



**Опыт построения информационной системы на основе онтологии *информационных ресурсов* количественной предметной области**

**Фазлиев А.З., Привезенцев А.И., Лаврентьев Н.А.  
(ИОА СО РАН)**



## Содержание

Абстрактная структура информационной системы

(данные, приложения и интерфейсы пользователя)

Слои информационной системы (e-Science)

Опыт построения слоя знаний в ИС с помощью OWL DL.

**Tbox.** Вычислительные задачи в области молекулярной спектроскопии. Данные и метаданные.

**Abox.** Данные измерений и вычислений. Открытый или замкнутый мир?

Отношения транзитивности и рефлексивности в количественной спектроскопии.

Примеры. Ограничения на значения и существование, среднеквадратические отклонения, упорядочение сущностей по значениям свойств.



## Абстрактная структура информационной системы

- Данные
- Приложения
- Интерфейсы

*Guarino N.* Formal Ontology and Information Systems // Proceedings of FOIS'98. Trento. Italy, 6–8 June 1998. Amsterdam: IOS Press. P. 3–15.



## Слои информационной системы (e-Science)

- Данные и вычисления
- Метаданные (информационный слой)
- Онтологии (слой знаний)

*De Roure D., Jennings N., Shadbolt N. A Future e-Science Infrastructure // Report commissioned for EPSRC/DTI Core e-Science Programme. 2001. 78 p.*



Опыт построения слоя знаний в ИС с помощью **OWL DL**

## Предметные области ИС

“**W**ater **@**ccessible **D**istributed **I**nformation **S**ystem”

D = D1 & D2 & D3

D1 – Молекулярная спектроскопия

D2 – Атомы и молекулы

D3 - Публикации



Открытый или замкнутый мир?

Точка зрения пользователя

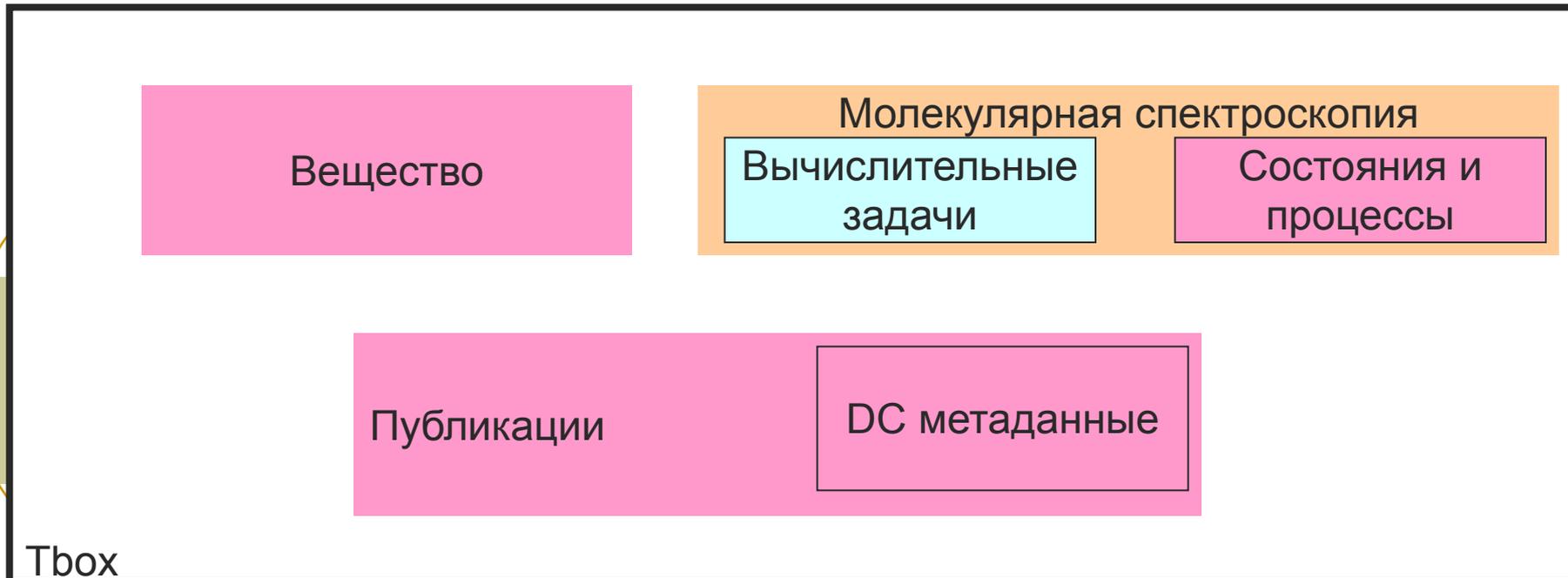
## **Tbox**

- Замкнутое множество понятий ПО
- Расширяемое множество понятий ПО

## **Abox**

- Возможность дополнения фактов (измерения и вычисления)
- Фиксация числа фактов





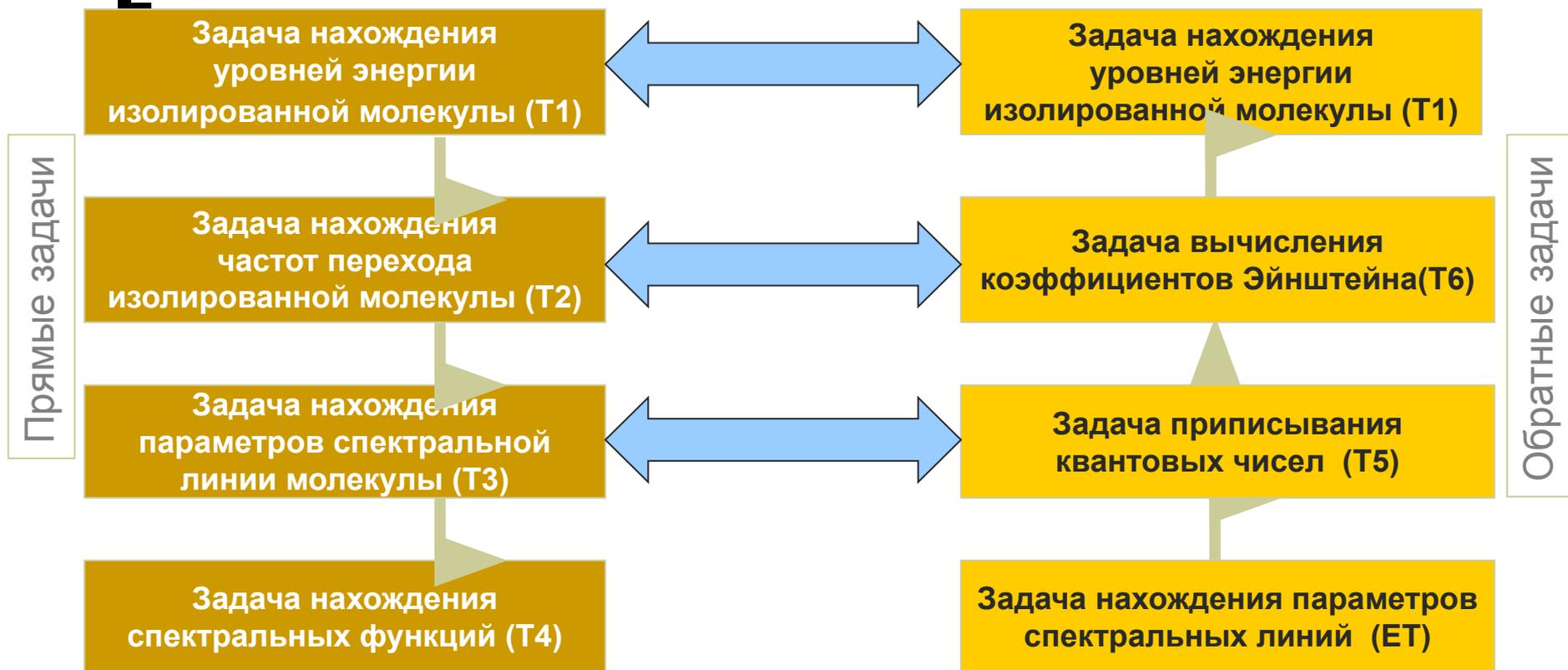
# Модель молекулярной спектроскопии

(0-ое приближение)



# Модель молекулярной спектроскопии

(1-ое приближение)



Стандартные отклонения



# Абох. Данные измерений и вычислений.

## Индивид «information source V4\_T5\_279\_1998\_ToBr\_H2\_170-H2O»

V4\_T5\_279\_1998\_ToBr\_H2\_170-H2O

isSolutionOf T5

hasMethod UNDEFINED

hasSubstance H2\_170

hasOutputData\_MD V4\_T5\_279-T5\_OutputData\_MD

hasReference **Toth R.A., Brown L.R., Self-broadened widths and frequency shifts of water vapor lines between 590 and 2400 cm<sup>-1</sup>. // Journal of Quantitative Spectroscopy and Radiation Transfer, 1998, v.59, p.529-562.**

V4\_T5\_279-T5\_OutputData\_MD

hasBroadeningSubstance\_MD

V4\_T5\_279\_BroadeningSubstances\_MD\_for\_H2O

hasWavenumbers\_MD V3\_T5\_279\_Wavenumbers\_MD

hasTransitions\_MD

V4\_T5\_279\_Transitions\_MD\_for\_NormalModes

hasPhysicalCondition\_MD

V4\_T5\_279\_PhysicalCondition\_MD

hasIntensity\_MD V3\_T5\_279\_Intensity\_MD

V4\_T5\_279\_BroadeningSubstances\_MD\_for\_H2O

hasBroadeningSubstance H2O

hasHalfwidth\_MD V3\_T5\_279\_Halfwidth\_MD\_for\_H2O

hasPressure\_MD

V4\_T5\_279\_PressureValue\_MD\_for\_H2O

hasPressureDependence\_MD

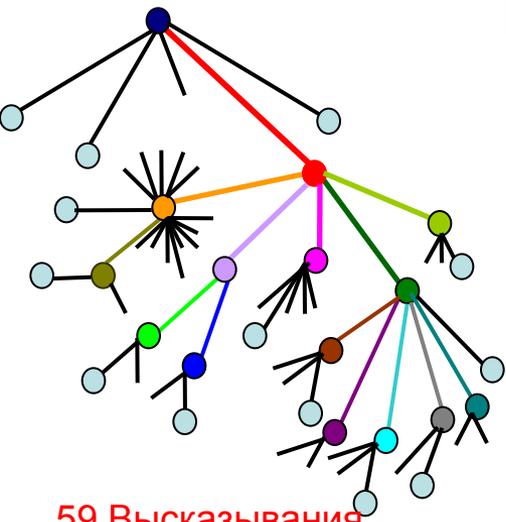
V4\_T5\_279\_PressureDependence\_MD\_for\_H2O

hasShift\_MD V3\_T5\_279\_Shift\_MD\_for\_H2O

hasTemperatureDependence\_MD

V4\_T5\_279\_TemperatureDependence\_MD\_for\_H2O

Субъектно-предикатная структура



59 Высказывания

V4\_T5\_279\_PhysicalCondition\_MD

hasTemperature\_MD

V4\_T5\_279\_TemperatureValue\_MD

hasPressure\_MD

V4\_T5\_279\_PressureValue\_MD

V4\_T5\_279\_PressureValue\_MD

hasUnit atm

hasFloatValue 1

V4\_T5\_279\_TemperatureValue\_MD

hasUnit K

hasFloatValue 296

H2O

hasSymmetryGroup C2v

hasPhysicalState SingleMolecule

V4\_T5\_279\_Intensity\_MD

hasUnit cm-1\_molecule

hasUncertainty false  
isPresented false

V4\_T5\_279\_Wavenumbers\_MD

hasUnit cm-1

hasUncertainty false  
hasMaxWavenumber 2010.911865  
hasMinWavenumber 1315.606567  
hasNumberOfWavenumbers 142

V4\_T5\_279\_Halfwidth\_MD\_for\_H2O

hasUnit cm-1\_atm-1

hasUncertainty true  
isPresented true

V4\_T5\_279\_PressureValue\_MD\_for\_H2O

hasUnit atm

hasFloatValue 1

V4\_T5\_279\_PressureDependence\_MD\_for\_H2O

hasUncertainty false  
isPresented false

V4\_T5\_279\_Shift\_MD\_for\_H2O

hasUnit cm-1\_atm-1

hasUncertainty true  
isPresented true

V4\_T5\_279\_TemperatureDependence\_MD\_for\_H2O

hasUncertainty false  
isPresented false

V4\_T5\_279\_Transitions\_MD\_for\_NormalModes

hasSpectralBand

V4\_T5\_279\_for\_NormalModes\_v1UP\_v2UP\_v3UP\_v1LOW\_v2LOW\_v3LOW\_SpectralBand

hasQuantumNumbersType NormalModes  
hasTotalMaxAngularMomentum 12 hasTotalMinAngularMomentum 0  
hasNumberOfInvalidTransitions 0 hasNumberOfValidWater-C2V-Transitions 142  
hasNumberOfRejectedTransitions 0 hasNumberOfUnassignedTransitions 0  
hasNumberOfValidTransitions 142 hasNumberOfUniqueTransitions 142  
hasNumberOfValidIdentifications 142 hasNumberOfInvalidWaterTransitions 0  
hasNumberOfInvalidWater-C2V-Transitions 0 hasNumberOfInvalidIdentifications 0  
hasNumberOfNonuniqueTransitions 0

V4\_T5\_279\_for\_NormalModes\_v1UP\_v2UP\_v3UP\_v1LOW\_v2LOW\_v3LOW\_SpectralBand

hasBandType  
v1UP\_v2UP\_v3UP\_v1LOW\_v2LOW\_v3LOW  
hasNumberOfSpectralBands 1

# Индивид «information source 1998\_ToBr\_H2\_170-H2O»

## V4\_T5\_279\_1998\_ToBr\_H2\_170-H2O

isSolutionOf T5 hasMethod UNDEFINED hasSubstance H2\_170 hasOutputData\_MD V3\_T5\_279-T5\_OutputData\_MD  
 hasReference Toth R.A., Brown L.R., Self-broadened widths and frequency shifts of water vapor lines between 590 and 2400 cm<sup>-1</sup>. // Journal of Quantitative Spectroscopy and Radiation Transfer, 1998, v.59, p.529-562.

## V4\_T5\_279-T5\_OutputData\_MD

hasBroadeningSubstance\_MD  
 V4\_T5\_279\_BroadeningSubstances\_MD\_for\_H2O  
 hasWavenumbers\_MD V3\_T5\_279\_Wavenumbers\_MD  
 hasTransitions\_MD  
 V4\_T5\_279\_Transitions\_MD\_for\_NormalModes  
 hasPhysicalCondition\_MD  
 V4\_T5\_279\_PhysicalCondition\_MD  
 hasIntensity\_MD V3\_T5\_279\_Intensity\_MD

## V4\_T5\_279\_BroadeningSubstances\_MD\_for\_H2O

hasBroadeningSubstance H2O  
 hasHalfwidth\_MD V3\_T5\_279\_Halfwidth\_MD\_for\_H2O  
 hasPressure\_MD V3\_T5\_279\_PressureValue\_MD\_for\_H2O  
 hasPressureDependence\_MD  
 V4\_T5\_279\_PressureDependence\_MD\_for\_H2O  
 hasShift\_MD V3\_T5\_279\_Shift\_MD\_for\_H2O  
 hasTemperatureDependence\_MD  
 V4\_T5\_279\_TemperatureDependence\_MD\_for\_H2O

## V4\_T5\_279\_PhysicalCondition\_MD

hasTemperature\_MD  
 V4\_T5\_279\_TemperatureValue\_MD  
 hasPressure\_MD  
 V4\_T5\_279\_PressureValue\_MD

## H2O

hasSymmetry Group C2v  
 hasPhysicalstate SingleMolecule

## V4\_T5\_279\_Halfwidth\_MD\_for\_H2O

hasUnit cm-1\_atm-1  
 hasUncertainty true  
 isPresented true

## V4\_T5\_279\_PressureValue\_MD

hasUnit atm  
 hasFloatValue 1

## V4\_T5\_279\_Intensity\_MD

hasUnit cm-1\_molecule  
 hasUncertainty false  
 isPresented false

## V4\_T5\_279\_Shift\_MD\_for\_H2O

hasUnit cm-1\_atm-1  
 hasUncertainty true  
 isPresented true

## V4\_T5\_279\_TemperatureValue\_MD

hasUnit K  
 hasFloatValue 296

## V4\_T5\_279\_Wavenumbers\_MD

hasUnit cm-1  
 hasUncertainty false  
 hasMaxWavenumber 2010.911865  
 hasMinWavenumber 1315.606567  
 hasNumberOfWavenumbers 142

## V4\_T5\_279\_PressureDependence\_MD\_for\_H2O

hasUncertainty false  
 isPresented false

## V4\_T5\_279\_TemperatureDependence\_MD\_for\_H2O

hasUncertainty false  
 isPresented false

## V4\_T5\_279\_Transitions\_MD\_for\_NormalModes

hasSpectralBand  
 V4\_T5\_279\_for\_NormalModes\_v1UP\_v2UP\_v3UP\_v1LOW\_v2LOW\_v3LOW\_SpectralBand  
 hasQuantumNumbersType NormalModes  
 hasTotalMaxAngularMomentum 12 hasTotalMinAngularMomentum 0  
 hasNumberOfInvalidTransitions 0 hasNumberOfValidWater-C2V-Transitions 142  
 hasNumberOfRejectedTransitions 0 hasNumberOfUnassignedTransitions 0  
 hasNumberOfValidTransitions 142 hasNumberOfUniqueTransitions 142  
 hasNumberOfValidIdentifications 142 hasNumberOfInvalidWater-Transitions 0  
 hasNumberOfInvalidWater-C2V-Transitions 0 hasNumberOfInvalidIdentifications 0  
 hasNumberOfNonuniqueTransitions 0

## V4\_T5\_279\_PressureValue\_MD\_for\_H2O

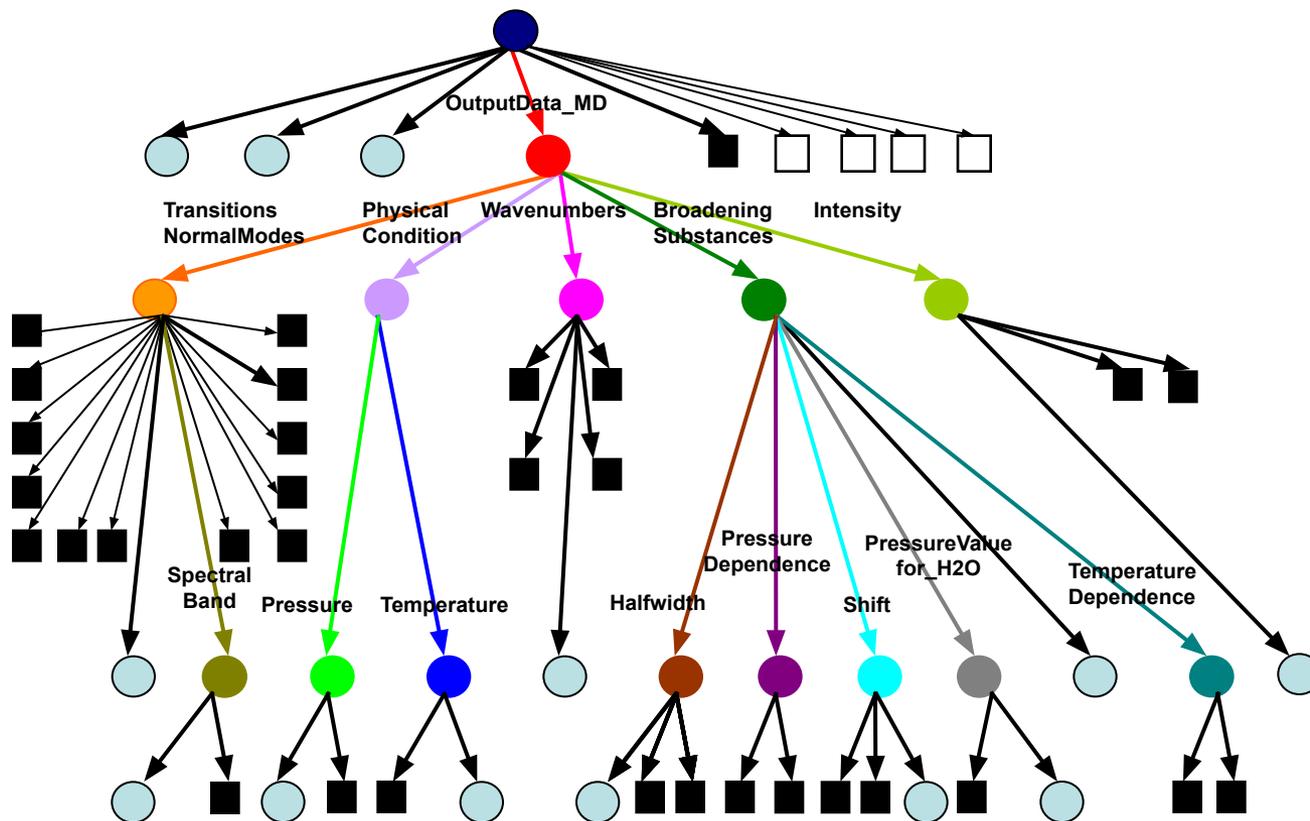
hasUnit atm  
 hasFloatValue 1

## V4\_T5\_279\_for\_NormalModes\_v1UP\_v2UP\_v3UP\_v1LOW\_v2LOW\_v3LOW\_SpectralBand

hasBandType  
 v1UP\_v2UP\_v3UP\_v1LOW\_v2LOW\_v3LOW  
 hasNumberOfSpectralBands 1

# Субъектно-предикатная структура «information source 1998\_ToBr\_H2\_170-H2O»

1998\_ToBr\_H2\_170-H2O



61 высказывания



# Количественная предметная область



$N_T$  – число высказываний в Tbox

$N_A$  – число высказываний в Abox (M - факты предметной области)

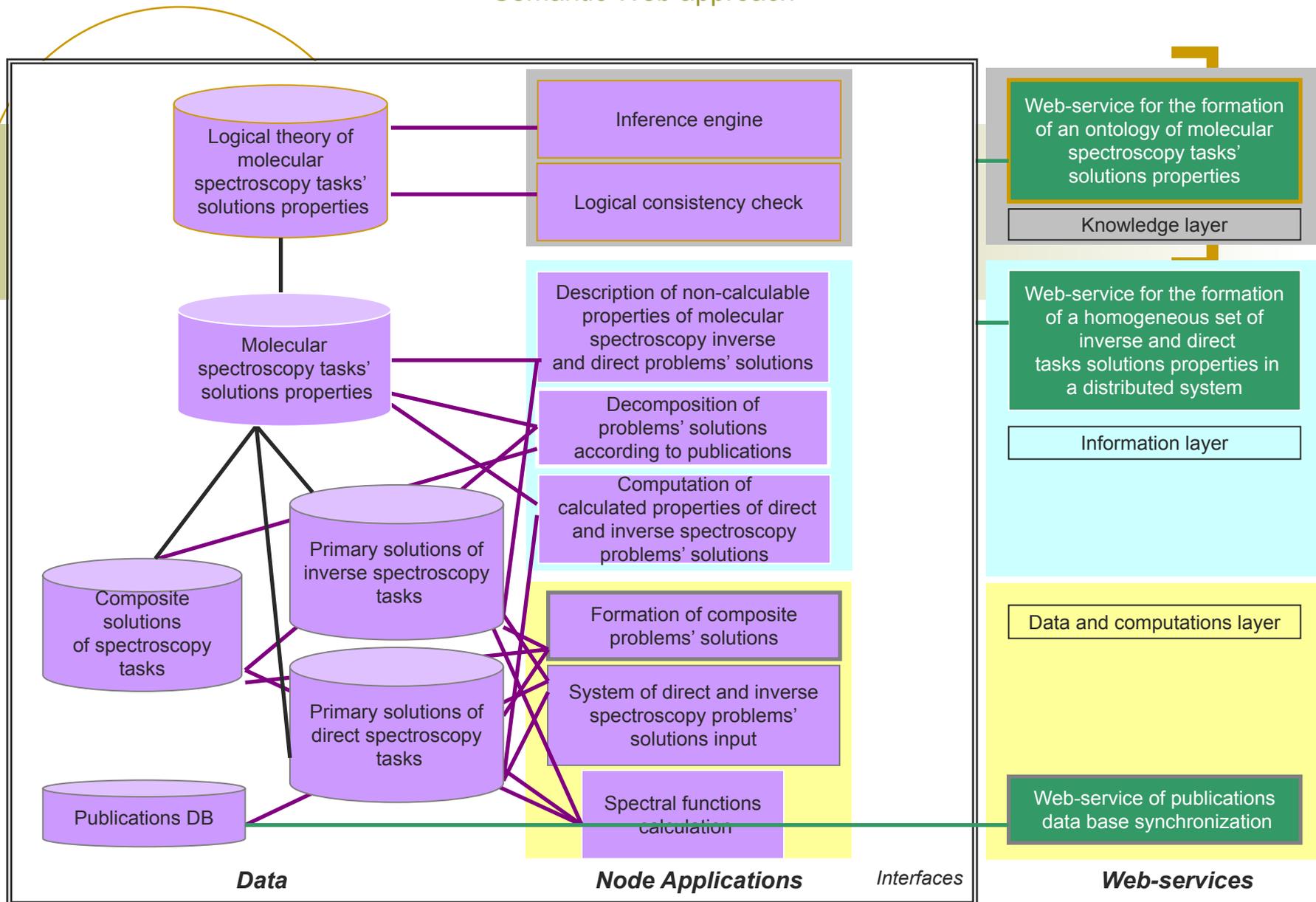
$N_{A(P)}$  – число высказываний в Abox (P - факты предметной области,  $P \subseteq M$ )

$$N_T \ll N_{A(P)} \ll N_A$$



# Архитектура распределенной информационной системы для молекулярной спектроскопии

Semantic Web approach



# Иерархия математических теорий



FIGURE 2.14 Обобщение иерархии теорий

Sowa J.F. Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

594 p.



# Отношения транзитивности и рефлексивности в количественной спектроскопии

- Транзитивность

$$a=b \ \& \ b=c \Rightarrow a=c$$

- Рефлексивность

$$a = a$$

- Симметрия

$$a=b \Rightarrow b=a$$

- Антисимметрия

$$a = < b \ \& \ b = < a \Rightarrow a = b$$



# Примеры. Ограничения на значения и существование

## Ограничения в задачах достоверности данных в молекулярной спектроскопии

### Формальные ограничения

#### Тип данных –

квантовые числа – натуральные числа,  
интенсивность, столкновительная полуширина, вакуумные волновые числа, уровни энергии – положительные действительные числа, ....

#### Интервалы изменения –

$0 < \text{волновые числа} < 45000 \text{ см}^{-1}$ ,  $10^{-17} \text{ см/mol} < \text{интенсивность} < 10^{-30} \text{ см/mol}$

#### Правила отбора -

нормальные моды -  $k_a + k_c = J$  or  $J+1$ , .....

точные квантовые числа –  $J < 60$ ,  $0 < s < 5$ , .....

### Ограничения опубликования

Решение задачи должно быть опубликованным информационным ресурсом

### Неформальные ограничения

#### Точки зрения экспертов



# Примеры. Сводка публикаций. Ограничения на значения

H <sub>2</sub> O		H <sub>2</sub> S											
Вещь	Вещь	T1		T2		T3		T5		T6		T7	
		HN	CN	HN	CN	HN	CN	HN	CN	HN	CN	HN	CN
H <sub>2</sub> O	CO <sub>2</sub>	-	-	19 (13)	4 (3)	14 (12)	1 (0)	52 (48)	-	47 (41)	-	-	-
H <sup>18</sup> OH	H <sub>2</sub> S	-	-	12 (6)	4 (3)	6 (6)	1 (1)	22 (21)	-	26 (23)	-	-	-
H <sup>17</sup> OH	HDS	-	-	4 (4)	2 (2)	2 (2)	-	11 (11)	-	14 (14)	-	-	-
HOD	D <sub>2</sub> S	-	-	12 (11)	4 (4)	5 (5)	1 (1)	14 (13)	-	21 (21)	-	-	-
D <sub>2</sub> O	H <sup>34</sup> S	-	-	10 (9)	4 (4)	5 (5)	1 (1)	11 (10)	-	16 (15)	-	-	-
H <sup>18</sup> OD	H <sup>33</sup> S	-	-	3 (3)	2 (2)	-	-	6 (6)	-	8 (8)	-	-	-
H <sup>17</sup> OD	HD <sup>34</sup>	-	-	1 (1)	-	1 (1)	-	5 (5)	-	5 (5)	-	-	-
D <sub>2</sub> <sup>17</sup> O	Итого	-	-	6 (4)	2 (2)	1 (1)	-	8 (7)	-	8 (7)	-	-	-
D <sub>2</sub> <sup>18</sup> O	Триады	-	-	1 (0)	-	-	-	-	-	1 (0)	-	-	-
HTO	A-Box	-	-	1 (1)	-	-	-	-	-	3 (3)	-	-	-
Итого	A-Box	-	-	4 (3)	-	-	-	3 (3)	-	7 (6)	-	-	-
	A-Box (корр.)	-	-	2 (1)	-	-	-	-	-	3 (3)	-	-	-
		-	-	-	-	1 (1)	-	1 (1)	-	1 (1)	-	-	-
		-	-	-	-	-	-	-	-	1 (1)	-	-	-
	Итого	-	-	75 (56)	22 (20)	35 (33)	4 (3)	133 (125)	-	161 (148)	-	-	-

Триады	T1	T7	T2	T6	T3	T5	Итого
A-Box	-	-	4064	8463	2461	10658	25646
A-Box (корр.)	-	-	425758	290139	290139	10658	2705897



# Примеры. Декомпозиция. Ограничения на существование

Сравнение массивов спектральных данных Hitran и GEISA в диапазоне 0-50 см<sup>-1</sup>

	0-10 см <sup>-1</sup>		10-30 см <sup>-1</sup>		30-50 см <sup>-1</sup>	
	HITRAN	GEISA	HITRAN	GEISA	HITRAN	GEISA
<b>H<sub>2</sub><sup>17</sup>O</b>		<u>0/1</u> 0%		<u>0/10</u> 0%	<b>9/25</b> 36%	<b>9/29</b> 36%
		<u>0/1</u> 0%		<u>0/10</u> 0%	<u>0/25</u> 0%	<u>0/29</u> 0%
<b>H<sub>2</sub><sup>18</sup>O</b>	<b>28/31</b> 90%	<u>0/1</u> 0%	<u>0/15</u> 0%	<u>0/15</u> 0%	<b>17/33</b> 52%	<b>17/33</b> 52%
	<b>8/31</b> 26%	<u>0/1</u> 0%	<u>0/15</u> 0%	<u>0/15</u> 0%	<u>0/33</u> 0%	<u>0/33</u> 0%
<b>HDO</b>	<b>20/65</b> 31%	<b>20/66</b> 30%	<b>39/76</b> 51%	<b>20/76</b> 26%	<b>60/76</b> 79%	<b>60/76</b> 79%
	<u>0/65</u> 0%	<u>0/66</u> 0%	<u>0/76</u> 0%	<u>0/76</u> 0%	<u>2/76</u> 3%	<u>2/76</u> 3%
<b>HD<sup>18</sup>O</b>	<b>26/53</b> 49%		<b>52/61</b> 85%		<b>28/45</b> 62%	
	<b>8/53</b> 15%		<b>3/61</b> 5%		<b>0/45</b> 0%	

Точность, с которой проводилось сравнение данных измерений задана шрифтом.

Подчеркнутые цифры означают, что сравнение проведено с точностью 0.01 см<sup>-1</sup>, цифры курсивом - с точностью *0.1* см<sup>-1</sup>, и выделенные жирным шрифтом – с точностью **1** см<sup>-1</sup>.



# Декомпозиция источников данных

Сравнение массивов спектральных данных Hitran и GEISA в диапазоне 50-20000 см<sup>-1</sup>

	50-200 см <sup>-1</sup>		200-10000 см <sup>-1</sup>		10000-20000 см <sup>-1</sup>	
	HITRAN	GEISA	HITRAN	GEISA	HITRAN	GEISA
<b>H<sub>2</sub><sup>17</sup>O</b>	<b>147/279</b> 53%	<b>147/279</b> 53%	<b>1309/4762</b> 28%	<b>1383/4828</b> 29%	<b>118/1916</b> 6%	<b>9/325</b> 3%
	<i>0/279</i> 0%	<i>0/279</i> 0%	<i>669/4762</i> 14%	<i>119/4828</i> 2%	<i>639/1916</i> 33%	<i>107/325</i> 33%
<b>H<sub>2</sub><sup>18</sup>O</b>	<b>205/342</b> 60%	<b>205/342</b> 60%	<b>1529/6889</b> 22%	<b>1380/7109</b> 20%	<b>164/2443</b> 7%	<b>23/863</b> 3%
	<b>2/342</b> 1%	<i>0/342</i> 1%	<b>302/6889</b> 4%	<b>317/7109</b> 4%	<b>208/2443</b> 9%	<b>49/863</b> 6%
<b>HDO</b>	<b>370/599</b> 62%	<b>370/599</b> 62%	<b>1674/8825</b> 19%	<b>2980/11045</b> 27%	<b>8/3483</b> 0%	<b>8/119</b> 7%
	<b>2/599</b> 0%	<b>0/599</b> 0%	<b>5/8825</b> 0%	<b>27/11045</b> 0%	<b>94/3483</b> 3%	<b>0/119</b> 0%
<b>HD<sup>17</sup>O</b>			7/175 4%	7/175 4%		
			0/175 0%	0/175 0%		
<b>HD<sup>18</sup>O</b>	<b>378/544</b> 69%		<b>434/908</b> 48%	<b>195/679</b> 29%		
	<b>16/544</b> 3%		<b>221/908</b> 24%	<b>221/679</b> 33%		

Точность, с которой проводилось сравнение данных измерений задана шрифтом.

Подчеркнутые цифры означают, что сравнение проведено с точностью 0.01 см<sup>-1</sup>, цифры курсивом - с точностью *0.1* см<sup>-1</sup>, и выделенные жирным шрифтом – с точностью **1** см<sup>-1</sup>.



# Вопросы

