

# Інтеграція даних в семантичному Грід

Лунченко Олена, ДА-71

# Mapping language

- Мова відображення (mapping language) - коментування XML структур даних за допомогою OWL понять і властивостей
- Ядро мови відображення (Mapping Language Engine) - реалізації мови відображення
- Компонент динамічного виклику веб-сервісу (Dynamic Web Service Invocation) - виконання веб-сервісів.

Надається можливість нарощувати існуючі системи веб-сервісів проміжним шаром для автоматичного виконання узгодження даних у випадку виникнення синтаксичної невідповідності.

# Mapping language.

## Переваги підходу

При підключенні двох компонентів, що містять однакові дані, але в різних форматах, разом, повинна бути надана додаткова інформація, щоб зазначити як дані було вибрано з однієї структури даних і передано у наступну.

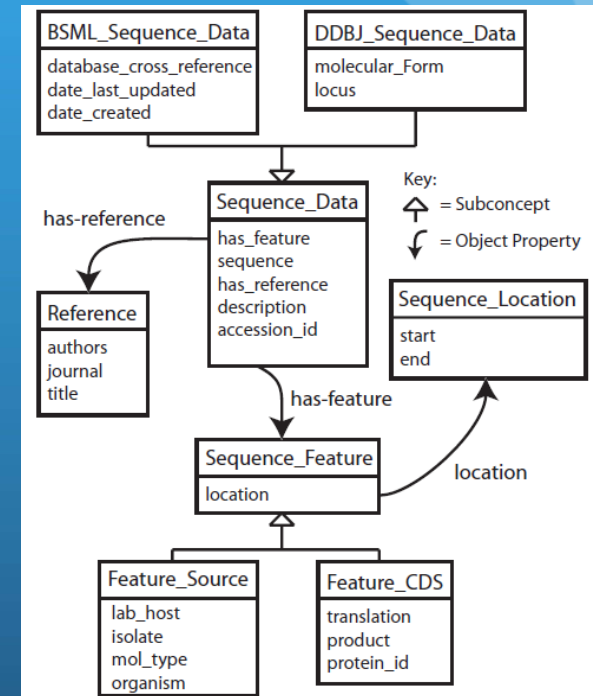
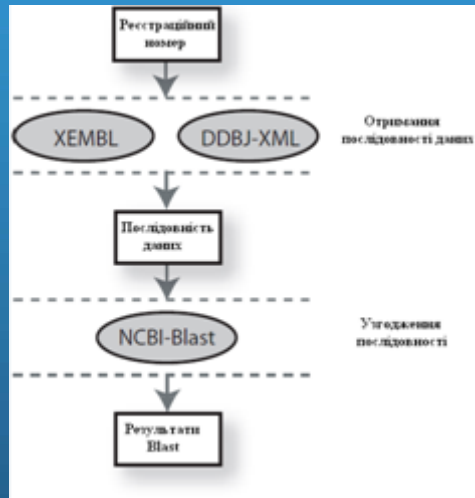
### Варіанти:

- *мови перетворення даних (XSLT, XQUERY)*
  - ручна специфікація всіх можливих перетворень. Для  $n$  сумісних форматів даних для максимальної сумісності необхідно  $(n - 1)n$  перетворень;
  - коли вводиться новий тип даних, повинні бути вказані відображення до і від усіх інших сумісних типів
- *використовуючи OWL поняття і властивості, щоб коментувати XML структури даних*
  - не вимагається відображень між усіма сумісними форматами, кожен формат даних потребує тільки одне відображення до онтологічної специфікації. Число відображень, необхідних для кожного сумісного формату даних, має складність  $O(n)$ ;
  - коли нові формати додаються до існуючої системи, тільки одне відображення необхідне для досягнення максимальної сумісності.

# Mapping language.

## Використання в біоінформатиці

Приклад використання в проєкті MyGrid: вирішення поширеної задачі біоінформатики: отримати послідовність даних з бази даних і передати її на інструмент узгодження для перевірки на подібність з іншими відомими послідовностями



Онтологія для опису послідовності даних

# SAS Data Integration Studio

У Data Integration Studio можна створювати перетворення, які викликають SAS Text Miner для обробки даних на веб-сторінках або документів різних форматів

- дані розбираються на два набори даних: список знайдених ключових слів і необроблені дані;
- витягуються стандартні ролі зі списку необроблених слів;
- застосовуються правила стандартизації щоб помістити слова у відомі набори;
- слова ставляться у відповідність до ролей.

Результатом є набір даних, які структуровані відповідно до існуючого сховища даних

# Процес інтеграції векторних даних у проекті SDI-Grid.

## Архітектура

У проекті SDI-Grid весь процес інтеграції даних розбивається на модулі в багаторазові, придатні до зміни, багатоцільові веб-сервіси

- Логіка програми знаходиться в невеликих модулях відповідних веб-сервісів.
- І програма, і потік даних організовуються легким ядром управління робочим процесом.
- Кожен модуль може бути замінений на інший модуль з аналогічним визначенням вихідних даних.
- Ядро управління робочим процесом саме по собі може бути замінене.

# Процес інтеграції векторних даних у проекті SDI-Grid.

## Робочий процес

### Модулі

*Попередня обробка* - максимізує схожість вхідних наборів даних, гарантує, що "набори даних мають однаковий формат даних, однакові проекції карт, однакові системи координат.

*Трансформація* предметного набору даних необхідна, якщо його елементи свідчать про систематичні розходження в позиції і / або орієнтації в порівнянні з еталонним набором даних.

Перевірка *цілісності даних* забезпечує взаємну цілісність обох наборів даних, елементи перевіряються на їх вірність, наприклад, обмеження атрибута на діапазон значень.

*Відповідність елементів* необхідна для визначення просторової відповідності елементів двох наборів даних за критеріями, такими як найближча відстань або топологічна подібність та пов'язаними з ними атрибутами.

*Переміщення* копіює елементи, які представлені тільки в предметному наборі даних, в еталонному наборі даних їх немає.

*Злиття* поєднує інформацію про елементи, що присутні і в предметному і в еталонному наборі даних.

*Гармонізація* накладає або оновлює обмеження, які постійно мають виконуватись.

# Процес інтеграції векторних даних у проєкті SDI-Grid.

Робочий процес GRID-обчислень

Щоб здійснити розроблені модулі інтеграції даних як Грід-сервіси, повинні бути виконані три вимоги:

- в середовищі, яке надається робочими вузлами, модулі повинні мати можливість виконуватись (Scientific Linux);
- для описів завдань повинні бути створені шаблони;
- сервіс має бути реалізований так, щоб мати можливість скопіювати шаблон опису завдання і замінити всі змінні відповідними значеннями вихідних файлів.



Дякую за увагу!