

Представление данных по  
теплофизическим свойствам веществ с  
использованием концепций и методов  
***Semantic WEB***

О. М. Атаева (ВЦ РАН),  
А. О. Еркимбаев, В. Ю. Зицерман,  
Г. А. Кобзев (ОИВТ РАН),  
В. А. Серебряков, К. Б. Теймуразов  
(ВЦ РАН), Р. И. Хайруллин (МФТИ)

- ❑ С учетом собственного опыта и круга интересов авторы выбрали в качестве конкретной области **теплофизику**, точнее теплофизические свойства веществ и материалов.
- ❑ Представляется особенно полезным отладить применение концепции в относительно узкой предметной области, с характерными для нее типами данных и стандартами их представления в БД и печатных источниках.
- ❑ Технология и практика интеграции данных с использованием концепции **Semantic WEB** находятся еще в стадии становления, несмотря на отдельные успешные примеры.

- ❑ Теплофизические данные (сжимаемость, энергетические и транспортные свойства), представлены в мировых БД для широчайшего круга веществ: чистых и растворов, органических и неорганических, наноструктур и материалов.
- ❑ Для каждого класса существуют свои модели, связанные с ними словари понятий и логические структуры данных.
- ❑ Особенность теплофизических данных – их преимущественное использование для работы вычислительных приложений, обеспечивающих моделирование различных технологических и природных процессов.
- ❑ Как следствие, потребность в методах обмена неоднородными данными, различающимися форматом и структурой, возникла в теплофизике задолго до того, как проблема интеграции приобрела актуальность для информационного сообщества.

*Особую актуальность проблеме интеграции теплофизических данных придает:*

- ❑ Массовое распространение теплофизических БД в науке, промышленности и образовании.
- ❑ Повышенное внимание к **метрологическим аспектам**, включающим накопление **больших массивов первичных данных**, изучение их неопределенности, воспроизводимости, согласованности и т. п.

**В тематике ОИВТ РАН подготовка справочных данных по свойствам всегда занимала важное место.**

- Широкую известность приобрели БД: **ИВТАНТЕРМО, ТЕРМАЛЬ, ЭПИДИФ, БД ударно-волновых экспериментов**, включающие обширные массивы информации по теплофизическим и термодинамическим свойствам.
- Накоплены обширные фонды данных по молекулярным постоянным, константам элементарных процессов в газах и плазме, спектрам атомов и ионов.
- Разработаны компьютерные системы хранения и обработки слабо структурированных данных, характерных для наноструктур и наноматериалов.

# Как и для других типов научных ресурсов, интеграция теплофизических данных включает:

- ❑ поддержку данных из множества неоднородных источников в терминах единой модели;
- ❑ стандартизацию процессов обмена;
- ❑ возможность публикации данных в WEB (Linked Open Data) с открытым доступом как для человека, так и для программных агентов.

## Цели работы:

- построить онтологию предметной области (при наложении некоторых ограничений);
- выявить на конкретном примере (теплофизических данных) возможности и преимущества онтологического моделирования

Онтология (по Груберу) это явная формальная **спецификация** разделяемой **концептуализации** предметной области

**□ Концептуализация** —  
строгое описание понятий предметной области, их связей и отношений средствами естественного языка.

**□ Спецификация** -  
означает формализацию понятий и связей, например в терминах OWL.

Хотя функция онтологии аналогична схеме БД, язык определения онтологий синтаксически и семантически богаче

### В практическом плане:

- ❑ онтология дает средства для совместного использования информации (технология «связанных данных», LOD);
- ❑ возможность за счет расширяемости онтологии и используемых словарей поддерживать данные с *a priori* неизвестной структурой (полуструктурированные данные).

**Концептуализация** (применительно к теплофизике) сводится к выделению нескольких базовых понятий. Прежде всего, это:



## Дополнительные понятия, уточняющие набор данных: **неопределенность, источник, статус.**

- ❑ Сведения о **неопределенности** включают: тип (среднеквадратическая, расширенная с указанием уровня значимости и т.п.) и значение. Помимо этого, оценка может относиться к набору в целом, отдельным свойствам из набора и даже зависеть от значения аргумента.
- ❑ Информация об **источнике**, стандартизованная (или нестандартизованная) библиографическая запись. Данные об источнике могут относиться к набору в целом, к отдельным свойствам из набора, или различаться для значений аргумента.
- ❑ **Статус** данных: экспериментальные, расчетные, справочные; для экспериментальных оправдано расширить набор данных сведениями об условиях эксперимента (метод измерений, прибор, подготовка образца и т.п.).

# Типовая форма набора данных

(из учебной версии БД ИВТАНТЕРМО)

- Термодинамические свойства воды в состоянии идеального газа
- Набор данных включает 2 константы и 4 функции от температуры (таблица из 5 столбцов)
- Пропущены сведения по источнику и неопределенности данных

**Water H<sub>2</sub>O(g)**

$\Delta_f H_0^*$	-238.923 kJ/mol	$\Delta_f H_{298}^*$	-241.826 kJ/mol
------------------	-----------------	----------------------	-----------------

T	C <sub>p</sub>	F	S	H
K	J/K*mol	J/K*mol	J/K*mol	kJ/mol
298.15	33.596	155.502	188.722	9.905
300.00	33.605	155.708	188.930	9.967
400.00	34.292	165.287	198.683	13.358
500.00	35.254	172.766	206.434	16.834
600.00	36.360	178.935	212.958	20.414
700.00	37.556	184.210	218.652	24.109
800.00	38.808	188.839	223.748	27.927
900.00	40.091	192.980	228.393	31.872
1000.00	41.385	196.738	232.684	35.946
1100.00	42.670	200.190	236.689	40.148
1200.00	43.927	203.390	240.456	44.479
1300.00	45.138	206.380	244.020	48.932
1400.00	46.287	209.190	247.407	53.504
1500.00	47.354	211.846	250.638	58.187
1500.00	47.273	211.846	250.634	58.181
1600.00	48.369	214.368	253.721	62.965
1700.00	49.344	216.770	256.683	67.851

# Пример более сложной организации набора данных - значения второго вириального коэффициента $B(T)$

**Argon** (cont.)

**Table 2.** (cont.)

$T$ К	$B_{\text{exp}} \pm \delta B$ $\text{cm}^3 \cdot \text{mol}^{-1}$	$B_{\text{exp}} - B_{\text{calc}}$ $\text{cm}^3 \cdot \text{mol}^{-1}$	Ref. (Symbol in Fig. 1)	$T$ К	$B_{\text{exp}} \pm \delta B$ $\text{cm}^3 \cdot \text{mol}^{-1}$	$B_{\text{exp}} - B_{\text{calc}}$ $\text{cm}^3 \cdot \text{mol}^{-1}$	Ref. (Symbol in Fig. 1)
398.15	$-1.2 \pm 1.0$	0.2	49-mic/wij(□)	573.15	$10.8 \pm 0.1$	0.1	67-kal/mil(■)
400.00	$-0.9 \pm 0.1$	0.3	92-ewi/tru-1(▲)	573.15	$10.6 \pm 0.1$	0.0	67-kal/mil(■)
400.00	$-1.1 \pm 0.1$	0.1	96-est/tru(●)	600.00	$11.8 \pm 0.1$	0.0	92-ewi/tru-1(▲)
423.15	$1.4 \pm 1.0$	0.4	49-mic/wij(□)	673.15	$14.3 \pm 0.1$	-0.2	67-kal/mil(■)
423.15	$1.4 \pm 1.0$	0.3	49-mic/wij(□)	673.15	$14.2 \pm 0.1$	-0.3	67-kal/mil(■)
450.00	$3.3 \pm 0.1$	0.0	96-est/tru(●)	700.00	$15.1 \pm 0.1$	-0.2	92-ewi/tru-1(▲)
473.15	$5.1 \pm 0.1$	0.1	67-kal/mil(■)	773.15	$17.1 \pm 0.1$	-0.1	67-kal/mil(■)
473.15	$5.2 \pm 0.1$	0.2	67-kal/mil(■)	773.15	$17.0 \pm 0.1$	-0.2	67-kal/mil(■)
500.00	$7.0 \pm 0.1$	0.2	92-ewi/tru-1(▲)				

**Further references:** [10-omn/cro, 25-hol/ott, 30-tan/mas, 53-wha/lup, 56-cot/ham, 60-lec, 62-fen/hal, 62-poo/sav, 66-cra/son, 67-wei/wyn, 68-byr/jon, 69-lic/sch, 70-bla/hal, 70-bos/col, 71-pro/can, 72-osb, 73-pop/cha, 74-bel/rei, 74-hah/sch, 74-sch/heb, 76-san/uri, 77-ren/sch, 77-sch/sch, 79-ewi/mar, 79-sch/leu-1, 80-per/sch, 80-sch/geh, 80-woo/kro, 82-ker, 84-ker/hae, 88-pat/jof, 89-ewi/owu, 91-lop/roz].

# Схема данных включает также древовидный перечень состояний вещества

- три агрегатных состояния;
- линии равновесия типа «жидкость-пар»;
- особые точки (тройная и критическая);
- выделение типа кристаллической решетки (кубическая, тетрагональная и т.п.) для твердого состояния.

# Логические связи между свойством вещества и его состоянием

- Пример 1. Свойство «вязкость» приложимо к веществу только в состояниях «газ» или «жидкость».
- Пример 2. Существует набор свойств, применимых к веществу, находящемуся только на линии фазового равновесия (скажем, «энтальпия испарения»).

Множество состояний вещества с наличием линий или точек равновесия идеальным образом отображается в онтологии с ее иерархией классов

- ❑ Все понятия, относящиеся к линии равновесия, принадлежат к классу, который является подклассом двух родительских классов.
- ❑ Каждый из этих классов объединяет понятия, принадлежащие родительским, и включает понятия, принадлежащие только ему.



Возможности онтологии в передаче логической структуры данных значительно богаче тех, которые предусмотрены в реляционной модели или модели «сущность-связь»

Специфика теплофизических данных - необходимость в **ОНТОЛОГИИ** учитывать, наравне с логическими ограничениями (типа «если....то»), те ограничения, которые связаны с **физическими принципами**

### Характерные примеры:

- равенство энергий Гиббса на линии фазового равновесия;
- связь энергии Гиббса, энтропии и энтальпии;
- интегральная связь энтальпии и теплоемкости

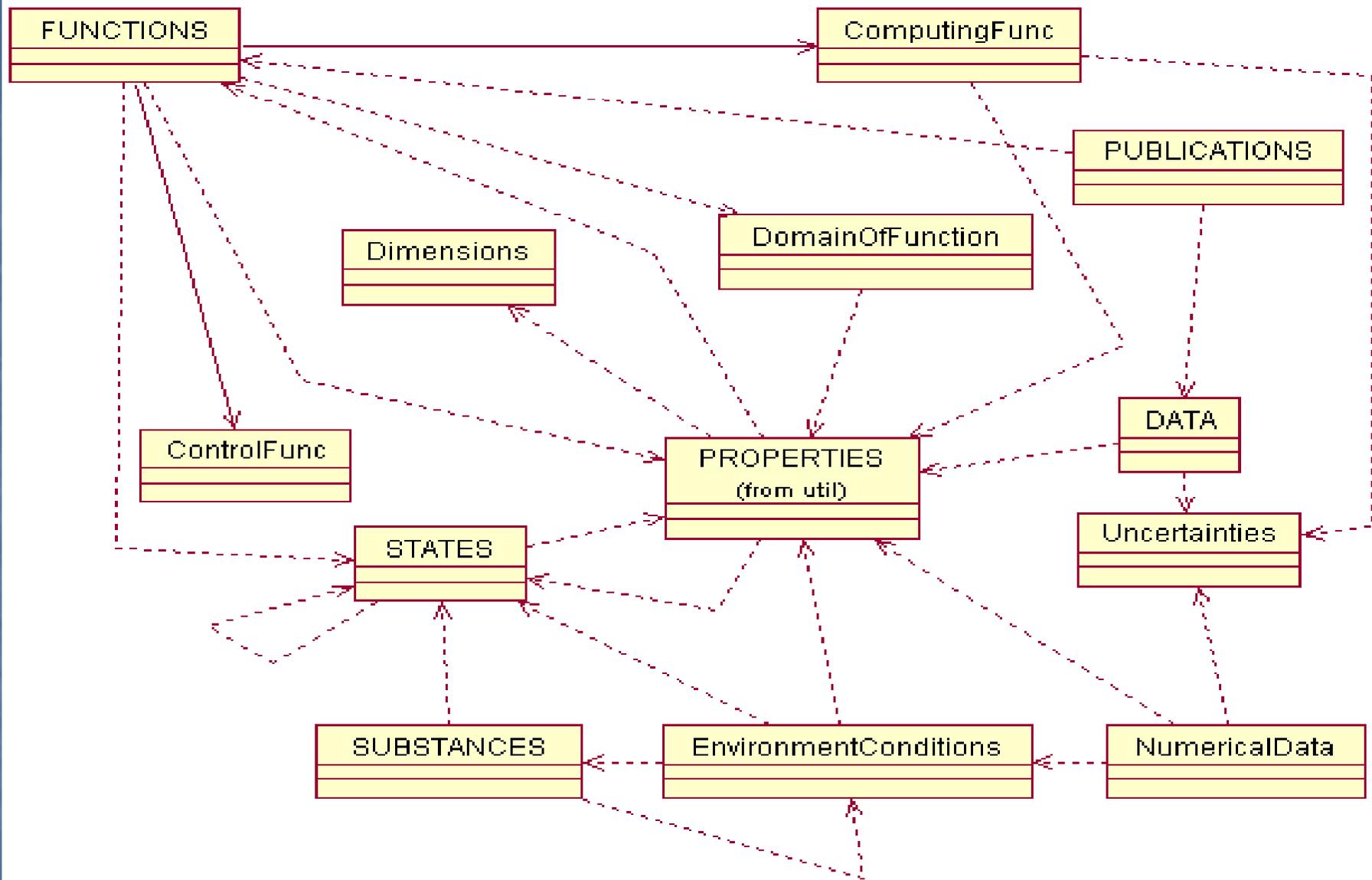
Реалистичность в построении онтологии обеспечивает стратегия “**bottom-up**” – сужение предметной области по ряду критериев с сохранением возможности последующего расширения.

- ❑ Отказ от формализации всех аспектов деятельности по подготовке данных (выбор моделей, постановка экспериментов и проч.) – только представление данных: исходных или полученных в ходе обработки.
- ❑ Два физических ограничения – (1) данные только для чистых веществ, исключая смеси и растворы; (2) условия и состояния, позволяющие игнорировать зависимость свойств от давления.
- ❑ Предложенные ограничения, не будучи универсальными, позволяют охватить значительный объем данных, публикуемых или представленных в электронных ресурсах.
- ❑ За счет этих ограничений все многообразие данных сведено к набору функций одной переменной, представленных в табличной или аналитической формах.

# Второй этап построения онтологии – спецификация концептуализации с выделением классов и свойств

## Классы онтологии

- **Substances, States, Properties, NumericalData.**
- **Dimensions, Uncertainties, EnvironmentConditions.**
- **Publication, Data** (внешние источники).
- **Functions, DomainOfFunctions** (аналитическое представление свойств и контроль области определения функций).
- Классы **ComputingFunc** и **ControlFunc** потомки по отношению к классу **Functions**.
- **ComputingFunc** - объединяет функции, используемые для вычисления свойств.
- **ControlFunc** - объединяет булевские функции, используемые для контроля соотношений между свойствами, которые следуют из физических принципов.



# Технология интеграции и связывания данных

- ❑ Конечная цель онтологии – обеспечить возможности интеграции, то есть связывания разнородных данных, отвечающих предметной области с учетом наложенных ограничений.
- ❑ Предложена технология **Linked Open Data (LOD)**, являющаяся порождением и развитием Semantic Web.
- ❑ Публикация в среде LOD предоставляет издателям и потребителям информации значительно больше возможностей в сравнении с простым размещением данных в традиционном “web of documents”.
- ❑ В отличие от гипертекста, где *линки* связывают отдельные документы, записанные в HTML, технология **LOD** обеспечивает связи между произвольными *сущностями*, отмеченными в документе посредством URI для идентификации любых объектов, персон или концепций.

Спасибо за внимание!