



# Семантичний Грід в молекулярних дослідженнях

Виконав студент групи ДА-52м: Османова Т.М.

Керівник проекту д.т.н., проф.: Петренко А.І.

Київ 2011

# Актуальність досліджень:

Грид- технології є новітнім напрямком комп'ютерних технологій, що набуває все більшого розвитку і застосування в молекулярних дослідженнях. На сьогоднішній момент обсяг інформації, що використовується в молекулярних дослідженнях зростає з величезною швидкістю, тому виникає велика потреба в інструментах для підтримки обміну знаннями, ресурсами, результатами та спостереженнями.



Перше, що вони роблять, це пошук он-лайн

# Проблематика досліджень:

- Віддалене співробітництво не поширене серед дослідників;
- Обмежене використання еТехнологій;
- OpenSource та OpenData відомі далеко не всім провідним дослідниками;
- Проблема інтелектуальної власності;
- Відсутня певна стандартизація щодо зберігання даних;
- Відсутній єдиний стандарт онтологій;

На даному етапі розвитку ці проблеми вирішувалися за допомогою:

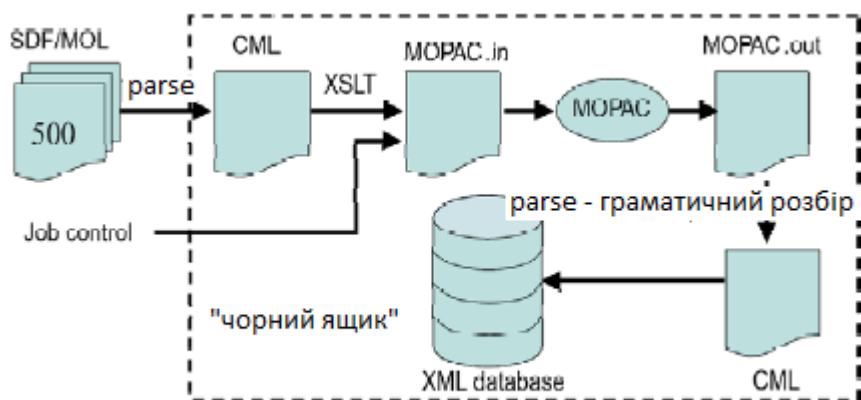
- Protein Data Bank;
- PubChem(NIH);
- NCI database;
- Molecular structures(EBI).

# Вирішення даних проблем:

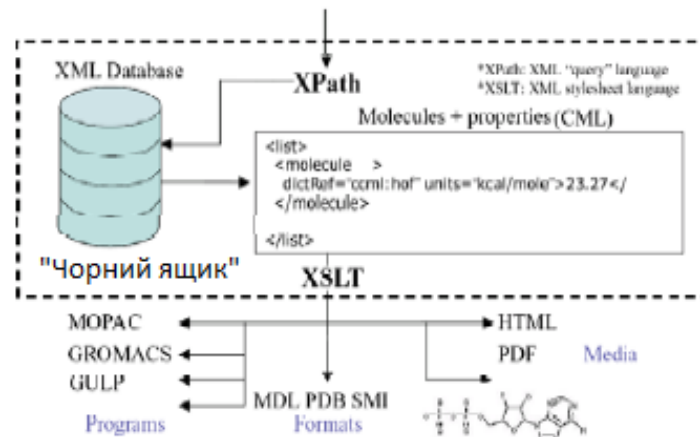
Впровадження технології семантичного Грід(СГ) в молекулярні дослідження:

1. СГ буде мати можливість знаходити смислову інформацію у запитах, наприклад:
  - Проаналізуй молекули з певною міжатомною енергією;
  - Підбери мені декілька молекул з подібними властивостями і певною структурою;
2. СГ визначає які ІТ – ресурси необхідні для розв`язання даної задачі, які ІТ – ресурси вже існують, які зовнішні ІТ – ресурси можуть бути використані, а які необхідно створити;
3. Дає можливість дослідниками працювати в кооперації, тим самим прискорюючи відкриття та ін.

# Molecular GRID (World Wide Molecular Matrix)



Знайти всі молекули з атомами більше 10 і  $\Delta H_f^?$



- Користувач завантажує дані у CML. За допомогою пакету MOPAC обчислюються властивості молекул і результат повертається до користувача. Паралельно результати публікуються на сайті, де їх можуть повторно використовувати.
- XPath запити направляються в базу даних XML, де вибирається навігація за потрібними вузлами, результати надаються користувачу.

# Molecular GRID (World Wide Molecular Matrix)

«Суміш розчиняється при 119 – 121 °C»

Інструменти які

використовуються у WMMM:

- Онтологія в даному проекті базується на XMLSchema, оскільки потрібне строкове введення даних , агрегація , програмована семантика та використання запитів XPath. Словники створюються кожним користувачем окремо;

- На 20 машинах встановлений Condor(77MHz), що дало змогу обробити вже більш як 250 000 молекул;
- Проміжне програмне забезпечення Globus;
- database (Xindice)
- Grid додаток MORAC

**cml namespace**

```
<entry id="mpt" name="Melting Point">  
<appinfo>  
<scalar dataType="xsd:float"  
unitType="unit:temp"/>  
</appinfo>  
<definition>The temperature...  
... a solid melts</definition>  
</entry>
```

**unit namespace**

```
<unit id="celsius"  
name="Celsius"  
parentSI="unit:k"  
unitType="temp"  
constantToSI="273.15"/>
```

```
<scalar dictRef="cml:mpt" units="unit:celsius"  
minValue="119" maxValue="121">120</scalar>
```

# The CombeChem

Проект CombeChem представлення семантичного Грід в еНауках.

CombeChem + онтології =

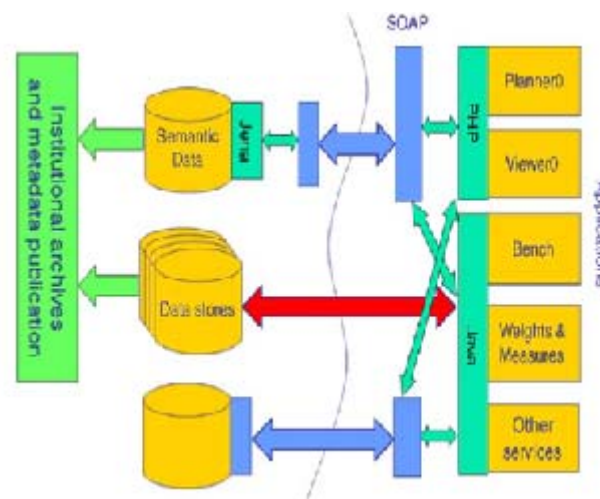
Інструменти які використовуються у проекті CombeChem:

Онтологія в даному описується мовою

OWL, а відношення між

метаданими(знаннями) описується у RDF;

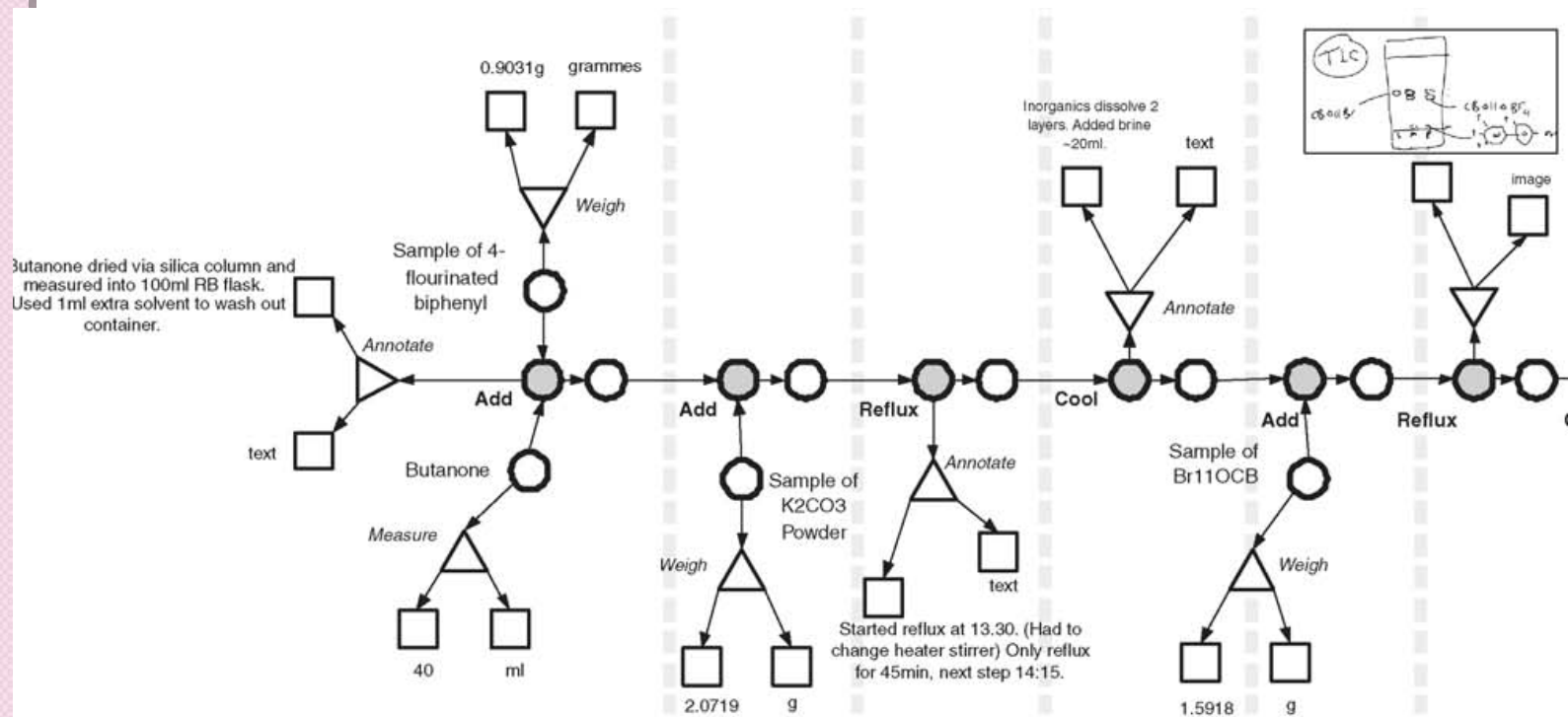
Database 3Store



SmartLab архітектура

# The CombeChem

## Фрагмент онтології





# Collaboratory for Multi-scale Chemical Science (CMCS)

Проект CMCS був створений американськими вченими для використання здобутків інформаційних технологій у хімічних проблемах горіння.

Для опису даних і метаданих використовується Dublin Core, дозволяє створювати словник як у XML так і у RDF описуючи такі властивості: **title**, **creator**, **dates**, **publisher**, **is-referenced-by**, **references**, **replaces**, **is-replaced-by**, **has-version**, та інші.

Проміжне програмне забезпечення SAM(Scientific Annotation *Middleware*)

Grid додаток : NWChem

# Висновки та рекомендації

1. Якщо розглядати моделі з точки зору сумісності, то кожна створювалася для конкретних задач.
2. Всі моделі мають у наявності потужні Грід додатки і направлені на вирішення задач пошуку, зберігання та обробки молекулярних даних.
3. Проект CombeChem та CMCS для опису метаданих використовують OWL та Dublin Core, що базується на RDF, це дозволяє використовувати словники в будь-якій прикладній області, що не можна сказати про XMLSchema.
4. Якщо порівнювати зберігання та доступ до даних то порівняно с реляційними базами даних спостерігається велика перевага у гнучкості, зміні словника.

# Висновки та рекомендації

В роботі проведений аналіз наукових досліджень в Україні так виявлено, що велика чисельність наукових інститутів використовує молекулярні дослідження в Україні.

Є два способи використання СГ в Україні:

- Створити власну СГ інфраструктуру;
- Використовувати вже існуючу.



Дякую за увагу!