydrate reactions are considered when reaction type is Kinetic and tree parameters are needed at this circumstance: **ActivationEnergy[double] -** The activation energy for the hydrate dissociation [J/ mol]**, IntrinsicRate[double] -** Intrinsic hydration reaction constant K0 [kg/ (m2. Pa.s)] and **AreaFctor[double]**- Area adjustment factor [dimensionless].
- **InhibitorEffect[ghml:InhibitorType] -** Inhibitor-related data. It includes **TMaxOff [double]**- The inhibitor-induced reference temperature depression, **CMaxOff[double]**- Reference inhibitor mole fraction in the aqueous phase, **MWInhib[double]**- Molecular weight of the inhibitor [g/ mol], **DInhib[double]**- Inhibitor density [kg/ m3], **HInhib[double]**- Specific enthalpy of inhibitor dissolution [kg/ m3] and **DifcoInhib[double]**- Diffusion coefficient of inhibitor in water [m2/ s].

  - **DefaultInitialCondition[DefaultInitialConditions] -** Default global initial conditions. Initial conditions are used to define the initial state of each cell. A hierarchy is used to determine the cell state; if defined at the cell, those values are used; if defined at the region, those are used; finally, the default model initial conditions will be used. Default initial conditions are always defined for the model. This complex element consists of **PhaseType [string]** - Phase state of the modeling, **P [double]** - Pressure, **T [double]** - Temperature, **SH [double]** - Hydrate saturation, **SG [double]** - Gaseous phase saturations, **SA [double]** - Fluid phase saturations, **SI [double]** - Ice phase saturation, **Xinhibitor [double]** - Mole fraction of the inhibitor and **XGasInWater [double] -** Mole fraction of gas in water.

  - **ModelCreation[ModelCreationType] -** How the model was created (Figure 26).

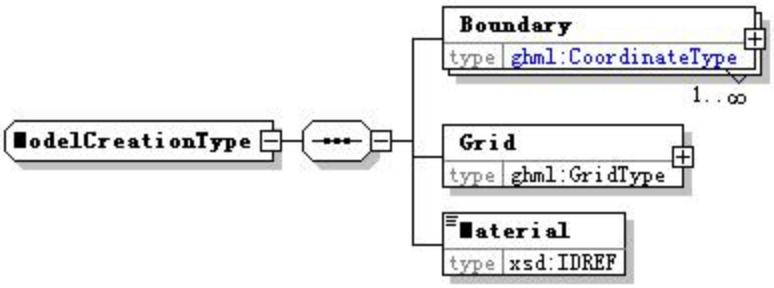

Figure 26. ModelCreationType.

- **Boundary[CoordinateType]** The first step in creating a model is to define the boundary. In this standard, we defined a common datatype - **CoordinateType** to describe all coordinates elements that appear in this standard. CoordinateType consists of five elements: **X[double]**, **Y[double]**, **Z[double]**, **StrikeAzinuth[double]**-The degrees from North (the positive Y axis) in a clockwise direction, and **DipAngle [double]** - Degrees from horizontal of the plane.
- **Grid [GridType],** this element describes how the region is subdivided into sub-regions/cells by indicating the cell number in each direction of increasing coordinate. **XNumber [integer]**-cell number in X axis. **YNumber [integer]**- cell number in Y axis **ZNumber [integer]**- cell number in Z axis.
- **Material [IDREF]** indicates the material that adopted in the model.

  - **Cell [CellType] -** Properties of a cell. Cells can be divided into two types: boundary cells and normal cells (Figure 27).

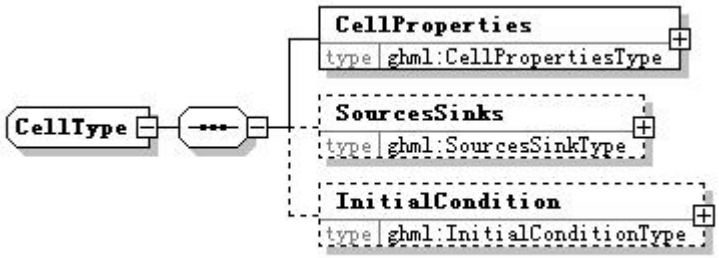

Figure 27. CellType.

  - **CellProperties[CellPropertiesType]-** Geometer properties of the cell (Figure 28).

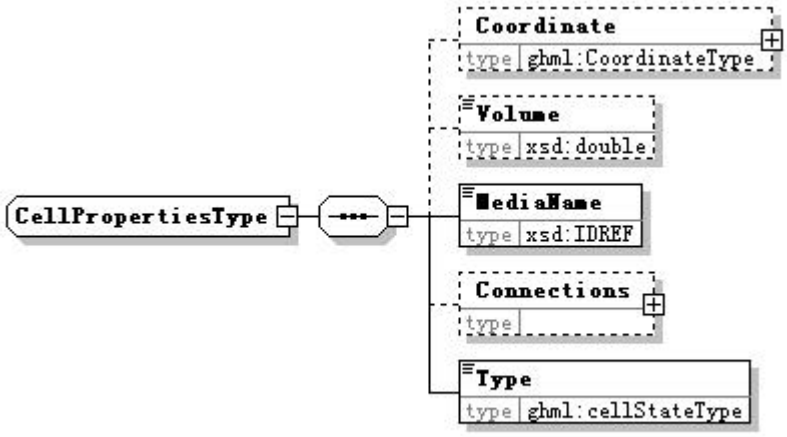

Figure 28. CellPropertiesType.

- **Coordinate [ghml:CoordinateType]**, we have described it above.
- **Volume[double]**- Volume of the cell.
- **MediaName[IDREF]**, the media used in this cell.
- **Connections** - Common connection area over which the elements communicate/interact. This elements has three sub-elements: **Area [double]** - Interface area [m2], **D1[double]** and **D2[double]**- D1 and D2 refer to distance [m] from first and second element, respectively, to their common interface.
- **Type [ghml:cellStateType]**, enumeration(Enabled, Disabled and Fixed State).

- **SourcesSinks[SourcesSinkType]** - This elements is used to define production from or injuction into a cell (Figure 29).

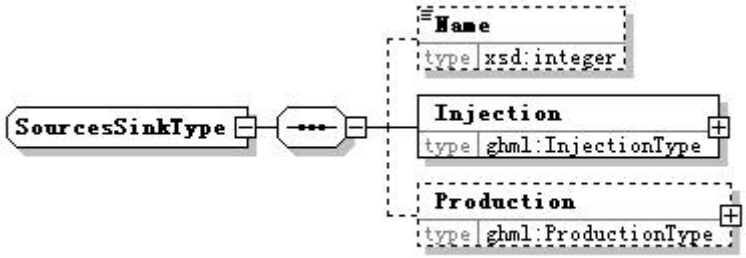

Figure 29. SourcesSinkType.

- **Name[string]** - Name of the well.
- **Injection [ghml:InjectionType]** - used to define production injection into a cell. Injection parameters will vary depending on the EOS being used (Figure 30). In general, the user will specify a rate and an enthalpy for each component to be

injected. The rates can be defined as constant (**Rate [double]** and **Enthalpy [double]**) or using a table to give time/rate pairs.

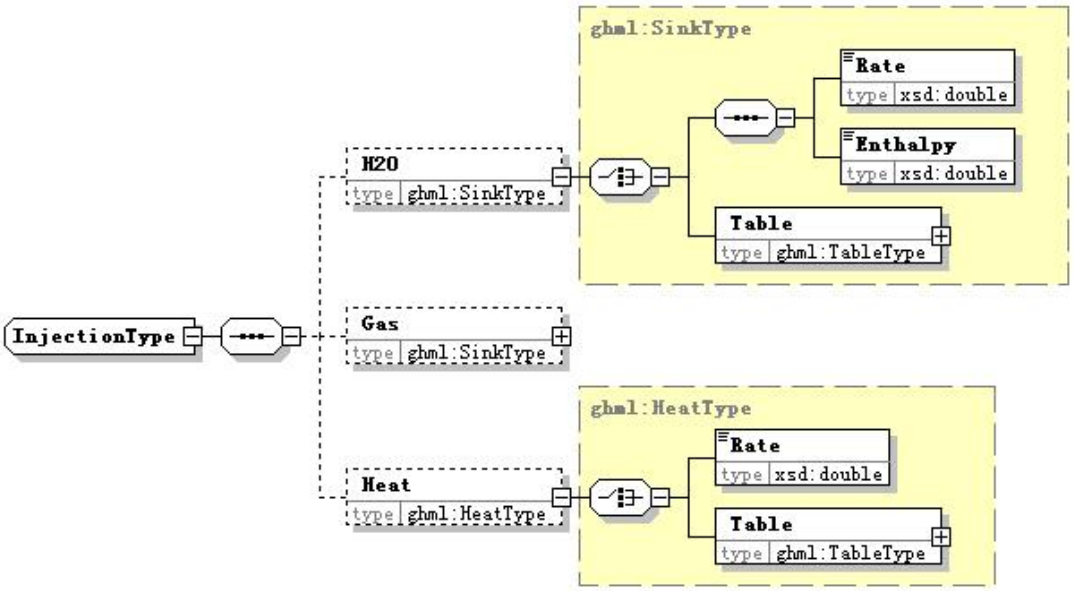

Figure 30. InjectionType.

- **Production[ghml:ProductionType]** - used to define production from a cell. **Mass[ghml:HeatType]** defines the mass produced from the cell. **WellOnDeliv** which includes two sub-elements (**ProcuctivityIndex[integer]** and **Pressure[double]**) defines a boundary condition where the cell produces to a fixed pressure.

- **InitialCondition[InitialConditionType]** - used to define the initial state of each cell. We have described before.

- **OutputParameter[ghml:OutputType]** - describes output parameters of the model (Figure 31). This element includes for complex sub-elements:

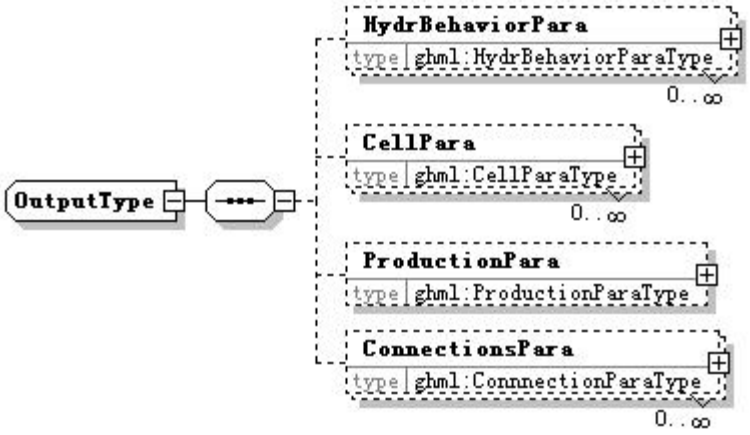

Figure 31. OutputType.

- **HydrBehaviorPara[ghml: HydrBehaviorParaType]** - Output file that includes data describing the evolution of hydrate dissociation over time.
  - **Time[string]** - Simulation time.
  - **Cum_Release_MRate[double]** - Mass rate of CH4 release from dissociation.
  - **Cum_Release_VRate[double]** - Volumetric rate of CH4 release from dissociation.
  - **Cum_Rel_Mass[double]** - Cumulative mass of released gas from the entire domain from the beginning of the simulation.
  - **Cum_Rel_Volume[double]** - Cumulative volume of released gas from the entire domain from the beginning of the simulation.
  - **Free_Reservoir_Gas[double]** - Volume of free gas in the reservoir at the time of observation.
  - **Rem_Hydrate[double]** - Mass of hydrate remaining in the deposit at the time of observation.
- **ProductionPara[ghml:ProductionParaType]** - Data describing gas production (rates and production stream composition at wells).
  - **Time[string]** - Simulation time.
  - **Qm_CH4_prod[double]** - mass rate of CH4 production.
  - **Qv_CH4 _prod[double]** - volumetric rate of CH4 production
  - **CumM_CH4_ prod[double]** - cumulative mass of produced gas since the inception of the simulation.
  - **CumV_CH4_ prod[double]** - cumulative volume of produced gas since the inception of the simulation.
  - **Qm_H2O_prod[double]** - mass rate of water production.

- **CumM_H2O_ prod[double]** - cumulative mass of produced water since the inception of the simulation.

○ **ConnectionsPara[ghml: ConnectionParaType]** - Output file that includes a list describing the evolution of connectionl-related parameters over time.

- **Time** [string]
- **HeatFlux[double]** - Heat flux.
- **GasFlux[double]** - Gas flux.
- **AquFlux[double]**-Liquid flux.
- **Ch4inGas_flux[double]** - Flux of CH4 in the gas phase.
- **CH4inAqu_flux[double]** - Flux of CH4 in the aqueous phase.
- **GasVeloc[double]** - Gas phase velocity.
- **AquVeloc[double]** - Aqueous phase velocity.
- **Gas_phi[double]** - Fugacity of the gas phase.
- **Aqu_phi[double]**-Fugacity of the aqueous phase.

○ **CellPara[ghml:CellParaType]** - Output file that includes a list describing the evolution of cell-related parameters over time.

- **Time[string]**
- **P[double]** - Pressure.
- **T [double]** - Temperature.
- **S_Hydrate[double]** - Hydrate saturation.
- **S_gas[double]** - Gas saturation.
- **S_aqu [double]** - Liquid saturation.
- **S-Ice[double]** - Ice saturation.
- **X_Inhibitor[double]** - Mass Fraction of inhibitor.
- **P_CH4[double]** - Partial pressure of CH4.
- **P_EqHydr[double]** - Equilibrium hydration pressure at temperature T.
- **P_SatWar[double]** - Saturation pressure of water at temperature T.
- **C_CH4inGas[double]** - Concentration of CH4 in the gas phase.
- **C_CH4inAqu[double]** - Concentration of CH4 in the aqueous phase.
- **Dens_Gas [double]** - Density of gas.
- **Dens_Aqu [double]** - Liquid Density.
- **Dens_Hydr[double]** - Hydrate density.
- **Visc_Gas [double]** - Gas viscosity.

- **Visc_Aqu[double] -** Liquid viscosity.
- **Phi [double] -** Porosity.
- **Pcap [double] -** Capillary pressures.
- **Krel_Gas [double] -** Relative permeability of the gas phase.
- **Krel_Aqu [double] -** Relative permeability of the aqueous phase.
- **H_Gas [double] -** Enthalpy of the gas phase.
- **H_Aqu [double] -** Enthalpy of the aqueous phase.
- **H_Hydr [double] -** Enthalpy of the hydrate phase.
- **H_Ice [double] -** Enthalpy of the ice phase.
- **ReacTn_Rate[double] -** Rate of the hydration reaction.
- **Reactn_Heat[double] -** Heat of dissociation/formation.
- **T_shift -** Temperature shift in the hydration reaction caused by inhibitors.

Acknowledgements

The authors gratefully acknowledge support by several agencies. CODATA contributed US$5,000 per year over the course of this project to support travel costs. From 2002 to 2004, ICSU supported the CODATA gas hydrate initiative by providing a US$100,000 grant over two years. It helped in organizing regional meetings in Russia, US, France, China and India (see Table 1). The Computer Network Information Center (CNIC) of the Chinese Academy of Sciences donated two IT professionals to develop the GHML. GeoForschungsZentrum (GFZ) - Potsdam donated one full time IT professional to the project in 2006. The U.S. Department of Energy has supported the participation of an IT consultant and provided partial travel support for some members of the CODATA task group.

References

Bernstein, W., (2004): *The Birth of Plenty*, New York, McGraw-Hill Co.

Cox, S., Daisey, P., Lake, R., Portele, C. and Whiteside, A., (2004): *Implementation Specification - ISO/TC 211/WG 4/PT 19136 Geographic information – Geography Markup Language (GML),* version 3.1.0, OpenGis© Recommendation Paper.

Dallimore, S.R., Collett, T.S., (Eds.),(2005): *Scientific Results from the Mallik 2002 Gas Hydrate Production Research Well Program, Mackenzie Delta, Northwest Territories, Canada*, Ottawa, Canada, Geological Survey of Canada Bulletin 585.

Dublin Core Metadadta, *http://dublincore.org/.*

Examples Using the PetraSim Pre- and Post-Processor for TOUGH-Fx/HYDRATE, *http://www.thunderheadeng.com/petrasim/manual_tetrad.pdf.*

Frenkel, M., Chirico, R.D., Diky, V.V., Dong,Q., Marsh, K.N., Dymond, J.H., Wakeham, W.A., Stein, S.E., Königsberger, Goodwin, R.H., (2006): *XML-based IUPAC standard for experimental predicted and critically evaluated thermodynamics property data storage and capture (ThermoML),* Pure Appl. Chem., 78(3), 541-512, doi:10;1351/pac200678030541.

GASHYDAT, (2001): initial hydrate database generated by J. Klerkx and J. Dimitrov, found at www.gashydat.org/.

ISO 19115 Geographic information – Metadata, *http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020&ICS1=35&ICS2=240&ICS3=70.*

Klauda, J., B., Sandler, S.I., (2005): *Global distribution of methane hydrate in ocean sediment*, Energy & Fuels, 19, 459-470.

Klump J., Bertelmann R., Brase J., Diepenbroek M., Grobe H., Höck H., Lautenschlager M., Schindler U., Sens I. and Wächter J., (2006): *Data Publication in the open access Initiative*, Data Science Journal, Volume 5, pages 79-83.

Löwner R. and Conze R., (2005): *Mallik Data and Information System – development of a scientific data exchange platform"; in Scientific Results from the Mallik 2002 Gas Hydrate Production Research Well Program, Mackenzie Delta, Northwest Territories, Canada*, (ed.) S.R. Dallimore and T.S. Collett; Geological Survey of Canada, Bulletin 585, 9 p.

Löwner, R., Cherkashov, G., Pecher, I., Makogon, Y. F., (2007): *Field Data and the Gas Hydrate Markup Language*, proceedings of the 20th International CODATA Conference, 22-25 October 2006, Beijing, China.

Makogon, Y.F., (1997): *Hydrates of Hydrocarbons*, Tulsa, Oklahoma, PennWell Books

Milkov, A.V. (2004): *Global estimates of hydrate-bound gas in marine sediments: How much is really out there?,* Earth Science Reviews, 66 (3-4) 183-197.

Radler, M., (2000): *World crude and natural gas reserves rebound in 2000*, Oil and Gas J., 98(51) p. 121-123.

Sachs, J.D., (2005): *The End of Poverty*, New York, Penguin Press.

Sloan, E.D., (1998): *Clathrate Hydrates of Natural Gases*, (2nd Ed.) New York, Marcel-Dekker, Inc.

Sloan, E.D., (2004): *Introductory overview: hydrate knowledge development,* American Mineralogist, 89, 1155-1161

Smith, T., Ripmeester, J., Sloan, D., Uchida, T., (2007): *Gas Hydrate Markup Language as it pertains to laboratory data*, proceedings of the 20th International CODATA Conference, 22-25 October 2006, Beijing, China.

Taylor, B., Barry, Kuyatt, E., Chris, (1994): *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, Physics Laboratory, National Institute of Standards and Technology, NIST Technical Note 1297.

ThermoML – An XML-Based Approach for Storage and Exchange of Experimental and Critically Evaluated Thermophysical and Thermochemical Property Data, *http://trc.nist.gov/ThermoML.html*.

Wang, W., Moridis, G., Wang, J., Xiao, Y., Li, J., (2007): *Modeling hydrates and the gas hydrate markup language*, proceedings of the 20th International CODATA Conference, 22-25 October 2006, Beijing, China.

van der Vlist E., (2002): *XML schema*, O'Reilly & Associates, Inc., First Addition, ISBN: 0-596-00252-1.