Dynamic Meta-Data Support for Information Integration and Sharing Across Heterogeneous Databases

By

Wiput Phijaisanit

Director

Dr. Larry Kerschberg

Committee

Dr. Alex Brodsky

Dr. David A. Schum

Dr. Xiaoyang Wang

May 22, 1997

George Mason University

Fairfax, VA.

Outline of Talk

- Problem Statement
- Dynamic Meta-Data
- Unit Value Mediator
- Mediation Data Model
- Multiple Unit Value (MUV) Concept
- The InfoFED Federated Database System
- Contributions
- Summary and Future Directions

Problem Statement

- Definition of Meta-data
  - Data about data
  - Additional information that is necessary for data to be useful in the context of an application
  - Example is a schema definition that describes database tables, that is, relation names, attribute names, data types, etc.
- Traditional data models do not have enough meta-data to assist in:
  - Integration or mediation of data from disparate databases
  - Assist users in understanding the “context” in which data is being used.

Dynamic Meta-Data

- Meta-data in a database system.
  - Meta-data for classes and associations.
    - Examples are Superclass, Subclass, Synonym, etc.
  - Meta-data for class instances.
    - Example is a schema definition, i.e., class names, attribute names, etc.
- Meta-data classification for class instances.
  - Static Meta-Data (SMD)
    - Meta-data that is common to all instances of a class.
    - Example is a traditional schema definition such as class names, relation names, attribute names.
  - ***Dynamic*** Meta-Data (DMD)
    - Meta-data that varies for each instance of the same class.
    - Examples are data unit type, data sources, etc.

Dynamic Meta-Data (Cont.)

- Example of DMD Specification
  - Unit-type DMD
    - shows unit type and unit value of the data

Unit-type *Currency* USD 1 *Time-range* Month -1

Unit type = Currency/Time-range

Unit value = US-Dollar/Month

Static Meta-Data Schema Definition	**Data** Data Element (Attribute.Value)	**Dynamic Meta-Data** Property Knowledge Package (Attribute.Property)
Attribute Salary	2500	: **Unit-type** Currency USD 1 Time-range Month -1 : **Source** URL http://gmu.edu/db.cgi : **Security** Class Unclassified :
Attribute Salary	3000	: **Unit-type** Currency JPY 1 Time-range Week -1 : **Source** URL http://gu.edu/db1.cgi : **Security** Class Classified :

Unit Value Mediator

- Mediator
  - A mediator is a software module that exploits encoded knowledge about certain sets or subsets of data to create information for a higher layer of application.

Ref: Wiederhold

- Unit Value Mediator
  - A mediator that provides unit value conversion services.

Unit Value Mediator: Architecture

Ref: Service Description Diagram, I3 Reference Architecture

Unit Value Mediator

Unit value Mediator function

f(a, u, u') = (a', u")

where

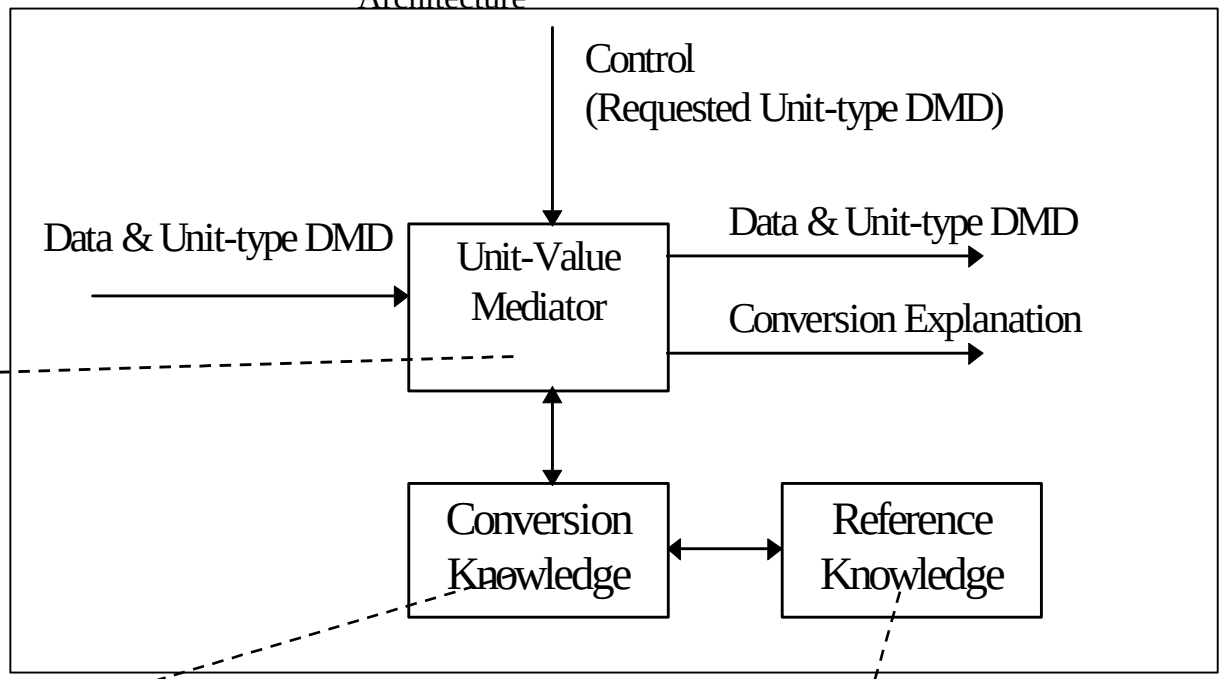
a ::= data value

u ::= data's Unit-type DMD

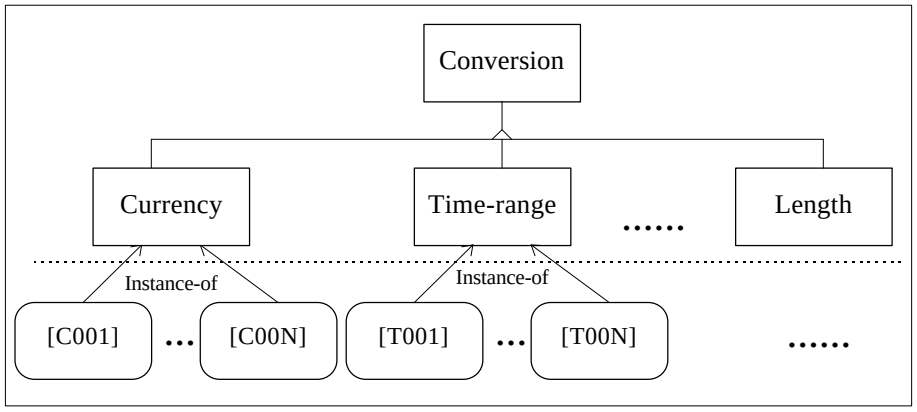
u'::= requested Unit-type DMD

a'::= converted data

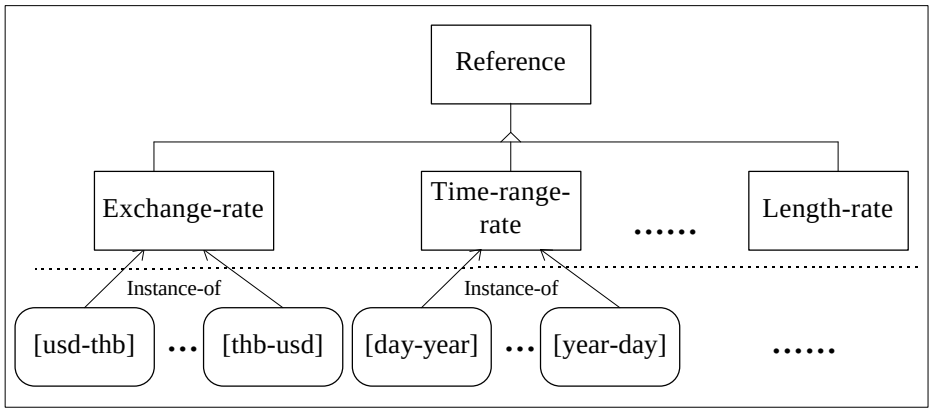
u"::=converted Unit type DMD



Conversion Knowledge



Reference Knowledge



Unit Value Mediator: Example

- Converting value 36000 from US_Dollar/year unit-value to Thai_Baht/month unit-value

a = 36000
u = ‘Unit-type Currency usd 1 Time-range year -1’
u’ = ‘Unit-type Currency thb 1 Time-range month -1’

f (a,u,u’) = (a’, u’’)
= (76110, ‘Unit-type Currency thb 1 Time-range month -1 by [c001] [usd-thb] [t001] [year-month]’)

Conversion Knowledge object ‘[c001]’

```
([c001] of Currency
  (Program.value c-convert)
  (Program.prop ": Unit-type None :
   Source URL http://isse.gmu.edu/wiput/query.cgi :")
  (Designer.value Jame Smith)
  (Designer.prop ": Unit-type None :
   Source URL http://isse.gmu.edu/wiput/query.cgi:")
  (Version.value "1/29/96")
  (Version.prop ": Unit-type None :
   Source URL http://isse.gmu.edu/wiput/query.cgi:")
)
```

Reference Knowledge object ‘[usd-thb]’

```
([usd-thb] of Exchange-rate
  (From.value usd)
  (From.prop ": Unit-type None :  Source URL gopher://una.hh.lib.umich.edu/00/ebb/monetary/noonfx.frb:")
  (To.value thb)
  (To.prop ": Unit-type None :  Source URL gopher://una.hh.lib.umich.edu/00/ebb/monetary/noonfx.frb:")
  (Rate.value 25.370000)
  (Rate.prop ": Unit-type None : Source URL gopher://una.hh.lib.umich.edu/00/ebb/monetary/noonfx.frb:")
  (Valid.value Y)
  (Valid.prop ": Unit-type None : Source URL gopher://una.hh.lib.umich.edu/00/ebb/monetary/noonfx.frb:")
  (Version.value 12/18/96)
  (Version.prop ": Unit-type None : Source URL gopher://una.hh.lib.umich.edu/00/ebb/monetary/noonfx.frb:")
)
```

Unit Value Mediator: Conversion Knowledge

- Unit-type conversion function

  f(a,v,v',+-1) = a'

  where a ::= data value

  v ::= data's unit value

  v'::= requested unit value

  a'::= converted data

  Example: time-range-f(10, day, hour, 1) = 240

- Conversion Properties
  - Total

    is one to which all conversion functions are defined for all arguments
  - Lossless

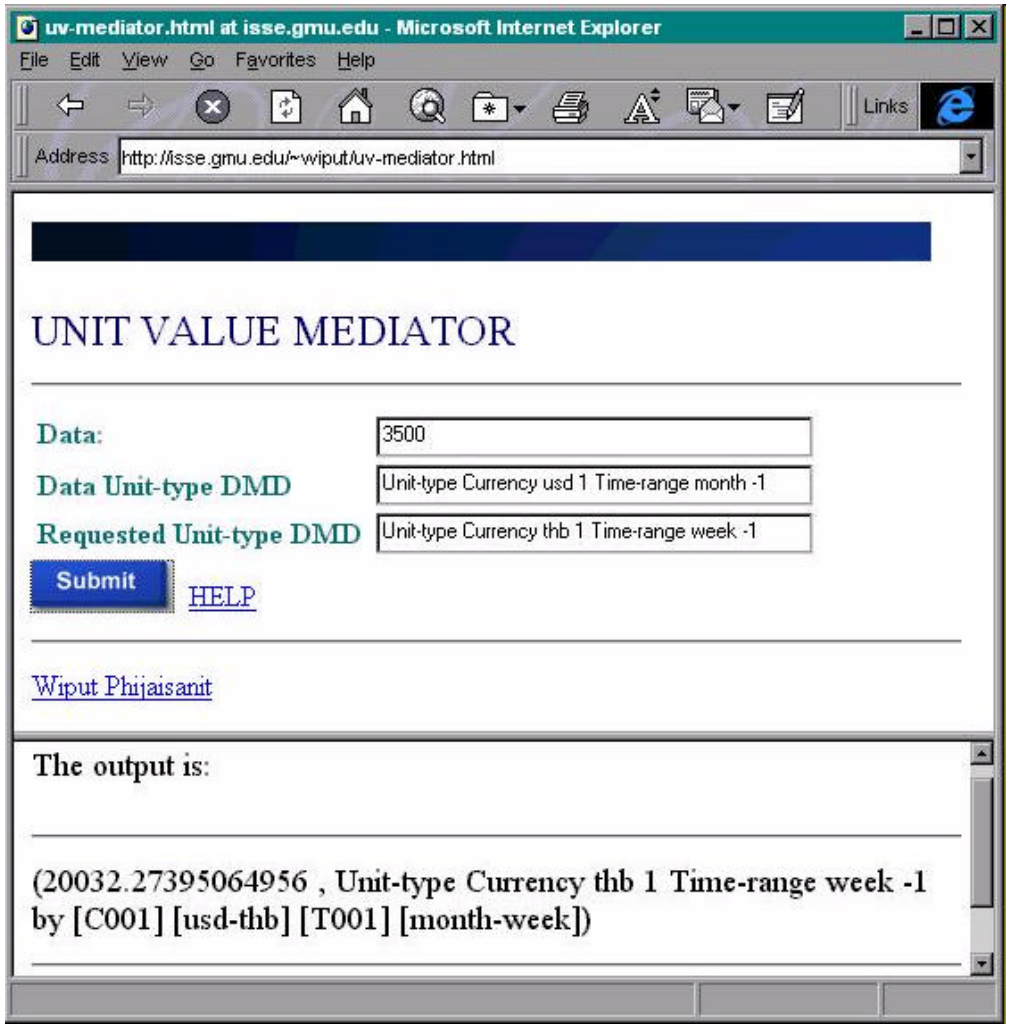
    is one to which it makes no difference whether a value is converted from one property value to another directly or sequentially.
  - Order Preserving

    is one that does not change the order of any two values after converting them.

Ref: Siegel & Sciore

Unit Value Mediator: Prototype

- Allows users to connect to the mediator by using any web browser.
- Provides unit conversion services over the Internet.
- Four unit types are supported that are Currency, Length, Time-range, and Weight.
- Provides “services” to other mediators.



Mediation Data Model

- Definition
  - Data model with additional constructs that supports static and dynamic meta-data (DMD) and incorporates mediators that use DMD.
- An attribute is described by a pair consisting of a value attribute and a property attribute.
  - Value attribute (Attribute.value) contains an actual data value.
  - Property attribute (Attribute.prop) contains dynamic meta-data that describes data in the value attribute.

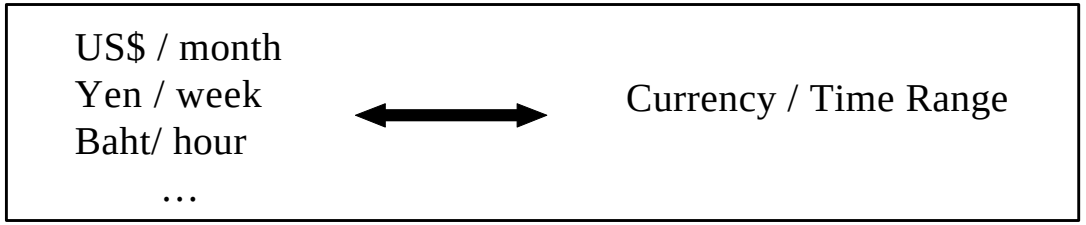
ObjectID	Aircraft-type.value	Aircraft-type.prop	Ceiling.value	Ceiling.prop
[gen1]	F-117A	: Source URL http://www.wpafb.af.mil/museum/modern_flight/mf22.htm : Unit-type None :	45000	: Source URL http://www.wpafb.af.mil/museum/modern_flight/mf22.htm : Unit-type Length foot 1 :
[gen2]	F-117A	: Source URL http://isse.gmu.edu/wiput/query.cgi : Unit-type None :	48000	: Source URL http://isse.gmu.edu/wiput/query.cgi : Unit-type Length foot 1 :
[gen5]	F-14	: Source URL http://isse.gmu.edu/wiput/query.cgi : Unit-type None :	15240	: Source URL http://isse.gmu.edu/wiput/query.cgi : Unit-type Length meter 1 :

Mediation Data Model (Cont.)

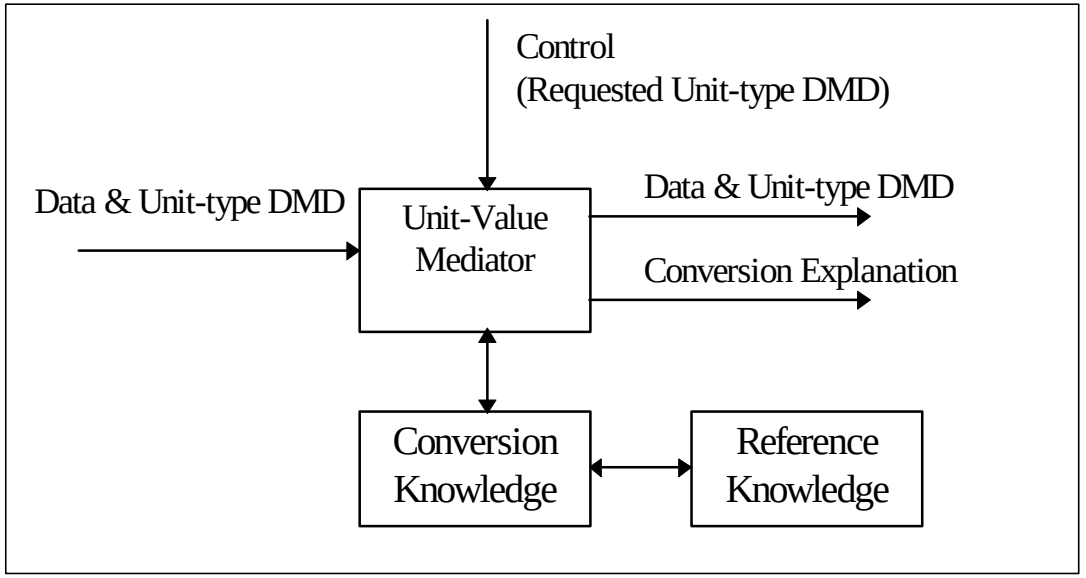
- Data is described in more detail, specifically, addition of dynamic meta-data
  - Assists users in understanding the role of data and its context semantics.
  - Provides conflict resolution support, and
  - Provides a framework for the specification of mediation services.
- Incorporation of DMD, mediators, and an extended data. manipulation language extends system capabilities.
  - Example: Multiple Unit Value (MUV) concept.

Multiple Unit Value (MUV) Concept

- **Concept:**
  - Object attribute values can be stored, retrieved and manipulated in more than one unit value.
  - Cooperation among Unit-Type DMD, UV mediator, and extended DML.
- **Unit-Type DMD:**
  - Extends the data model with Dynamic Meta-Data.
  - Specifies the unit type & unit value of a data element.
  - Provides the specification for processing by mediation services (UV Mediator).
- **Unit Value Mediator**
  - Special agent that is responsible for converting the unit values.



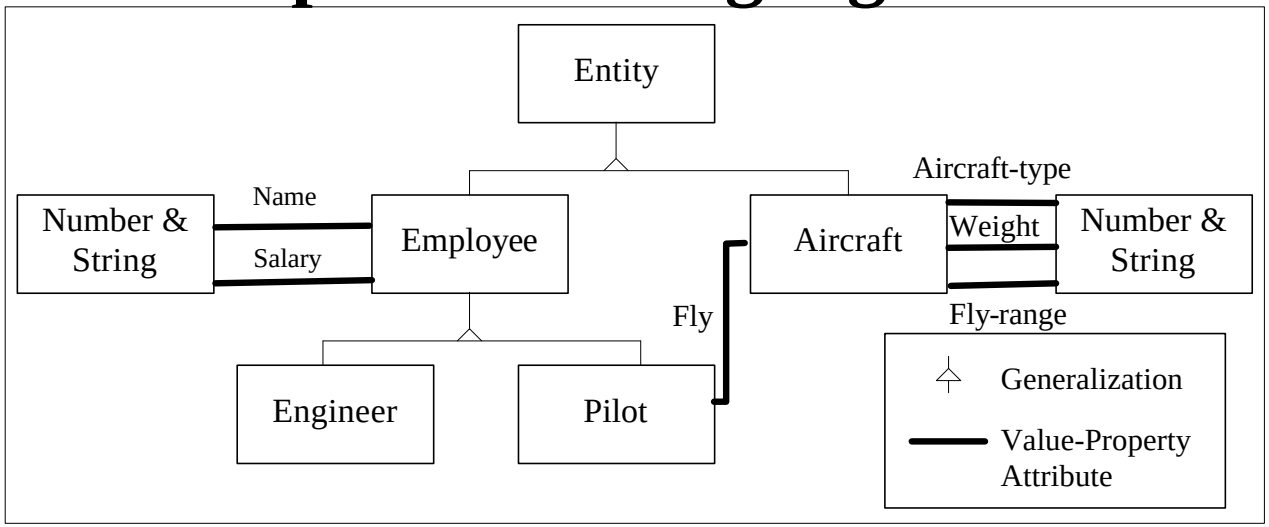
Unit-type DMD = Unit-type Currency jpy 1 Time-range Week -1



Extended Data Manipulation Language

Concept

Extension of CLIPS Object-Oriented Language (COOL) to incorporate MUV features into query language as an SQL-like template.



Question

Find Pilots who can fly aircraft that weight less than 12,000 kilograms. Report the pilot's name, how many Yen he earns per year, the aircraft type, and the range that the aircraft can fly in kilometers.

Query in the standard SQL DML --- not possible???

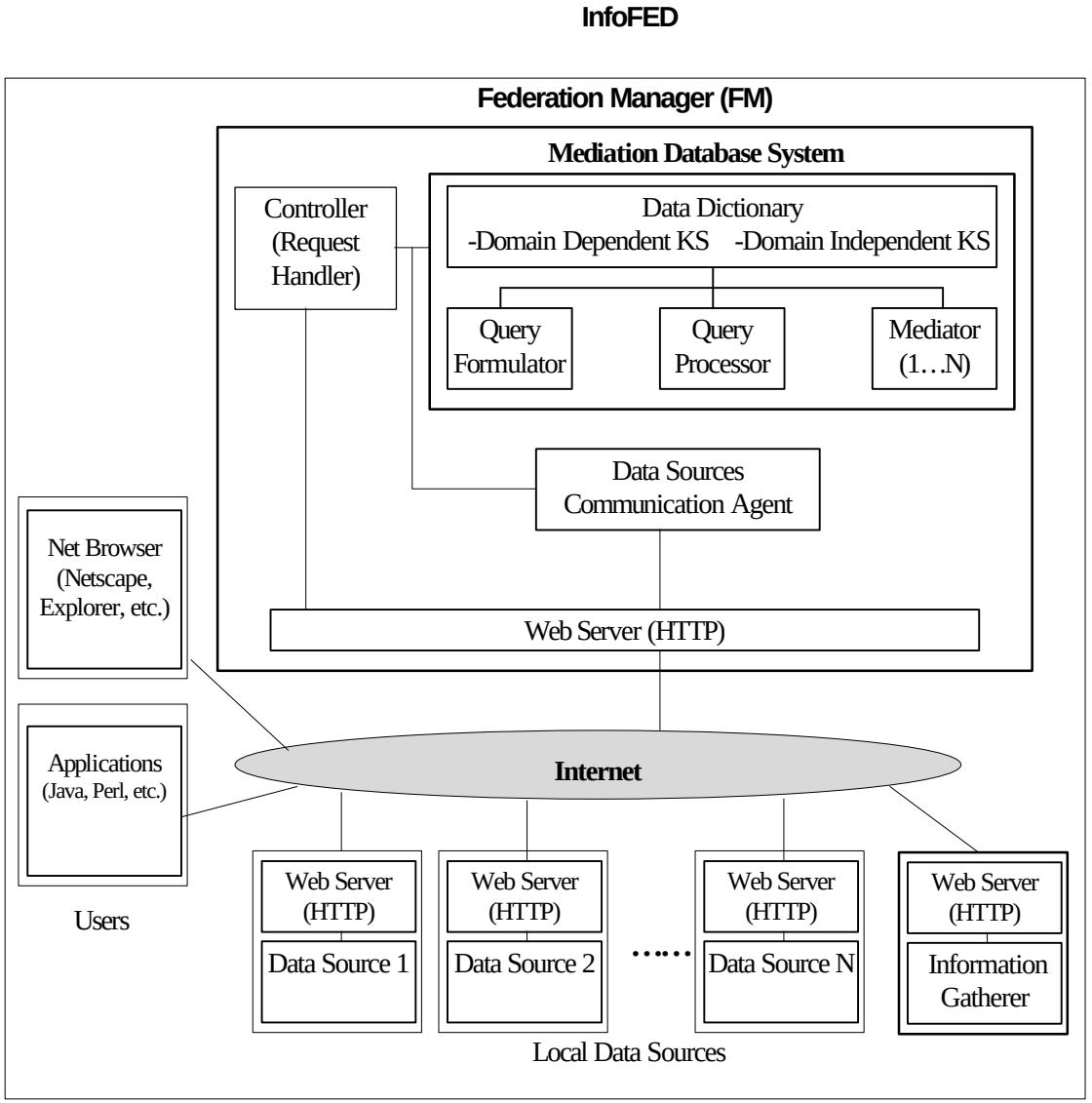
```
SELECT    p.Name.value, p.Salary, a.Aircraft-type.value, a.Fly-range
FROM      Pilot p, Aircraft a
WHERE     p.Fly = a.Aircraft  and a.Weight > 12000
```

Query in the extended COOL DML

```
SELECT    (?p:Name.value ? p:Salary  ?a:Aircraft-type.value  ?a:Fly-range)
UNIT-AS   ((= ?p:Salary.prop.Unit-type  'Unit-type Currency jpy 1 Time-range year -1')
          (= ?a:Fly-range.prop.Unit-type 'Unit-type Length kilometer  1')
FROM      ((?p Pilot) (?a Aircraft))
WHERE     (and  (eq  ?p:Fly.value ?a:Aircraft.value)
          (< ?a:Weight.value  ?a:Weight.prop.Unit-type  12000   'Unit-type Weight kilogram 1' ) )
```

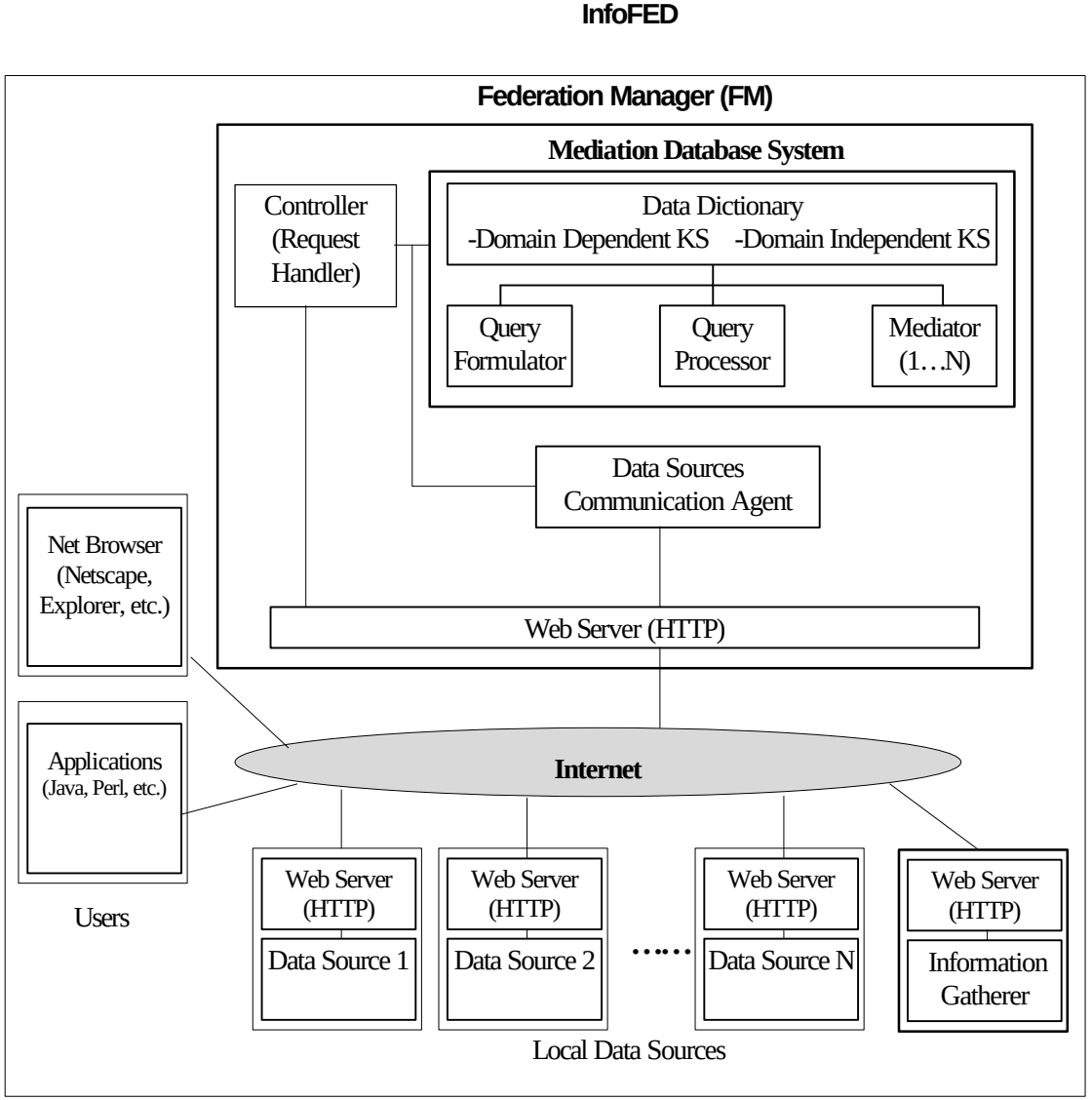
The InfoFED Federated Database System

- Definition
  - A virtual database system that integrates data from heterogeneous and autonomous data sources over the Internet.
- Provides Object-Oriented browser for data and meta-data:
  - Domain Dependent Knowledge Source (ontology).
  - Domain Independent Knowledge Source (conversion and reference knowledge).
- Use of Mediation Data Model
  - Supports data integration by reducing meta-data conflicts.
  - Intelligent DML based on Extended COOL.



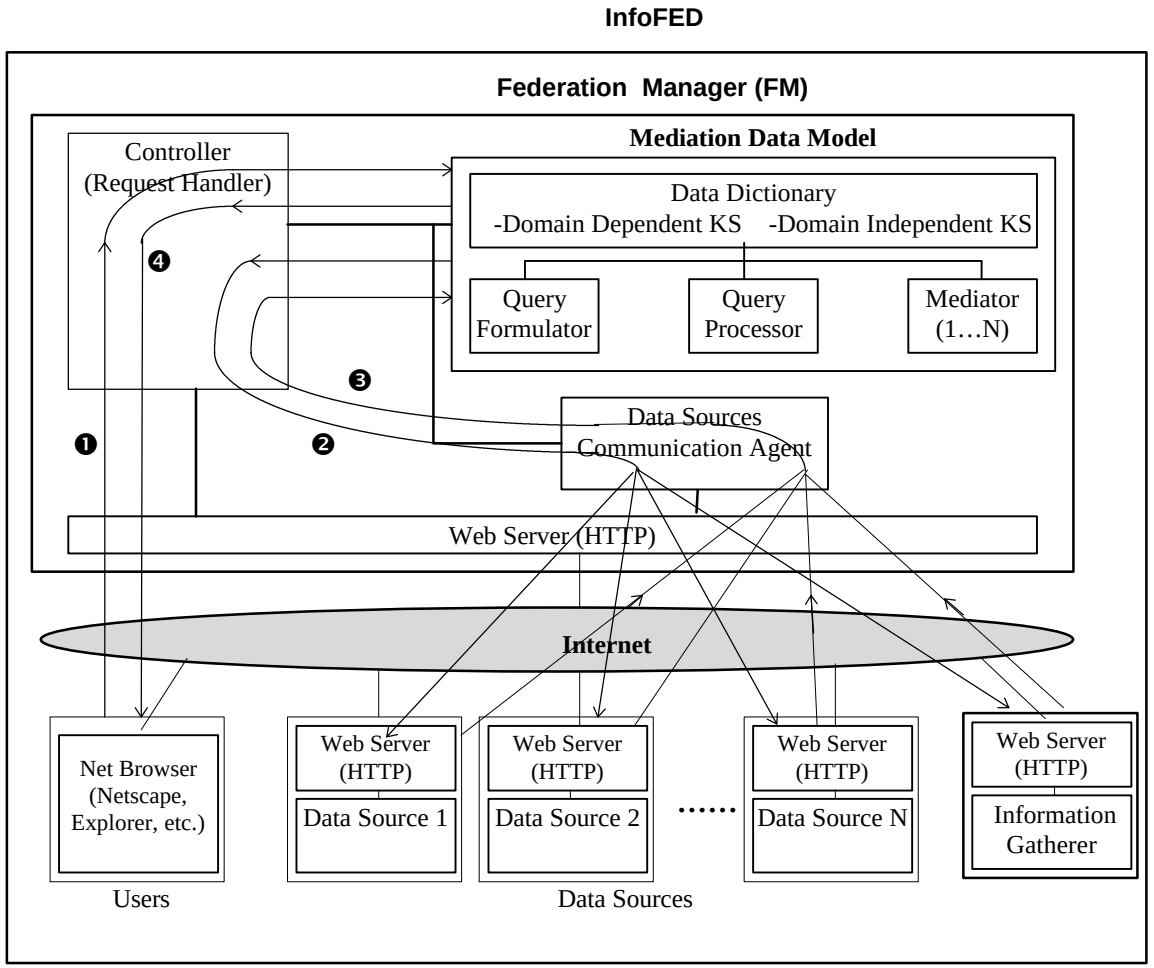
The InfoFED Federated Database System (Cont.)

- Efficient data retrieval in query processing
  - Retrieves only data that is related to the objects specified in the query.
- Easy data source management
  - Federation Manager can add or remove data source by editing the schema mapping knowledge source.
  - Data providers can register new sources/objects with InfoFED.
- Supports 3rd party agents, e.g.., information gatherer, to gather information for InfoFED (Data Sources Pull).



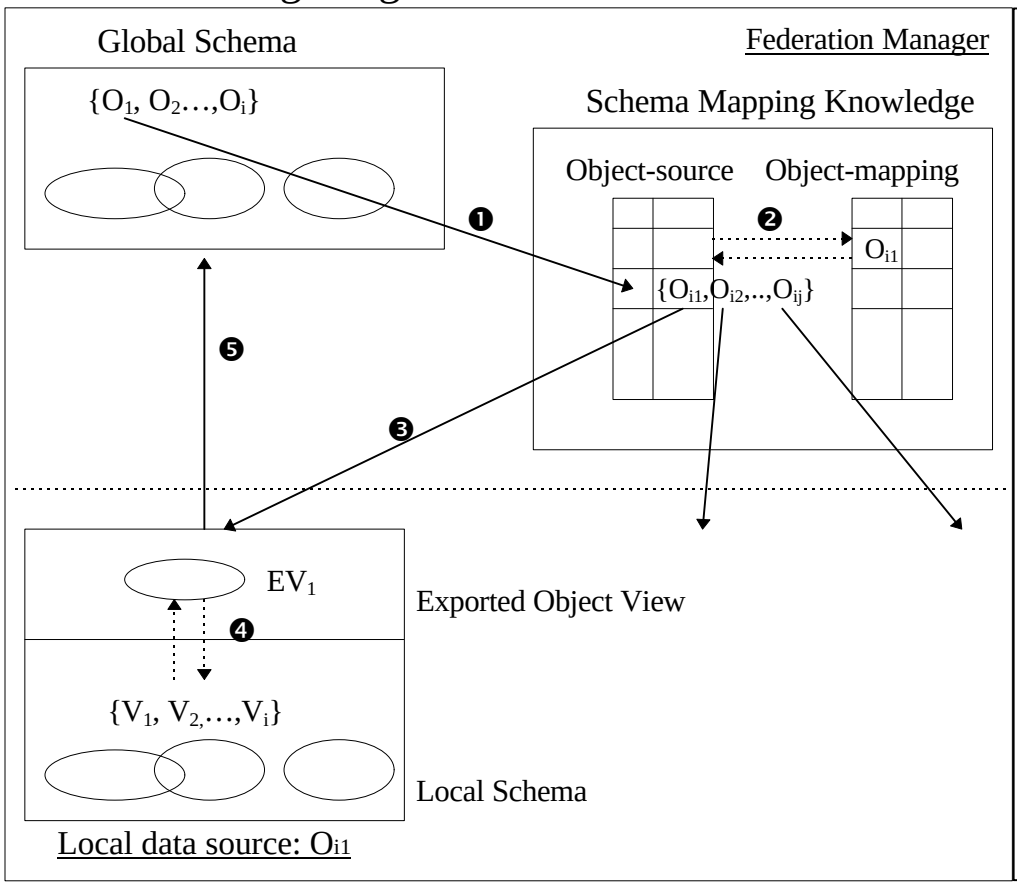
InfoFED: Working Scenario

1) The user submits a query to Federation Manager.
2) If external data is needed, Federation Manager connects to external data sources through the Data Source Communication Agent for retrieving data.
3) Each data source returns data to the Federation Manager.
4) The Federation Manager integrates the information from multiple sources and provides the final result to the user.



InfoFED: Local Data Sources Accessing Process

Data Accessing Diagram



Schema mapping Knowledge

Object-source

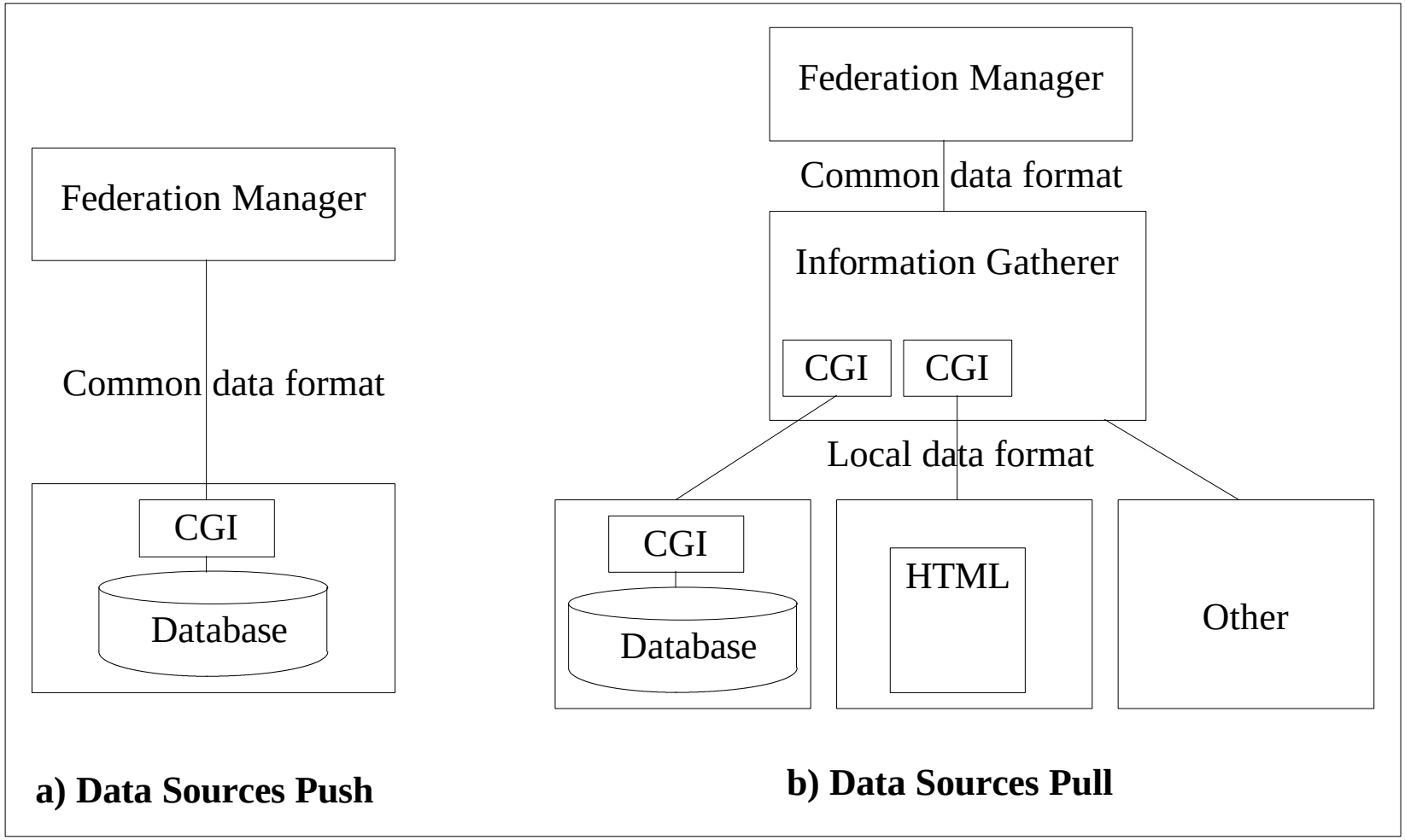
```
(Term01 of Object-source
    (Term Pilot)
    (Source-list Pi001 Pi002))
```

Object-mapping

```
(Pi001 of Object-mapping
    (Term Pilot)
    (Url-source  "http://site.gmu.edu/wiput/oracle.cgi")
    (Map-attribute
      : Ssn.value Attr Column-1
      : Ssn.property Default ": Source URL http://site.gmu.edu/wiput/oracle.cgi : Unit-type None :"
      : Last-name.value Attr Column-2
      : Last-name.property Default ": Source URL http://site.gmu.edu/wiput/oracle.cgi : Unit-type
        None :"
      : First-name.value Attr Column-3
      : First-name.property Default ": Source URL http://site.gmu.edu/wiput/oracle.cgi : Unit-type
        None :"
      : Salary.value Attr Column-4
      : Salary.property Default ": Source URL http://site.gmu.edu/wiput/oracle.cgi : Unit-type
        Currency usd 1 Time-range month -1 :"
      : Fly.value Attr Column-5
      : Fly.property Default ": Source URL http://site.gmu.edu/wiput/oracle.cgi : Unit-type None :": )
    (Map-query    "SELECT   e.Ssn, e.Last-name, e.First-name, e.Salary, p.Fly
                  FROM     Employee e, Pilot p
                  WHERE   e.Ssn = p.Ssn;" ))
```

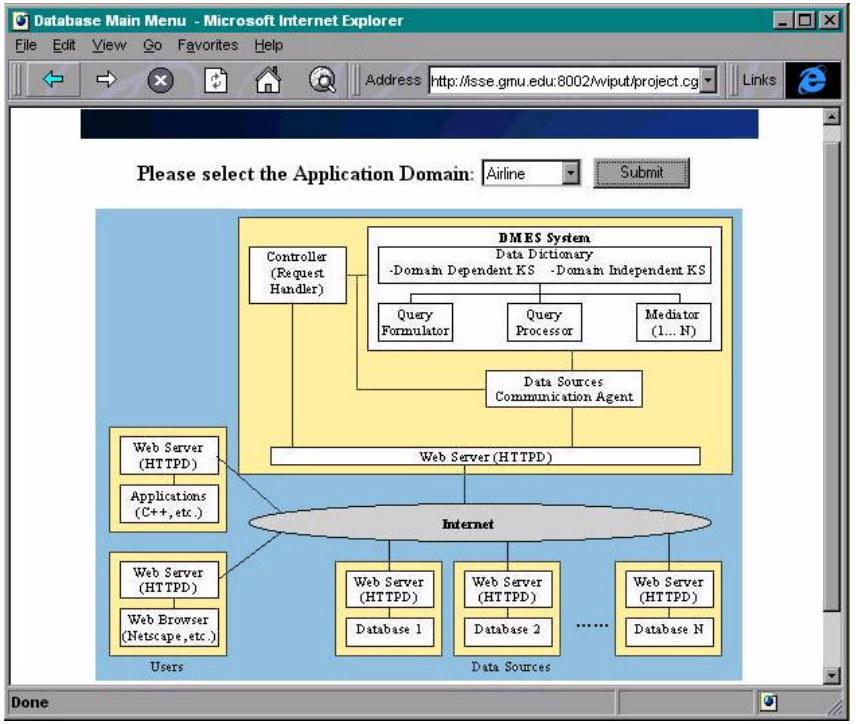
1) Federation Manager internally queries Schema Mapping Knowledge for the necessary information to retrieve data from these data sources supplying data for relevant objects.
2) Schema Mapping Knowledge provides mappings of global to local schema objects for retrieving data from each data source.
3) Federation Manager connects and retrieves data from each data source based on the information that is provided by Schema Mapping Knowledge.
4) Each data source internally prepares data.
5) Each data source returns data to the Federation Manager.

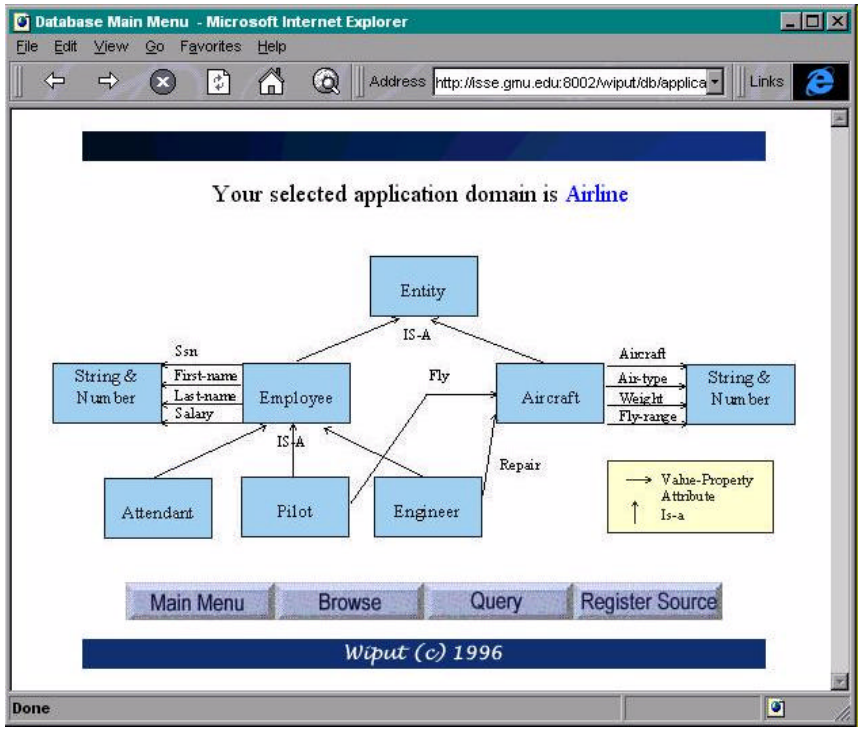
InfoFED: Data Obtaining Process



InfoFED: Prototype

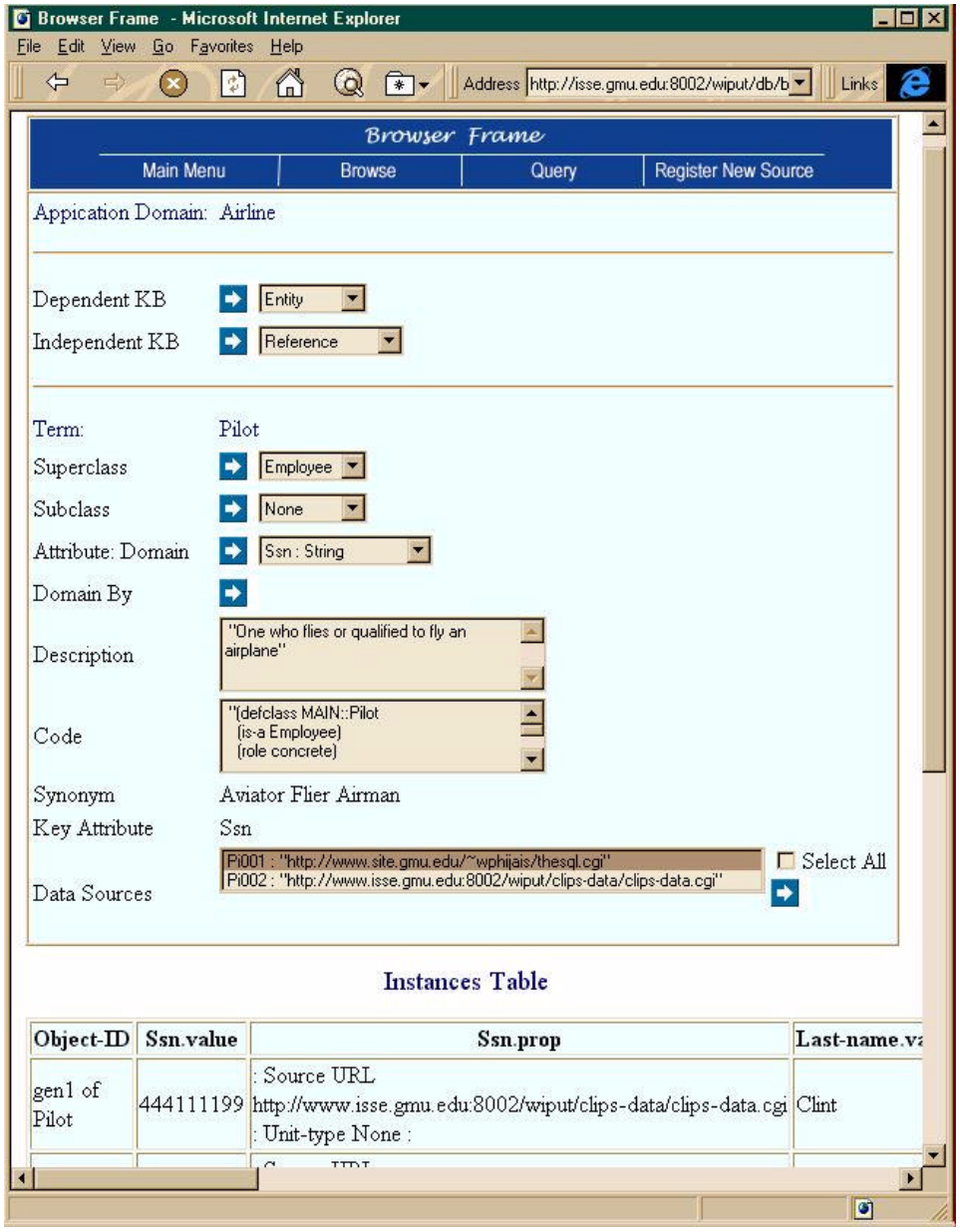
- Allows users to connect to InfoFED by using any web browser.
- Supports multiple application domains.
- Provides three modes of interaction
  - Browser Mode
  - Query Mode
  - Data Source Registration Mode.





InfoFED: Browser Mode

- Allows users to browse meta-data of both Domain Dependent Knowledge Sources and Domain Independent Knowledge Sources.
- Allows users to move along links to other classes for further investigation within the object schema.



InfoFED: Query Mode

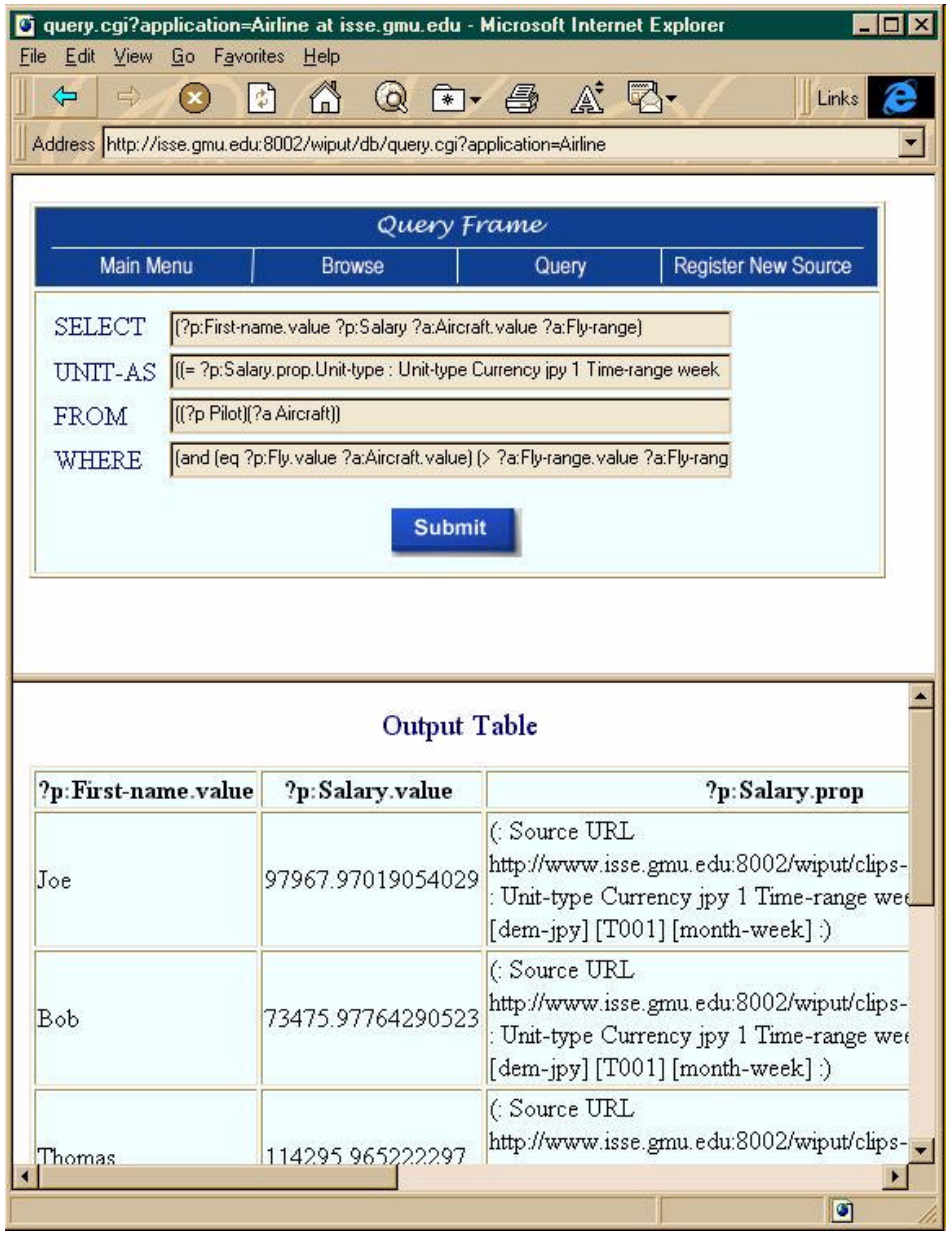
- Allows users to specify complex queries to InfoFED.
- Supports intelligent query language.
- Explain notation as the COOL DML extended with template:

  SELECT

  UNIT-AS

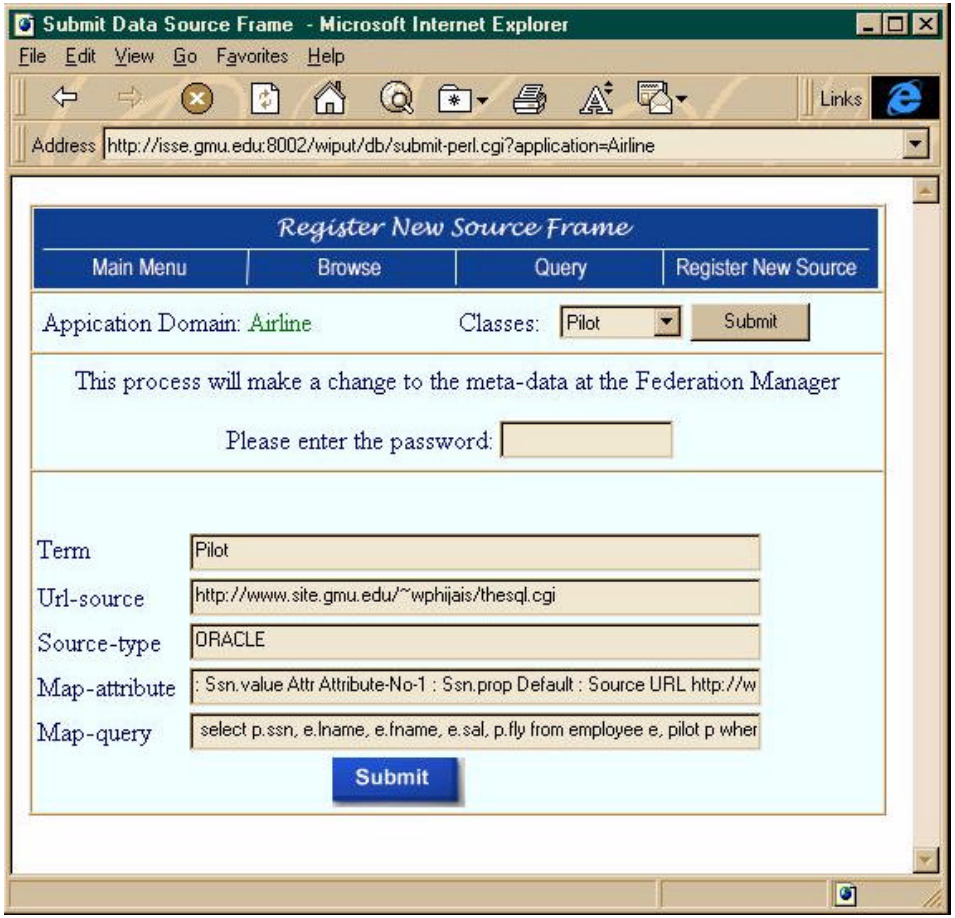
  FROM

  WHERE



InfoFED: Data Source Registration Mode

- Allows data providers to register and link their information sources to InfoFED by means of a straight-forward registration process.



Contributions

- Investigation of the role of meta-data in multidatabase systems to assist in the intelligent integration of information.
- Introduction of Dynamic Meta-Data (DMD) and its specification into a data model.
- Unit Value Mediator architecture.
- Introduction of Mediation Data Model which:
  - supports schema extensions that include Dynamic Meta-Data.
  - provides a framework for incorporating mediators into the data model.
- Multiple Unit Value Concept to allow automatic conversion mediation.
- InfoFED, a Federated Database System architecture and prototype that supports DMD and Mediation.

Contributions (Cont.)

- InfoFED Prototype Software
  - Incorporation of Mediation Data Model into the Federated Database System architecture
    - to support intelligent query language.
    - to reduce semantic integration conflicts.
  - Object Oriented support for
    - Federated Global Schema and Constituent Local Schemas.
    - Mapping from Global Schema query to Local Schema queries.
  - Dynamic linking to local data sources.
  - Registration Service for new sources.

Summary

- Dynamic Meta-Data (DMD) allows properties of objects to be stored, retrieved and manipulated.
- Mediation Data Model using DMD
  - provides better understanding of the semantics of objects,
  - enables InfoFED to handle more complicated queries, and
  - provides a framework for the specification of mediation services.
- Multiple Unit Value (MUV) concept allows object attributes to be expressed in convertible units.
- Multiple Unit Value concept provides an example of how we can utilize DMD (Unit-Type DMD) incorporated with a mediator and extended DML to enhance system capability.

Summary (Cont.)

- The InfoFED system makes use of Mediation Data Model to provide:
  - a browser for both data and meta-data,
  - object-oriented federated schema,
  - data integration and query processing, and
  - data source registration and management.

Future Directions

- Investigation of new types of Dynamic Meta-Data and mediators to support new capabilities.
  - Data source DMD and data source mediator to support the filtering services.
  - Media type DMD and Media type mediator to support multimedia objects in the data model.
- Performance and Query Optimization Strategy for dealing with queries that involve mediation services (mediation cost).

The InfoFED Federated Database System

- Definition
  - A virtual database system that integrates data from heterogeneous and autonomous data sources in the Internet environment
- Provides Object-Oriented browser for data and meta-data:
  - Domain Dependent Knowledge Source (ontology)
  - Domain Independent Knowledge Source (conversion and reference knowledge)
- Use of Mediation Data Model
  - Supports data integration by reducing meta-data conflicts
  - Intelligent DML based on Extended COOL