



Collaboratory for Multi-scale Chemical Science (CMCS)



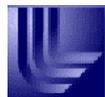
CMCS Task Group: Data Flow Between Scales

Branko Ruscic & Theresa Windus

ruscic@anl.gov & theresa.windus@pnl.gov

ANL & PNNL

CMCS Workshop II – May 11-12, 2002





Introduction

- **Purpose:** Explore and develop applicable modes of data flow between applications and/or participants at various chemical scales.
- **Members:** Theresa Windus, Branko Ruscic, Elena Mendoza, Reinhardt Pinzon, Eric Stephan, Karen Schuchardt, John Hewson, Sandra Bittner, Gregor van Laszevski
- **Background info:** The essence of multi-scale science is the transfer of needed information from one scale to the other. For example, species-specific molecular (spectroscopic) data and species-interconnecting data (e.g. kinetic equilibria, bond dissociation energies, etc) feeds directly into the Active Tables (as well as others), which in turn produces thermochemical data that will be used by other applications (e.g. by programs that fit this data and supply polynomials to modeling suites). Therefore, the Active Tables represent both a consumer and producer of data. This consumer/producer model continues up all the chemical scales, up to the numerical simulation stage (for the current scope of this project).



Systematization

- **Some examples of systematization:**

- **Generator & recipient of data**

[both can be a person (P) or an application (A)]

- In some case this can be simplified by selecting a common exchange medium (M) (e.g. DAV)

- **Manual vs. automatic** (both are needed in CMCS)

- **Synchronous vs. asynchronous**

(async. is applicable to P → and → P xfers, as well as some A → A xfers, while other A → A xfers will be synchronous)

- **Solicited vs. unsolicited**

(e.g. solicited may require search capabilities, unsolicited will require subscription to notification)

- **Exchange medium**

(through DAV, through portal, through local file system, direct)

- **Index of indirection**

(depth of tree, when one xfer causes other dependent data xfers, etc.)

- **Initial focus (after discussion with Applications Workgroup):**

development of API calls used by A → xfers to DAV,
putting data used by Active Tables into xml format for use by others,
posting output data, notifying & launching dependent applications

P → P	P → A
A → P	A → A

P → M	M → P
A → M	M → A



Active Thermochemical Tables

- **Active Tables** (“intelligent” databases centered around data dependencies and statistical self-consistency)
- **Active Thermochemical Tables (ATcT) are the first example of the Active Tables category and are a *new paradigm* to thermochemistry**
 - ATcT addresses and solves some important problems associated with the traditional sequential approach to evolving thermochemical values, such as:
 - cumulative errors (lack of feedback to values obtained in previous steps)
 - uncertainties that do not properly reflect the globality of data available
 - hidden dependencies that prevent proper updates of “static” tables because
 - Some advantages of Active Thermochemical Tables:
 - output of the ATcT is superior to conventional tables because its values are globally consistent with all input data and error bars properly reflect all relationships presented as input
 - allows painless propagation of new data and its consequences
 - opens a new venue of rapid sharing of latest information (including tentative information)
 - allows “what if” tests of new data for consistency with existing knowledge
 - allows educational explorations, including Gedanken experiments
 - can suggest new high-impact experiments/calculations
 - pedigree documents are easily incorporated for examination

(see separate PowerPoint presentation for more information on Active Tables)



Development

- **Active Tables development: datastores**

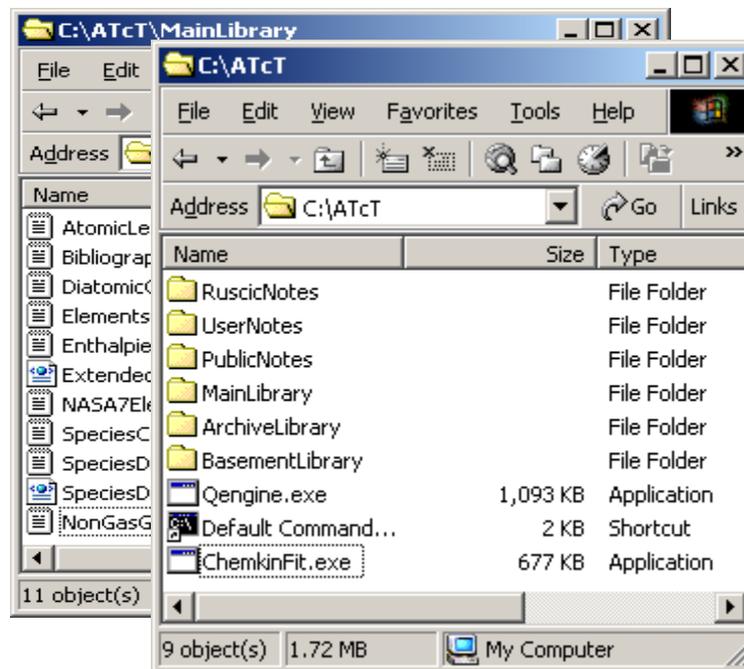
- **Blueprinting ATcT input datastores**

- Hierarchical organization of information into **Libraries**:

- UserNotes
 - PublicNotes
 - MainLibrary
 - ArchiveLibrary
 - BasementLibrary

- Hierarchical organization of information **within a Library**:

- General information:
 - SpeciesDictionary, SpeciesCookbook, ElementsCookbook, ...
 - Bibliography
 - Thermochemical data:
 - ReactionDictionary, NetworkEncyclopedia, ...
 - AtomicLexicon, DiatomicCompositeLexicon, DiatomicCountLexicon, DiatomicHerzbergLexicon, DiatomicRRHOLexicon, PolyatomicCompositeLexicon, PolyatomicCountLexicon, PolyatomicHerzbergLexicon, PolyatomicRRHOLexicon, ...





Development

- **Active Tables development: datastore sharing**
 - Collaborative sharing of datastores used by ATcT as input
 - Export of data in xml format
- example: SpeciesDictionary.txt ⇒ SpeciesDictionary.xml (DAV)
⇒ ExtendedSpeciesDictionary.xml (DAV)

```
SpeciesDictionary.txt - Not...
File Edit Format Help
10035-10-6
HBr
Hydrogen bromide
Hydrobromic acid
Bromohydric acid
Bromohydrogen
.
10034-85-2
HI
Hydrogen iodide
Hydriodic acid
Hydroiodic acid
Iodohydric acid
Iodohydrogen
.
7732-18-5
H2O
Water
HOH
Dihydrogen oxide
Dihydrogen monoxide
Hydrogen oxide
.
```



```
C:\Documents and Settings\Ruscic.CHM_CSG\My Documents\Projects\FTN projects\ATcT\Qengine\Debug\ - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address C:\Documents and Settings\Ruscic.CHM_CSG\My Documents\Projects\FTN projects\ATcT\Qengine\Debug\MainLibrary\ExtendedSpeciesDictionary.xml
- <species>
- <identifiers>
  <UniqueID type="CAS">7732-18-5</UniqueID>
  <ChemicalFormula type="AlphaSortedGross">H2O</ChemicalFormula>
  <ChemicalFormula type="ZSortedGross">H2O</ChemicalFormula>
  <ChemicalFormula type="PrincipalIsotopomer" reference="ATcT_1.0_Alpha:Internal:3">1H2 16O</ChemicalFormula>
- <names>
  <Name type="Preferred">Water</Name>
  <Name type="Alternate">HOH</Name>
  <Name type="Alternate">Dihydrogen oxide</Name>
  <Name type="Alternate">Dihydrogen monoxide</Name>
  <Name type="Alternate">Hydrogen oxide</Name>
</names>
</identifiers>
- <composition>
  <Charge>0</Charge>
  <TotalNumberOfAtoms>3</TotalNumberOfAtoms>
- <element>
  <Symbol type="Chemical">H</Symbol>
  <ElementName type="Chemical">Hydrogen</ElementName>
  <Name type="Isotopic">Hydrogen-natural</Name>
  <AtomicNumber>1</AtomicNumber>
  <MassNumber>NaturalIsotopicMixture</MassNumber>
  <NumberOfAtoms>2</NumberOfAtoms>
</element>
- <element>
  <Symbol type="Chemical">O</Symbol>
  <ElementName type="Chemical">Oxygen</ElementName>
  <Name type="Isotopic">Oxygen-natural</Name>
  <AtomicNumber>8</AtomicNumber>
  <MassNumber>NaturalIsotopicMixture</MassNumber>
  <NumberOfAtoms>1</NumberOfAtoms>
</element>
</composition>
<MolecularMass uncertainty="0.00033" reference="ATcT_1.0_Alpha:Internal:3">18.01528</MolecularMass>
</species>
- <species>
```



Development

- **Active Tables development: ATcT Kernel**
 - ATcT Kernel has a number of Engines (Q engine, Search engine, Network engine, Output engine, etc)
 - ~ **15 000 lines of Fortran 95 code** so far (~12 500 lines of permanent code & ~ 2 500 lines of testing code)
 - **Q engine** is currently nearing the end of phase I (the goal of phase I is to implement all strategies to calculating Q function used in JANAF and Gurvich et al.)
 - Implemented various functions needed to perform:
 - direct (exact) count for gas phase atoms
 - RRHO for all types of species
 - NRRAO corrections for diatomics
 - recovery of thermochemistry from empirical expressions for condensed and gas phases (Gurvich, NASA7)
 - Currently implementing:
 - direct (exact) count for gas phase diatomics and polyatomics
 - computed direct count for diatomics
 - NRRAO corrections for polyatomics
 - recovery of thermochemistry from empirical expressions for condensed and gas phases (Burcat, Chemkin)



Development

- **Active Tables development: ATcT Kernel (cont.)**
 - **Search engine** v 1.0 is in place
 - Strategy of “increasing despair” (hierarchical/non-hierarchical & desperate/non-desperate approach) searches along the sequence:
 - CAS #
 - existing name or “chemist choice” formula (restrictive, “as is”)
 - existing name or “chemist choice” formula (case insensitive)
 - formula (“as is”, specific isotopes, non-isotopic, relaxed)
 - case insensitive substring search

```
Default Command Prompt - qengine
Search string? c
3 hits
hit # 1
7440-44-0
C
Carbon atom
hit # 2
7782-42-5
C
Graphite
hit # 3
7782-40-3
C
Diamond
Must narrow search
```



Development

- Active Tables development: **ATcT Kernel (cont.)**
 - **Output engine**
 - user-determined or convenient fixed temperature schedules (JANAF/Gurvich/big Gurvich/ATcT/big ATcT)
 - user choice of various units (J/cal/R)
 - general format, or application specific format (e.g. ready for Chemkin III FitDat application)
 - will implement flexible format (user choice of quantities) and xml-annotated format

```
input.dat - Notepad
File Edit Format Help
Ar
AR 1
**
1000.000
0.000
0.000
100.000 4.968 31.582 -5.701
200.000 4.968 35.025 -5.204
250.000 4.968 36.134 -4.955
298.150 4.968 37.009 -4.716
300.000 4.968 37.040 -4.707
350.000 4.968 37.806 -4.459
400.000 4.968 38.469 -4.210
450.000 4.968 39.054 -3.962
500.000 4.968 39.578 -3.713
600.000 4.968 40.483 -3.217
700.000 4.968 41.249 -2.720
800.000 4.968 41.913 -2.223
```

```
Default Command Prompt - qengine
*****
* ARGONNE NATIONAL LABORATORY *
*****

Ar demo

T Cp S H(T)-H(ref)
K J/mol/K J/mol/K kJ/mol
0.00 0.000 0.000 -6.197
100.00 20.786 132.138 -4.119
200.00 20.786 146.546 -2.040
250.00 20.786 151.185 -1.001
298.15 20.786 154.846 0.000
300.00 20.786 154.974 0.038
350.00 20.786 158.179 1.078
400.00 20.786 160.954 2.117
450.00 20.786 163.403 3.156
500.00 20.786 165.593 4.196
600.00 20.786 169.382 6.274
700.00 20.786 172.587 8.353
800.00 20.786 175.362 10.432
900.00 20.786 177.810 12.510
1000.00 20.786 180.000 14.589
1100.00 20.786 181.982 16.667
1200.00 20.786 183.790 18.746
```



Development

- **Active Tables development: Network**

- **Test network:** assembled principal data needed to define the thermochemical network that includes GRI-Mech species (Excel file)
- Assembled small network pertinent to refinement of $\Delta H_f^\circ(\text{OH})$ (Excel) (combined MICS and Chemical Sciences effort)
- For testing purposes developed Excel add-in that implements a simplified network solution strategy (applied to $\Delta H_f^\circ(\text{OH})$ refinement problem)

Microsoft Excel - species1.xls

Species	CAS#	Name	R	E	A	C	T	I	O	N
O	17778-80-2	Oxygen, atomic	1						2	
O2	7782-44-7	Oxygen								
O3	10028-15-6	Ozone	1.5						1	0.3
H	12385-13-6	Hydrogen atom	1						2	
H2	1333-74-0	Hydrogen								
OH	3352-57-6	Hydroxyl	1						1	0
H2O	3170-83-0	Hydroperoxy	1						2	0
H2O	7732-18-5	Water	1						1	1
H2O	7732-18-5	Water	1						1	1
H2O	7732-18-5	Water	1						1	1
H2O	7732-18-5	Water	1						1	1
H2O	7732-18-5	Water	1						1	1
H2O	7732-18-5	Water	1						1	1
H2O2	7722-84-1	Hydrogen peroxide	1						1	2
H2O2	7722-84-1	Hydrogen peroxide	1						1	2
H2O2	7722-84-1	Hydrogen peroxide	1						1	2
H2O2	7722-84-1	Hydrogen peroxide	1						1	2
H2O2	7722-84-1	Hydrogen peroxide	1						1	2
N	17778-88-0	Nitrogen atom	1						2	
N2	7727-37-9	Nitrogen								
NO	10102-43-9	Nitrogen oxide	1						1	1
NO	10102-43.9	Nitrogen oxide	1						1	1
NO	10102-44.0	Nitrogen oxide	1						1	1
NO2	10102-44.0	Nitrogen dioxide	1						1	2
NO2	10102-44.0	Nitrogen dioxide	1						1	2

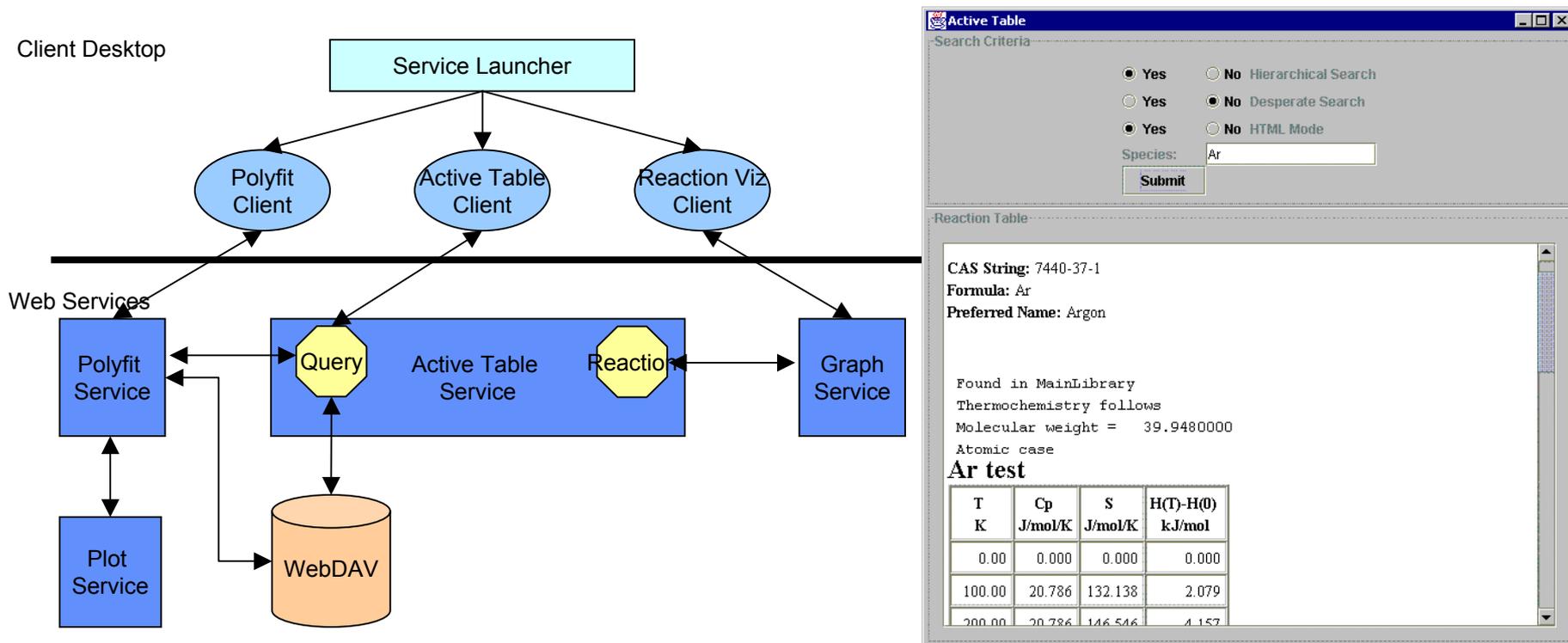
Microsoft Excel - H2O2_vanilaplay3.xls

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Constants																						
eV2cm 8065.545																						
eV2kJmo 96.48534																						
cal2J 4.184																						
22.5 cm-1																						
original data																						
41260 ± 15 cm-1 ΔH 0																						
36418.3 ± 1 cm-1 ΔH 0																						
-285.83 ± 0.040 kJ/mol ΔH 298																						
-285.195 ± 0.040 kJ/mol ΔH 298																						
-285.85 ± 0.33 kJ/mol ΔH 298																						
-44.004 ± 0.002 kJ/mol ΔH 298																						
-44.016 ± 0.010 kJ/mol ΔH 298																						
-2.903 ± 0.001 kJ/mol ΔH x																						
35420 ± 15 cm-1 ΔH 0																						
35450 ± 100 cm-1 ΔH 0																						
18.11778 ± 0.008 eV ΔH 0																						
18.11889 ± 0.0035 eV ΔH 0																						
18.11929 ± 0.002 eV ΔH 0																						
18.11779 ± 0.005 eV ΔH 0																						
19.4989 ± 2 cm-1 ΔH 0																						
13.01 ± 0.01 eV ΔH 0																						
13.0 ± 0.1 eV ΔH 0																						
31151 ± 5 cm-1 ΔH 0																						
37.082 ± 0.67 kJ/mol ΔH 0																						
-98.05 ± 0.07 kJ/mol ΔH 298																						
47.95 ± 4.4 kJ/mol ΔH 298																						
51.92 ± 0.15 kJ/mol ΔH 298																						
47.51 ± 3.1 kJ/mol ΔH 298																						
51.75 ± 0.16 kJ/mol ΔH 298																						
52.2 ± 0.16 kJ/mol ΔH 298																						
51.925 ± 0.073 kJ/mol ΔH 298																						
-5.59 ± 0.001 kJ/mol ΔH x																						
17051.8 ± 3.4 cm-1 ΔH 0																						



Development

- **Active Tables development: GUI & ATcT Portal (early stages)**
 - **JAVA Application** (can be modified into JetSpeed portlet or HTML)
 - creates a portal environment to ATcT
 - launches ATcT, passes input, displays output
 - launches dependent applications (e.g. Chemkin III FitDat)
 - graphically displays data (network, fit results)





Development

- **Active Tables development: Network Visualization (early stages)**
 - Graphical network visualization is an important component of the ATcT user interface that can be used during network analysis
 - the graph visualization component is developed uses Graph Viz Engine
 - nodes/edges clickable for more information
 - launched through the JAVA GUI/Portal
 - has the potential to display huge networks
 - will work in a collaborative environment between remote researches

The screenshot displays the 'Active Table Reaction Viewer' software interface. The main window shows a complex network graph with nodes representing chemical species and edges representing reactions. Nodes are labeled with numbers (1-26) and chemical formulas (H₂, O₂, H₂O, OH, H, O). Edges are labeled with numerical values (0.5, 1.0, 2.0). A 'CMCS Demo' window is open, showing buttons for 'Active Table', 'Reaction Parser', and 'Polynomial Fit'. The 'Reaction Parser' window is also open, displaying the following information:

Reaction Parser

Graph Generation

`java/cmc/srj/javalexamples/newgraph.in`

Modification

Reaction	Enthalpy	Uncertainty
<chem>H2O(l,298) <=> H2O(g,298)</chem>	-285.850	0.330

Meta Data

Water from liquid to gas state



Acknowledgement

- Supported by the

**Division of
Mathematical, Information and Computational Sciences,
Office of Advanced Scientific Computing Research**

and

**Division of
Chemical Sciences, Geosciences and Biosciences,
Office of Basic Energy Sciences,**

U.S. Department of Energy