



FIZ CHEMIE BERLIN
Fachinformationszentrum Chemie GmbH

Freie Universität



Berlin

SEMANTIC TECHNOLOGIES APPLIED TO CHEMISTRY

ChemCloud - Chemical e-Science Information Cloud

Adrian Paschke, Freie Universität Berlin

Stephan Heineke, FIZ CHEMIE

About FIZ CHEMIE

- **1830 Founding of Pharmaceutisches Centralblatt**
 - **Reestablished 1981 by the German Government**
- **German Chemistry Information Center**
- **Focus:**
 - Information Competence**
 - Educational Systems**
 - Reaction Databases**
 - Thermophysical Databases**

About Corporate Semantic Web

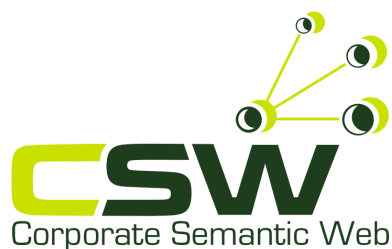
1. Application of Semantic Web technologies in enterprise information systems (Semantic Enterprise)

- Collaborative workflows and (business) process management (e.g. e-Science workflows, Semantic Business Process Management)
- Knowledge Management (e.g. Semantic Knowledge Management, Semantic Corporate Memory)

2. Corporate = Business Context

- Application of Semantic Web technologies under economical considerations and business conditions (e.g. cost models, return on investment)

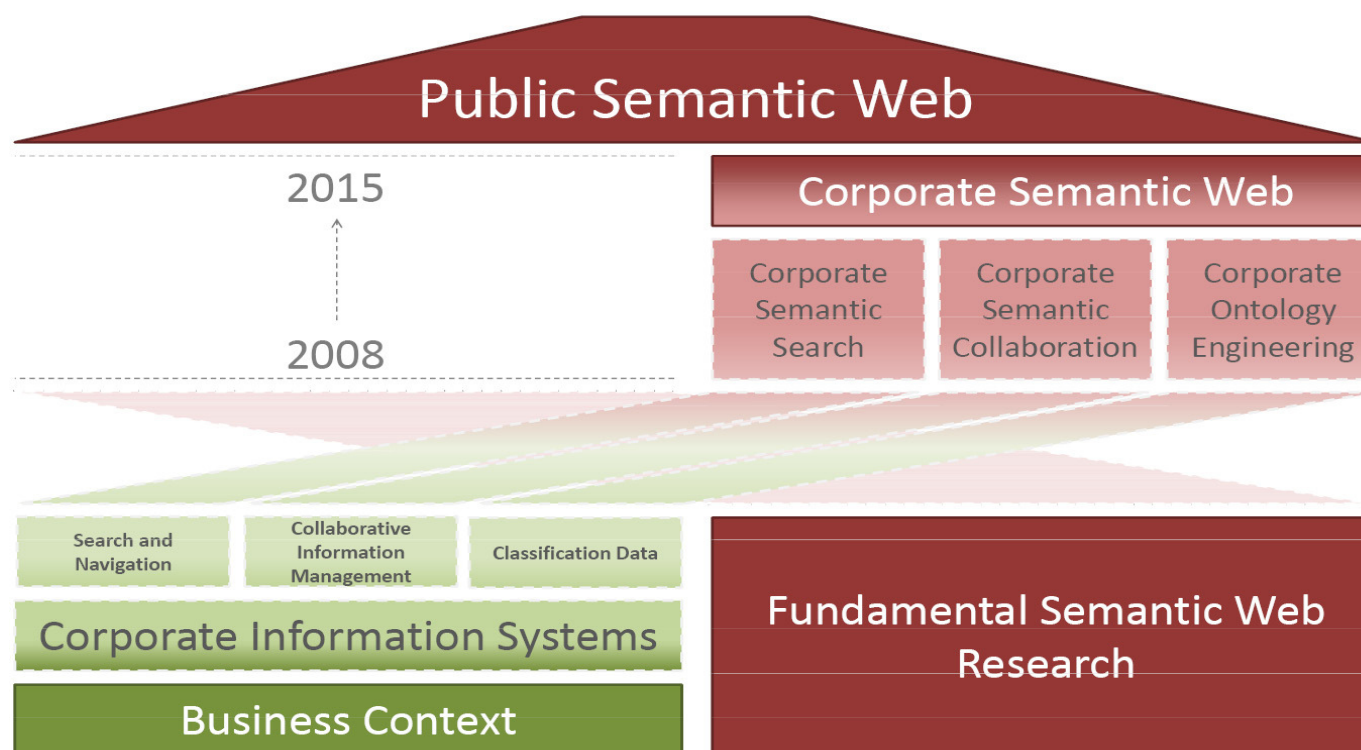
About Corporate Semantic Web at Freie Universität Berlin



Freie Universität



Berlin



Agenda

- **ChemCloud – Chemical e-Science Information Cloud**
- **Semantic Model of the ChemCloud**
- **Exemplary Use Cases**
- **Conclusion**

Support of Knowledge Intensive eScience Processes

Research &
Development



Knowledge

→ Event & Process
Context

Knowledge
Management:

Semantics

→ Relations
& Interpretation

Information
Sources:



Literature



Colleagues



Experts

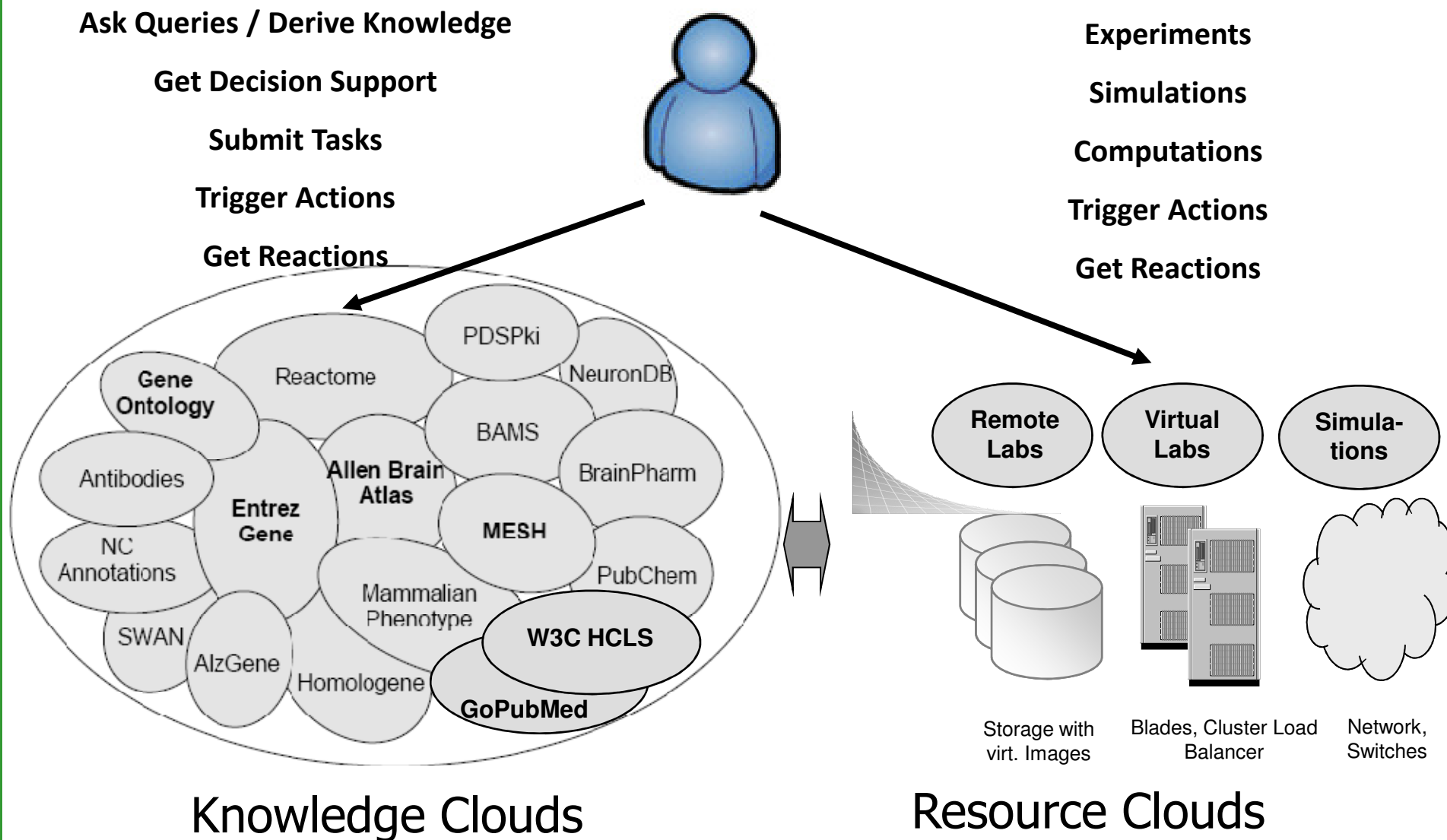


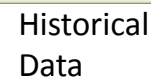
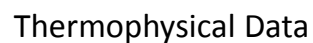
Databases

Information

→ Content

Motivation: Emerging e-Science Infrastructure





Health Care and Life Sciences Knowledge Base

We are going to host parts of the [W3C HCLS KB](#) at the Freie Universität Berlin. Currently we are in the setup process and more information will be available soon.

What is W3C HCLS ?

The W3C Health Care and Life Science Interest Group (W3C HCLS IG) is chartered to develop and support the use of Semantic Web technologies and practices to improve collaboration, research and development, and innovation adoption in the Health Care and Life Science domains. It is structured into six task-forces:

1. BioRDF (Structured Data to RDF)
2. Scientific Discourse
3. Clinical Decision Support
4. Clinical Observations Interoperability
5. Linking Open Drug Data
6. Terminology

Knowledge Bases

Following Knowledge Bases are currently imported in the Triplestore:

- SenseLab
- TCMOpenDT

Backend:

Backend:
We use an [AllegroGraph triplestore](#) with an integrated Sesame SPARQL Endpoint.

Team:

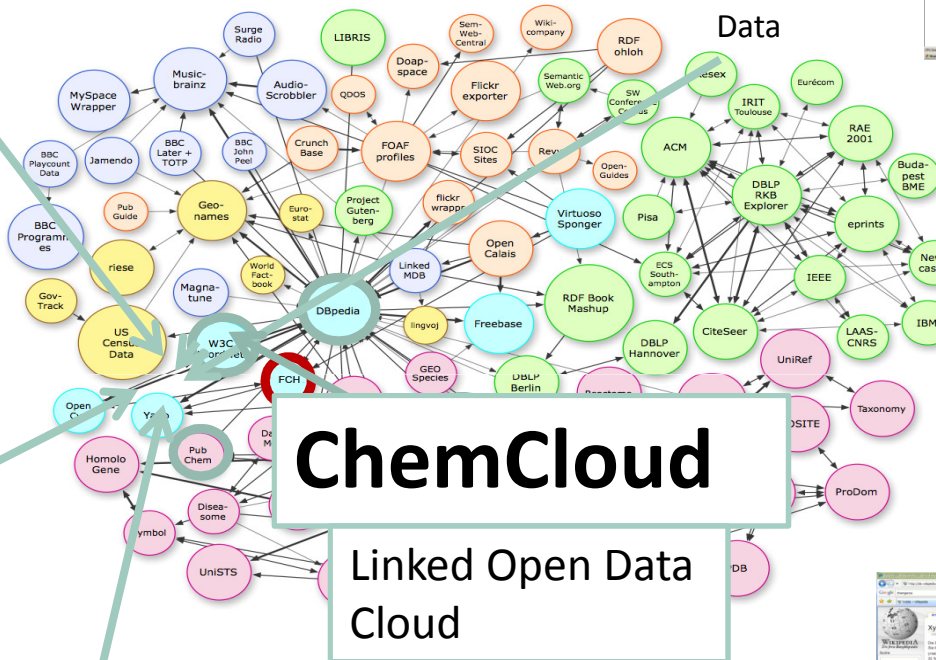
Team:
Mario Rothe (contact)
Prof. Dr. Adrian Pasch



FU Berlin W3C HCLS KB

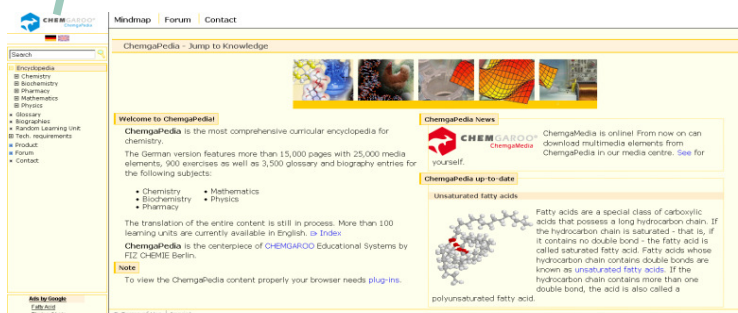


Literature Data

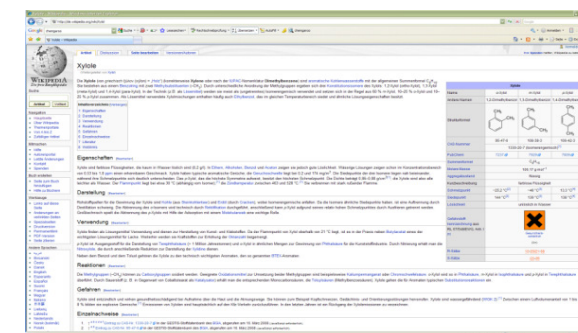


ChemCloud

Linked Open Data Cloud



eLearning Content



FU Berlin DBPedia (Germany)

Example 1 - Thermophysical Linked Data Cloud

■ Type of data

- Experimental thermodynamic and physical properties of mixtures and pure substances from a total of 22,000 compounds

■ Knowledge base growth

- About 1,000 tables per month

■ Content

- 450,000 tables and diagrams of PVT state values, phase equilibrium data, transport and surface properties, caloric, optical and acoustic properties

■ Material types

- 75% organic, 20% inorganic and 5% organometallic compounds with the focus on solvents

■ Sources

- 37,900 publications from journals and reports as well as measurement protocols and data collections in printed and electronic format, from 1919 to present, ~ 800 predefined therms

Example 2 - eLearning Linked Data Cloud

■ Type of Data

- Curricular Encyclopedia for Chemistry

■ Content

- 1,800 learning units with 15,000 content pages, 25,000 media elements, 900 exercises as well as 3,500 glossary and biography entries

■ Subjects

- Chemistry, Biochemistry, Pharmacy, Mathematics, Physics

Example 3 – Literature Linked Data Cloud

- **Type of data**
 - STM Literature
- **Content**
 - About 500,000 documents
- **Subject**
 - Science, Technology and Medicine
- **Sources**
 - eBooks and papers

Example 4 – DBpedia (Germany) Linked Data Cloud

■ Type of data

- German part of Wikipedia

■ Content

- Wikipedia articles and meta data

<http://de.dbpedia.org>

■ Material Subject

- All subjects included



■ Sources

- Wikipedia
Germany

Start

Mitmachen

Onlinezugriff

Technik

Impressum

Willkommen

DBpedia Deutschland stellt die strukturierten Informationen der [deutschen Wikipedia](#) frei zur Verfügung. Dabei ist dies ein Teil des internationalen [DBpedia](#) Projektes, wobei wir uns auf die Extraktion und Verfügungsstellung der Informationen der deutschen Wikipedia beschränken. Damit wird es ermöglicht, optimale strukturierte Informationen zu Nutzen die auf Anwendungen für deutsche Benutzer zugeschnitten sind. Das Ziel ist die Integration der deutschen Informationen in eine DBpedia Linked Data Wolke, welche die nationalen Datensätze miteinander verbindet. Mehr Informationen über das DBpedia Projekt befinden sich auf <http://dbpedia.org>

mit freundlicher Unterstützung:



Neuigkeiten

02.07.2012 13:09 von Sebastian Krebs
[Update auf AllegroGraph 4.0](#)
[Weiterlesen ...](#)

09.07.2010 15:46 von Sebastian Krebs
[Kategorien und Skos-Konzepte](#)
[Weiterlesen ...](#)

Agenda

- ChemCloud – Chemical e-Science Information Cloud
- **Semantic Model of the ChemCloud**
- Exemplary Use Cases
- Conclusion

Top Level Modular Ontology Model

■ General Ontologies

- General concepts commonly used in the domain ontologies, e.g. content, product, term

■ Chemistry Domain Ontologies

- Semantic Chemistry Concept, e.g. thermo-physical properties, reactions, compound classes

■ Product Domain Ontologies

- Product/Services Descriptions

■ Meta Domain Ontologies

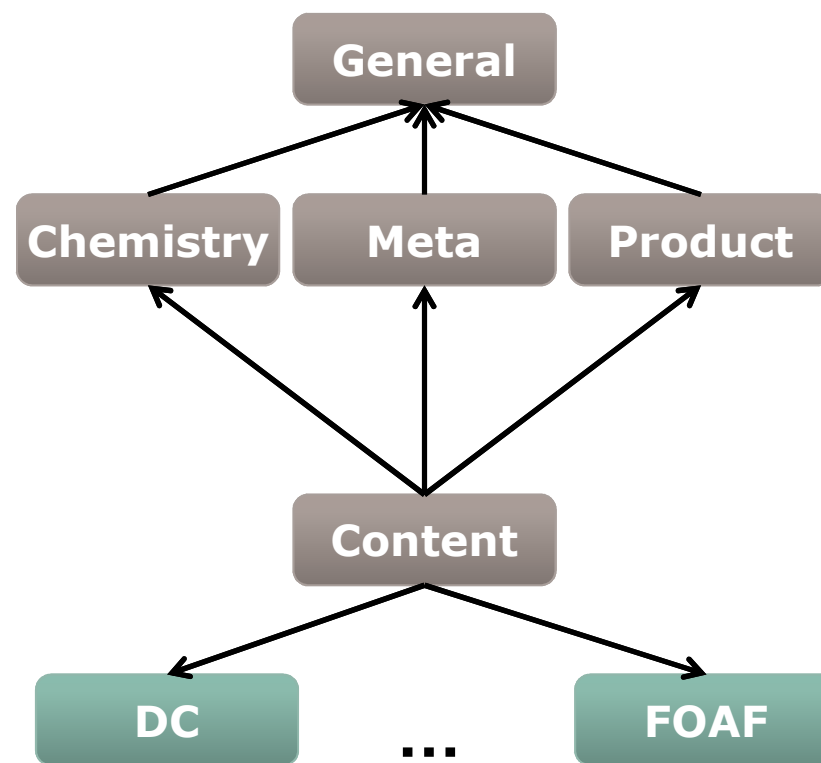
- e.g. provenance of data for DBPedia Deutschland, eBooks, eLearning; mapping ontologies

■ Content-Ontologies

- e.g. content e-Learning, eBooks, InfoTherm

■ Standard-Ontologies

- FOAF (Persons), Dublin Core (Content)



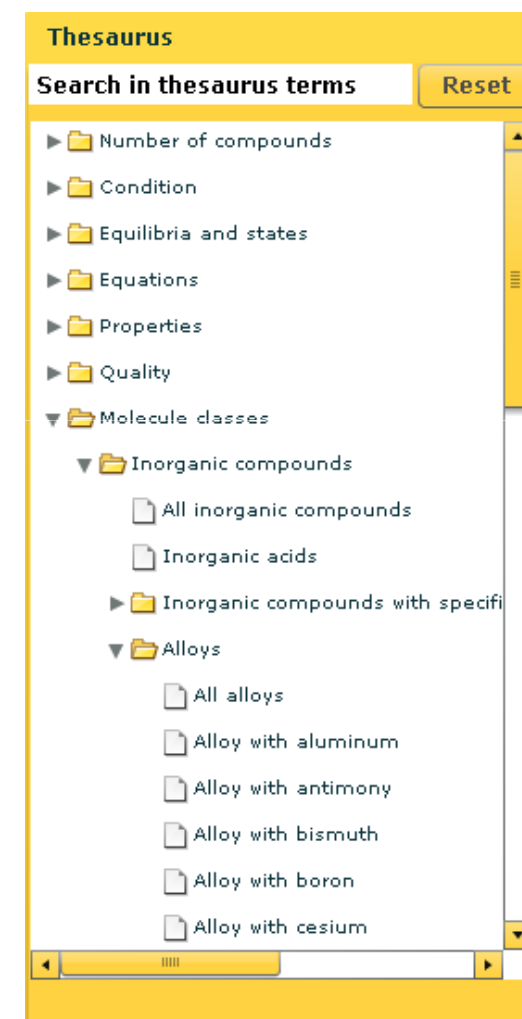
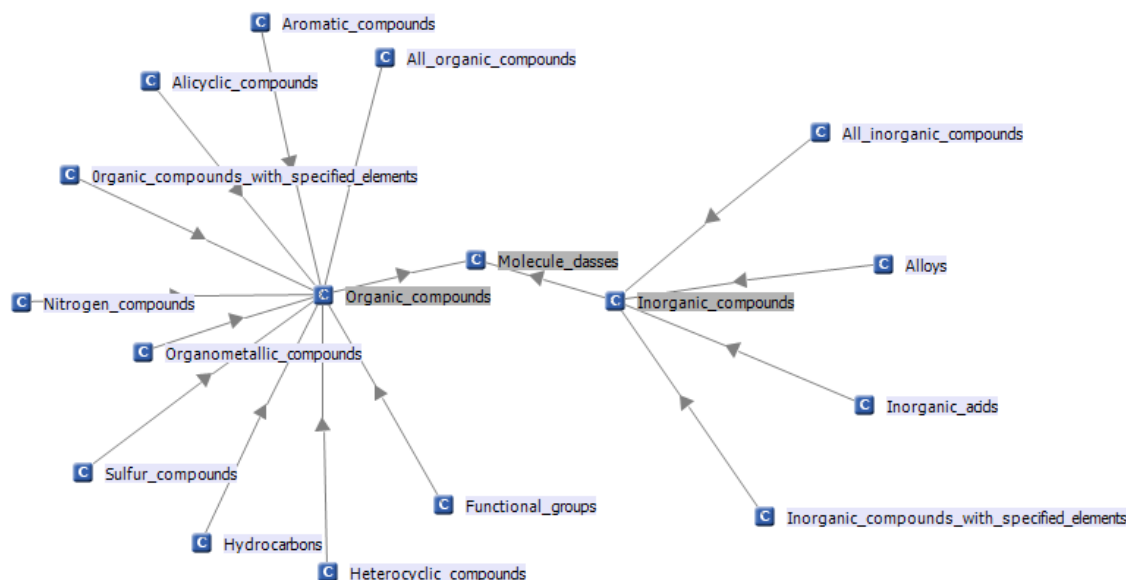
Thermo – Physical Domain Ontology

■ >800 pre-defined terms in Thesarus

- e.g. category "Molecule classes" approx. 500 molecule classes
- e.g. "metal organic compounds" or "inorganic acids" and their thermophysical properties



■ Semantic Web Thesarus



ThermoML Chemistry Domain Ontology



ThermoML:Property

```
[ a    ThermoML:PropertyType ;  
  ThermoML:PropPhaseID  
[ a    ThermoML:PropPhaseIDType ;  
  ThermoML:ePropPhase "Liquid"  ] ;  
  ThermoML:PropRepeatability  
[ a    ThermoML:PropVarRepeatabilityType ;  
  ThermoML:eRepeatMethod  
"Standard deviation of a single value (unbiased)" ;  
  ThermoML:sRepeatEvaluator  
"Author"  ] ;  
  ThermoML:Property-MethodID  
[ a    ThermoML:Property-MethodIDType ;  
  ThermoML:PropertyGroup  
[ a    ThermoML:PropertyGroupType ;  
  ThermoML:PhaseTransition  
[ a    ThermoML:PhaseTransitionType ;  
  ThermoML:ePropName "Molar enthalpy of vaporization or sublimation, kJ/mol" ;  
  ThermoML:sMethodName  
"Correlation Gas Chromatography"
```

CHEMAXIOM+CHEBI+THERMOML.OWL

- Annotate **ThermoML:ePropName** with terms from **ChemaxiomProp**
- Annotate **measuring units** with terms from **ChemAxiomUnits**
- Annotate measuring methods with terms from **ChemAxiomMetrology**
- Align ThermoML.owl and ChemaxiomChemDomain for Compound names and metadata
- Add the roles from Chebi Role Ontology, structure from Molecular Structure Ontology and chemistry relationships

Top Level Modular Ontology Model

■ General Ontologies

- General concepts commonly used in the domain ontologies, e.g. content, product, term

■ Chemistry Domain Ontologies

- Semantic Chemistry Concept, e.g. reactions, compound classes

■ Product Domain Ontologies

- Product/Services Descriptions

■ Meta Domain Ontologies

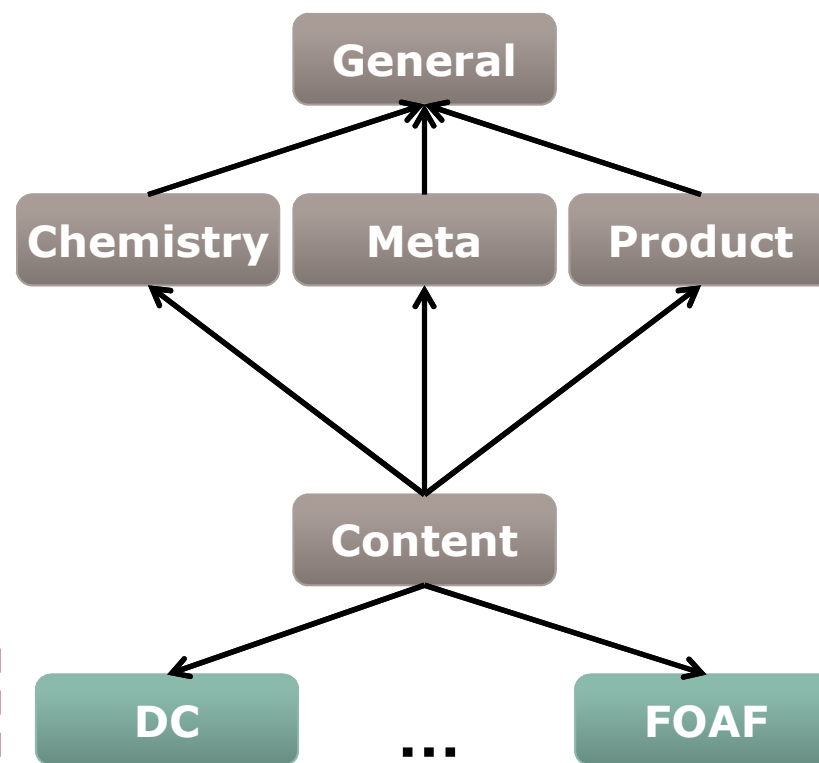
- e.g. provenance of data for DBPedia Deutschland, eBooks, eLearning; mapping ontologies

■ Content-Ontologies

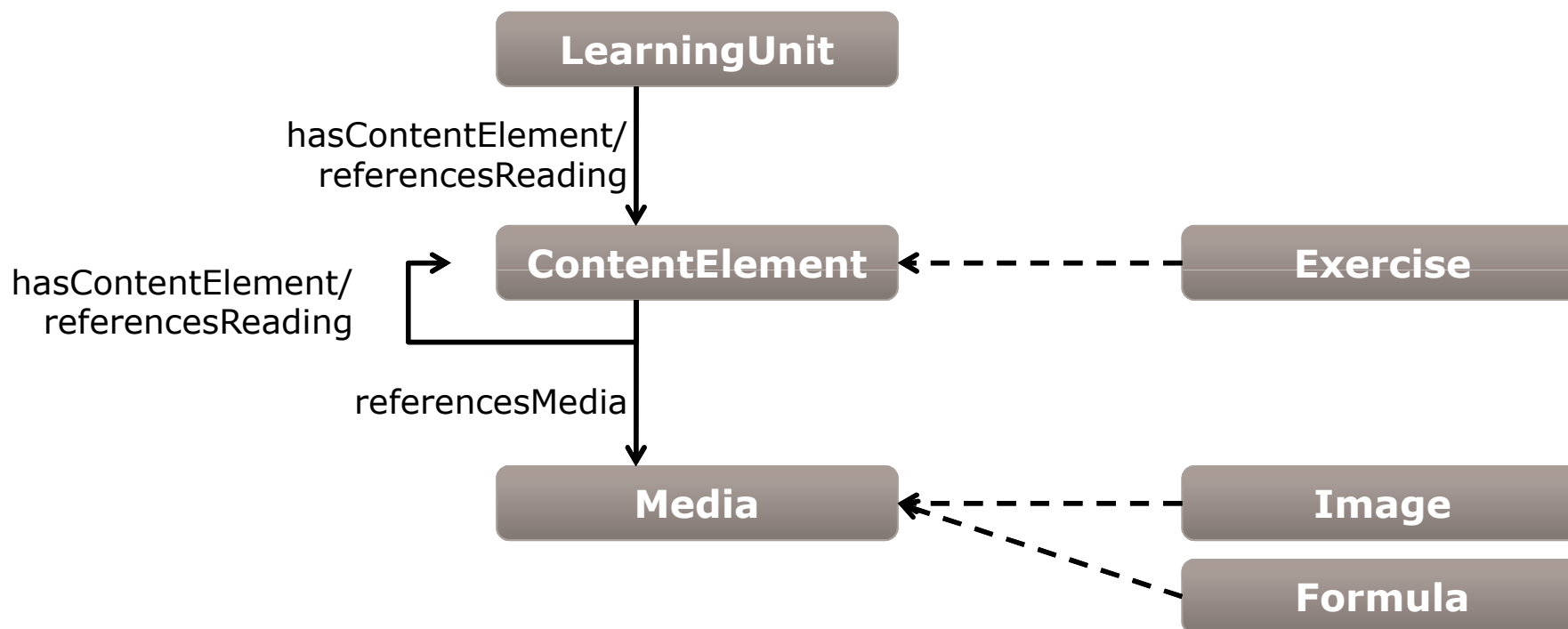
- e.g. content e-Learning, eBooks

■ Standard-Ontologies

- FOAF (Persons), Dublin Core (Content)



E-Learning-Ontology: Content



Example – Linked Data Pubby Interface eLearning Data Cloud

PeriodensystemDarstellung at FIZ-Chemie

<http://de.dbpedia.org/fiz-chemie/resource/ontologies/2010/2/learning.owl%23PeriodensystemDarstellung>

Property	Value
dcterms:abstract	■ Aufbau, Einteilung und Darstellungsformen von Periodensystemen.
dcterms:creator	■ learning:grPaderborn ■ learning:vsPaderborn
dcterms:educationLevel	■ learning-meta:School_B ■ learning-meta:Training_B ■ learning-meta:University_A
dcterms:extent	■ learning:_30min
learning-meta:highPriorityKeyword	■ chemie:Atombau
is learning-meta:objective of	■ learning:PeriodensystemDarstellung
learning-meta:objective	■ learning:PeriodensystemDarstellung
learning-meta:premise	■ learning:Atombau
learning-meta:recommendedReading	■ learning:PeriodensystemAtomradien ■ learning:PeriodensystemAubauprinzip ■ learning:PeriodensystemEinteilung ■ learning:PeriodensystemElektronenaffinitaet ■ learning:PeriodensystemHistorie ■ learning:PeriodensystemIonisierungsenergie
dcterms:source	■ https://vscms.ernetztes-studium.de/webdaw/files/vsc/de/ch/11/aac/vorlesung/kap_3/vlu/periodensystem.vlu
dcterms:subject	■ chemie:Periodensystem
dcterms:title	■ Darstellungsform des Periodensystems
rdf:type	■ learning-meta:LearningUnit

Example DBpedia Deutschland Triplestore Web View

AllegroGraph Web View browsing database dbpedia

« | **Overview** | Queries: [new](#), [saved](#), [recent](#) | Namespaces | User: [logout](#), [delete](#)

fu-berlin

[show namespaces](#)

```
select ?v ?o where {<http://dbpedia.org/resource/Aceton> ?v ?o}
```

as (optional) ☐ Shared

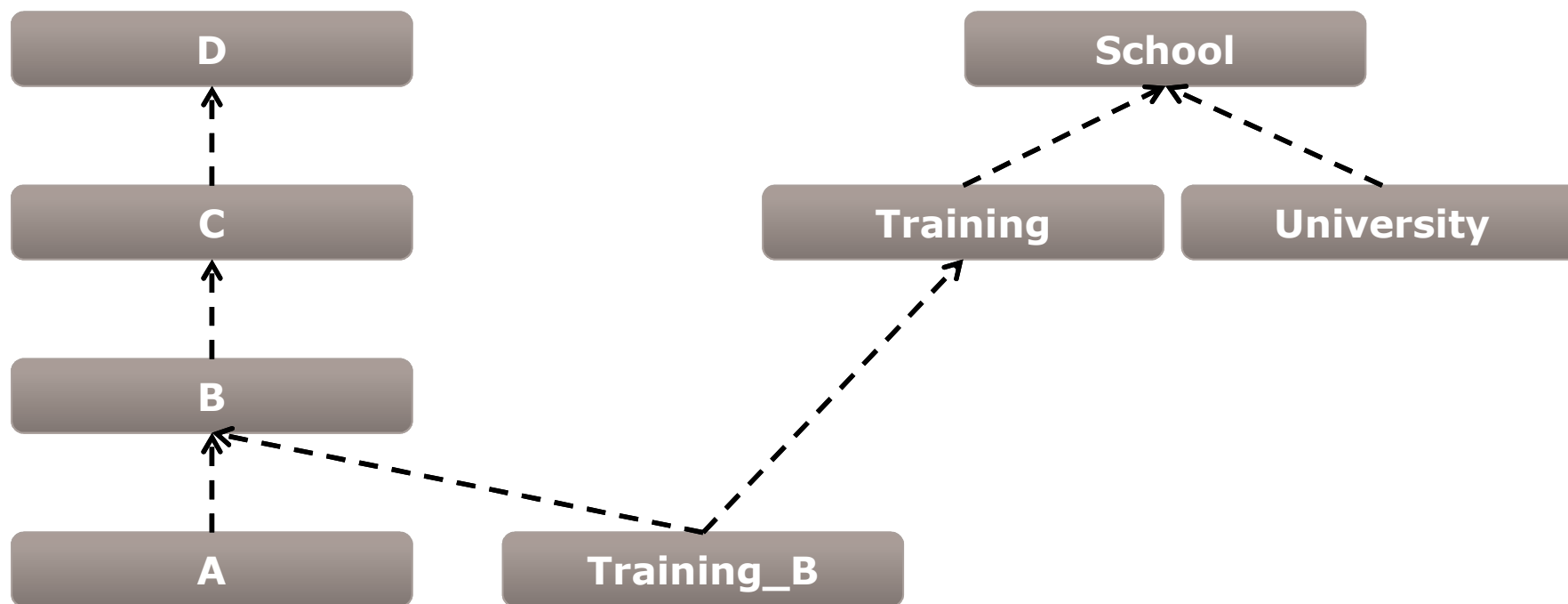
Result

?v	?o
l_percent_C3_percent_B6slichkeit	"mischbar mit Wasserund vielen organischen Lösungsmitteln"
beschreibung	"farblose Flüssigkeit"
siedepunkt	"56.0"
andereNamen	"* Propanon\n* Propan-2-on\n* 2-Propanon\n* Dimethylketon\n* Acetonum"
aggregat	"flüssig"
page	Aceton
wikilink	Acetonperoxid
wikilink	Hydroperoxide
wikilink	Ketonk%C3%B6rper
wikilink	Datei: Synthesis_Diacetone_alcohol.svg
wikilink	Cumol
wikilink	Edukt
wikilink	Diabetes_mellitus
wikilink	Celluloseacetat
wikilink	Autokatalyse
wikilink	Methylgruppe

Agenda

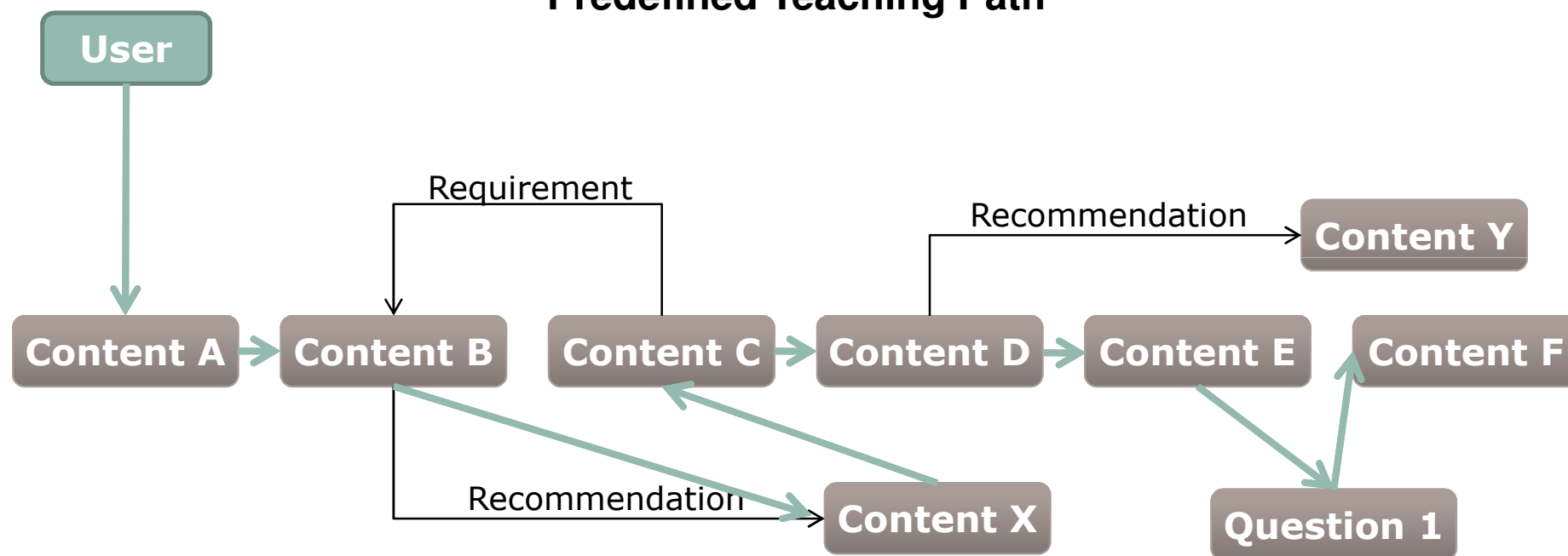
- ChemCloud – Chemical e-Science Information Cloud
- Semantic Model of the ChemCloud
- **Exemplary Use Cases**
- Conclusion

E-Learning-Ontology: Learner



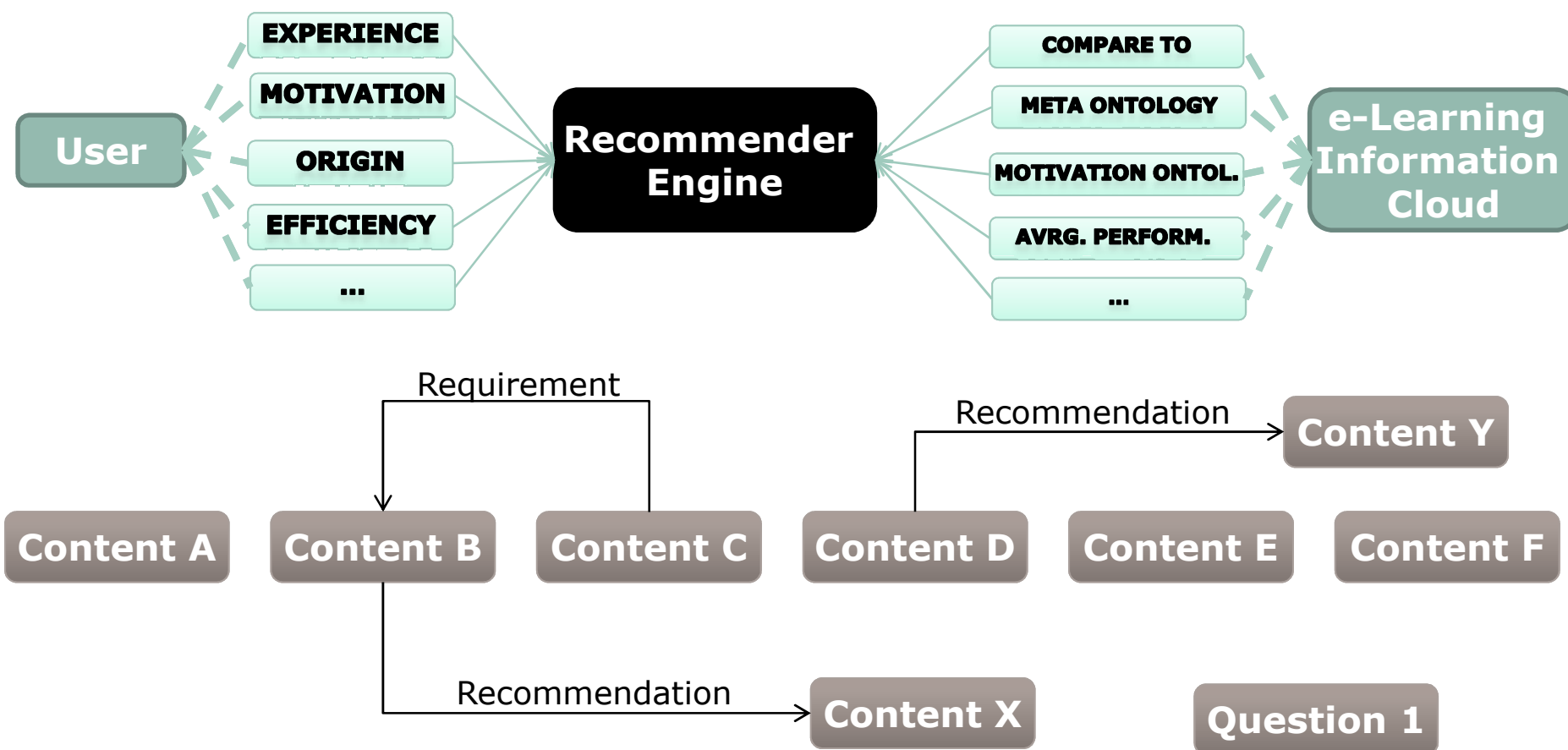
Example 1: User e-learning path recommendation

Predefined Teaching Path

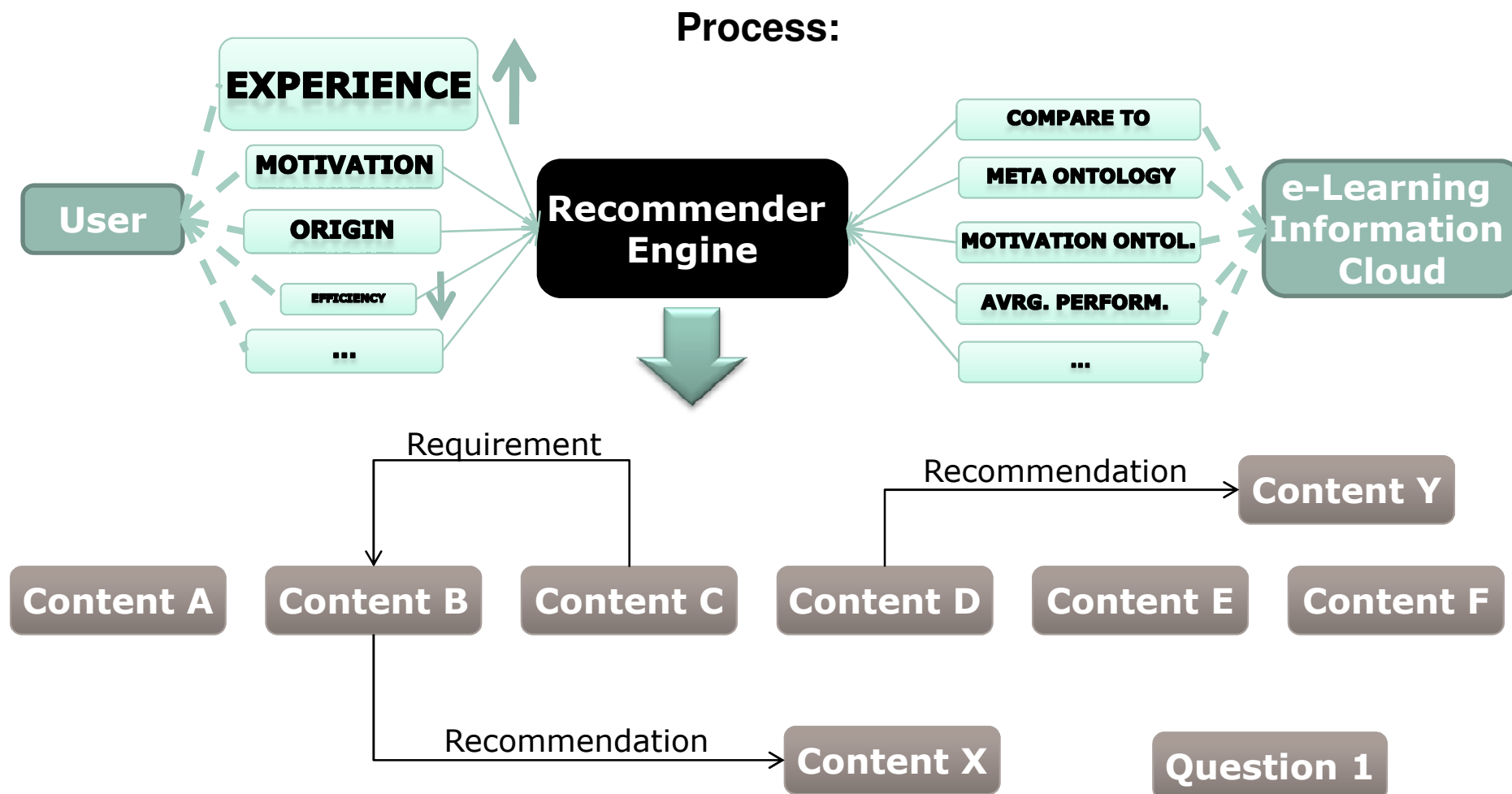


Example: Semantic e-learning path recommendation

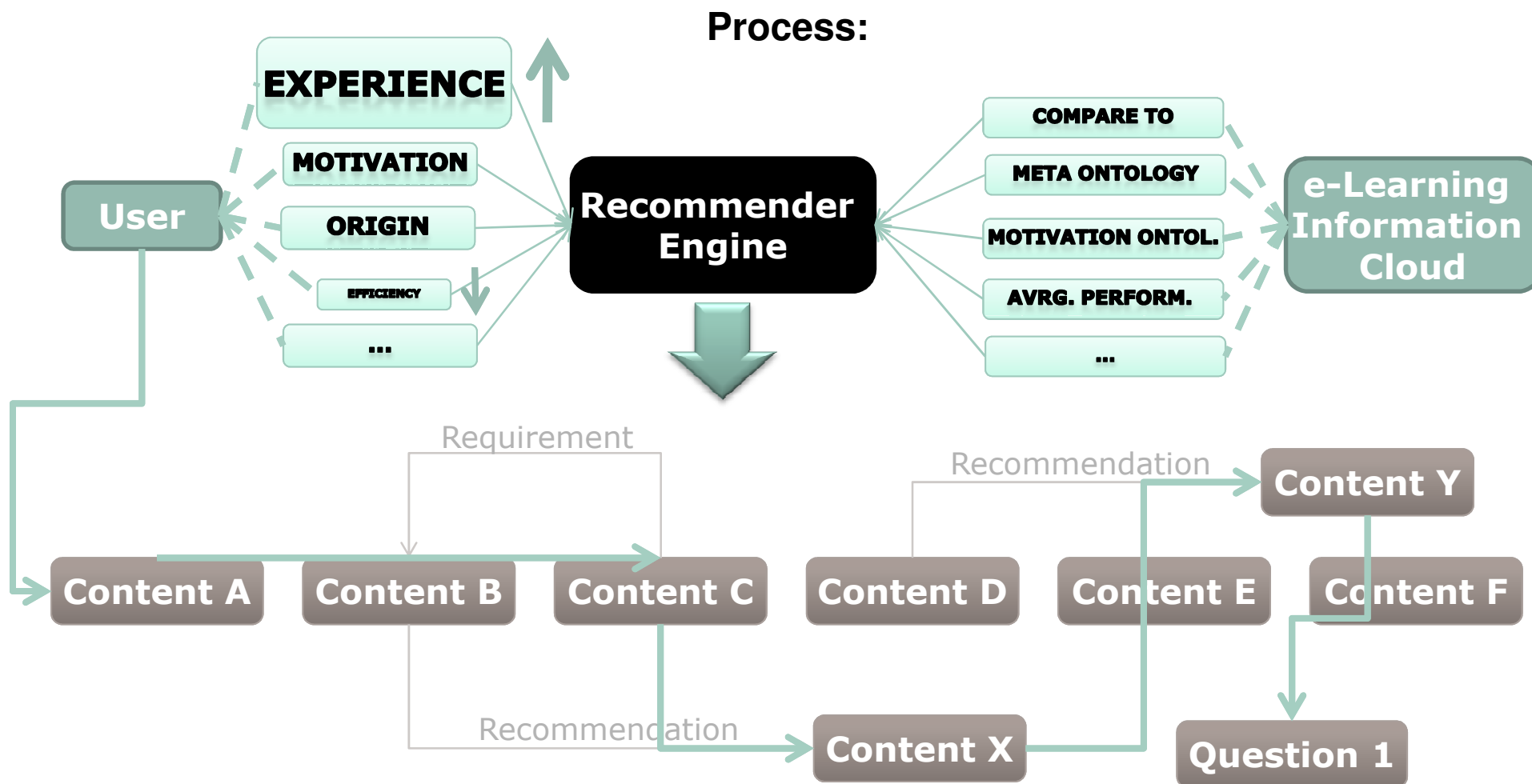
Dynamic, semantic based Teaching Path



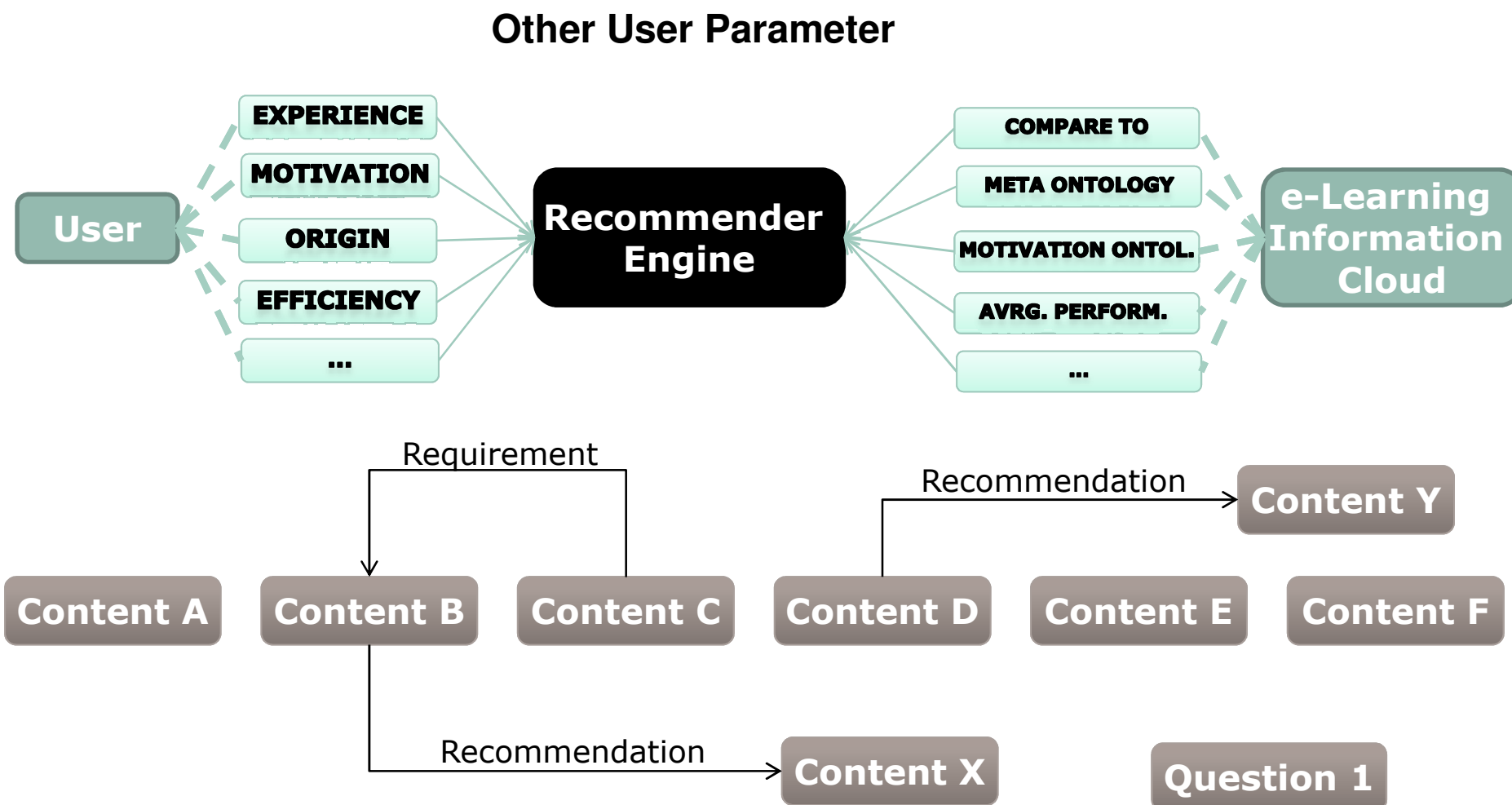
Example: Semantic e-learning path recommendation



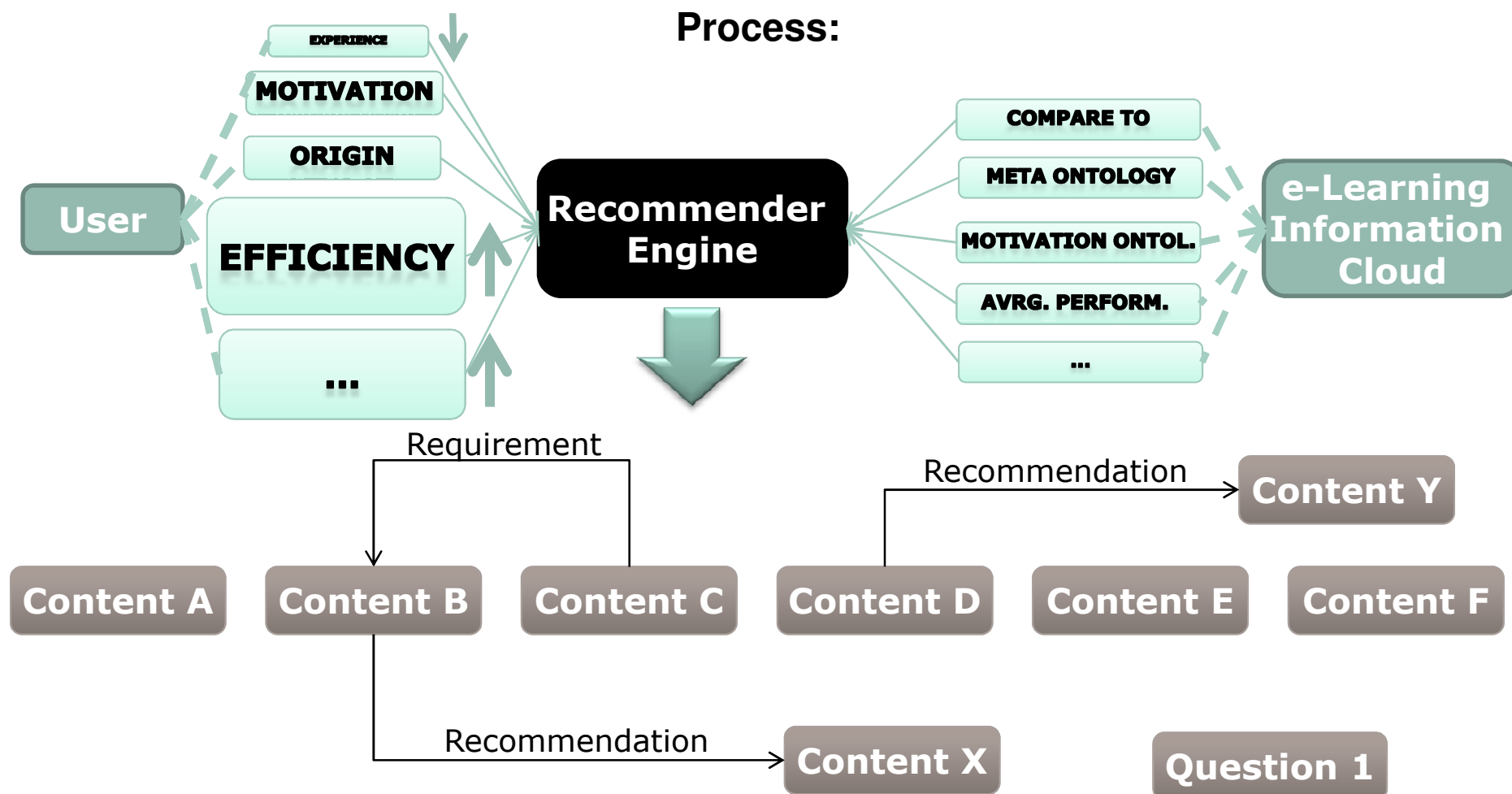
Example: Semantic e-learning path recommendation



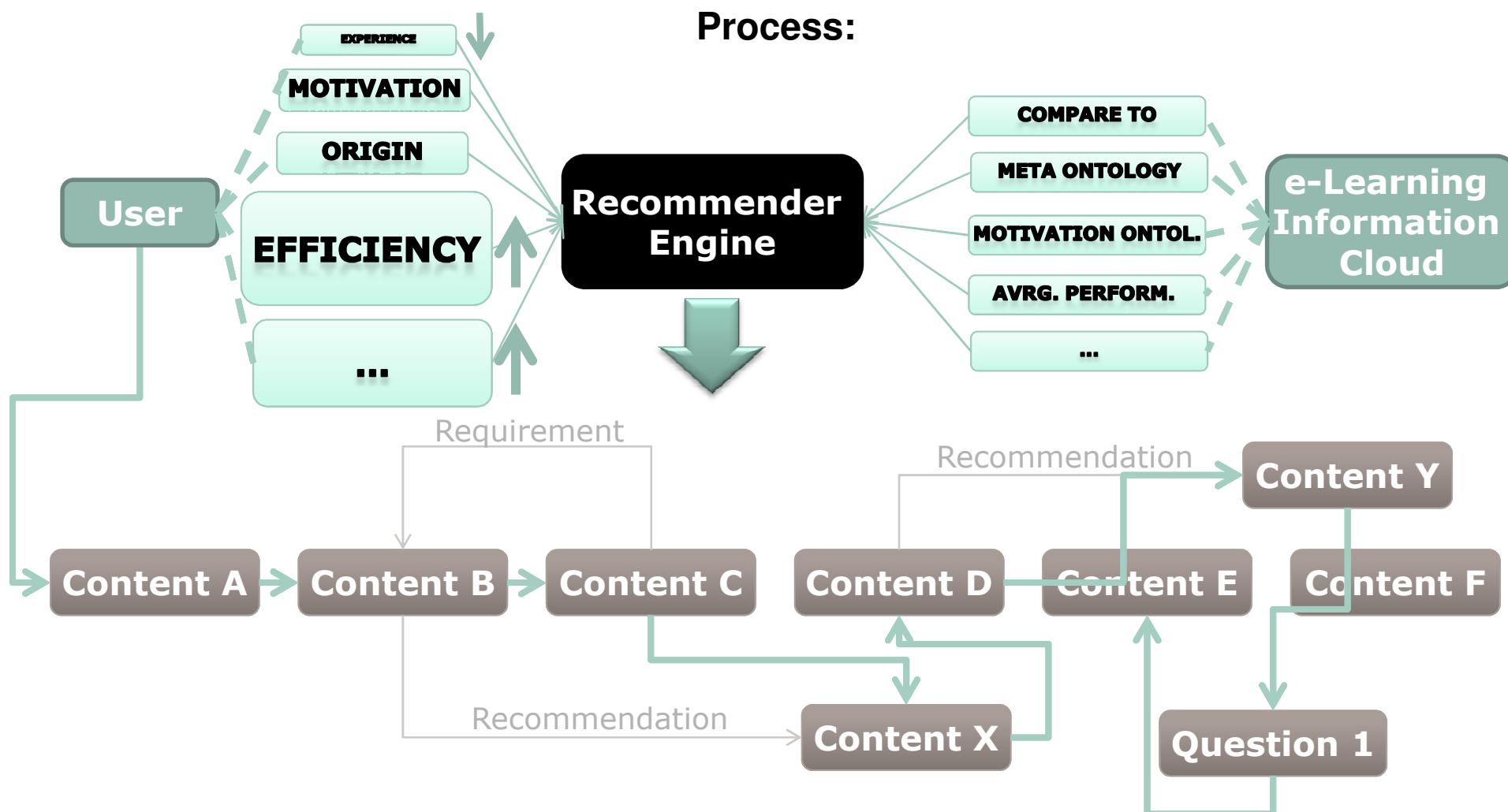
Example: Semantic e-learning path recommendation



Example: Semantic e-learning path recommendation

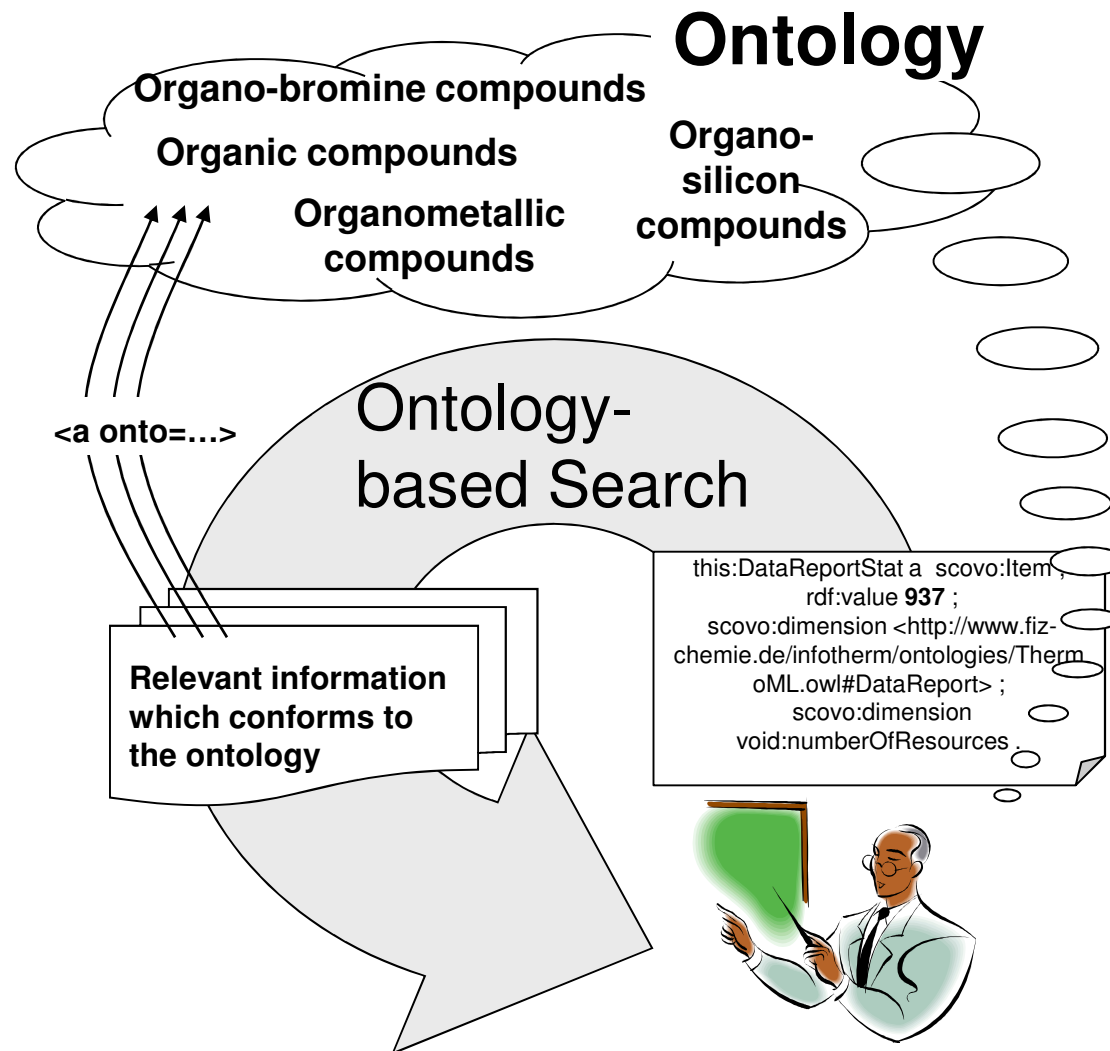


Example: Semantic e-learning path recommendation



Example 2: Thermo-Physical Semantic Search

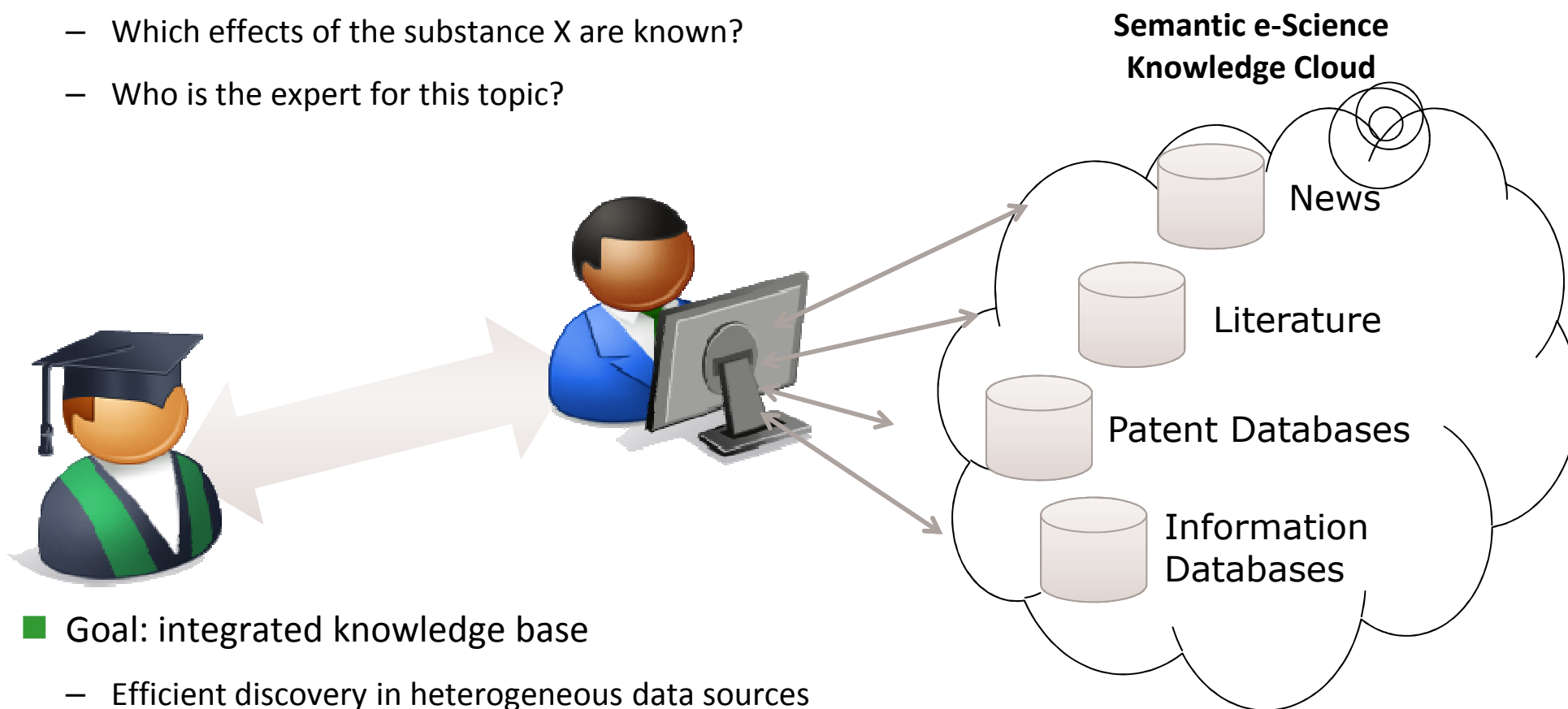
- Ontology representation of Infotherm Thesaurus with classifications from other data sources
 - Add Classification from Chebi Ontology
 - Add Synonyms from Chebi, Pubchem
- Enabling Multilingual Search
 - DBPedia multilingual terms
- Enabling accurate search with chemical identifiers
 - "4095094" , "C H3 Br3 Si" ;
- Adding Semantic Auto Complete
- Adding Semantic Facets



Example 3: e-Science Infrastructure

■ Scientific Question

- Which effects of the substance X are known?
- Who is the expert for this topic?

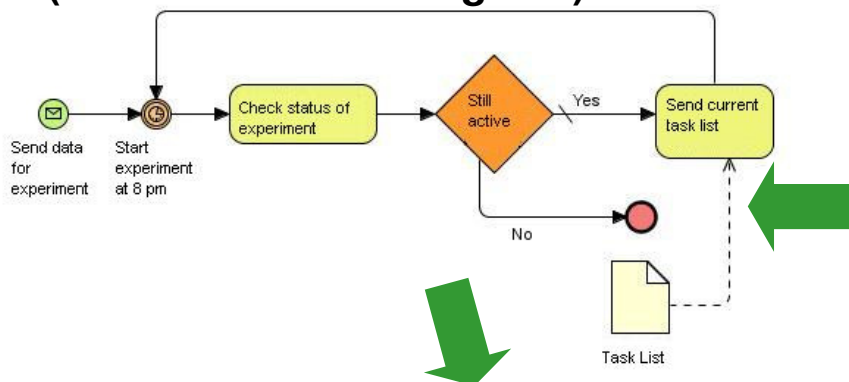


■ Goal: integrated knowledge base

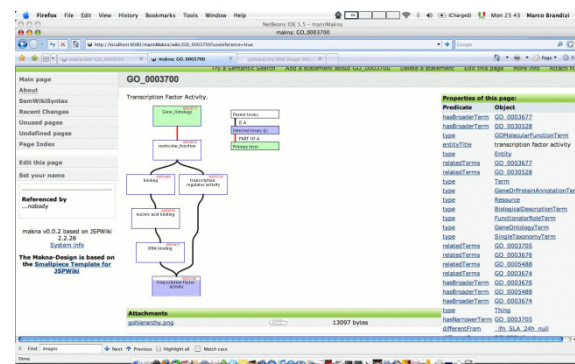
- Efficient discovery in heterogeneous data sources

Example: Knowledge Intensive e-Science Processes

e-Science Processes (event-based rule engines)

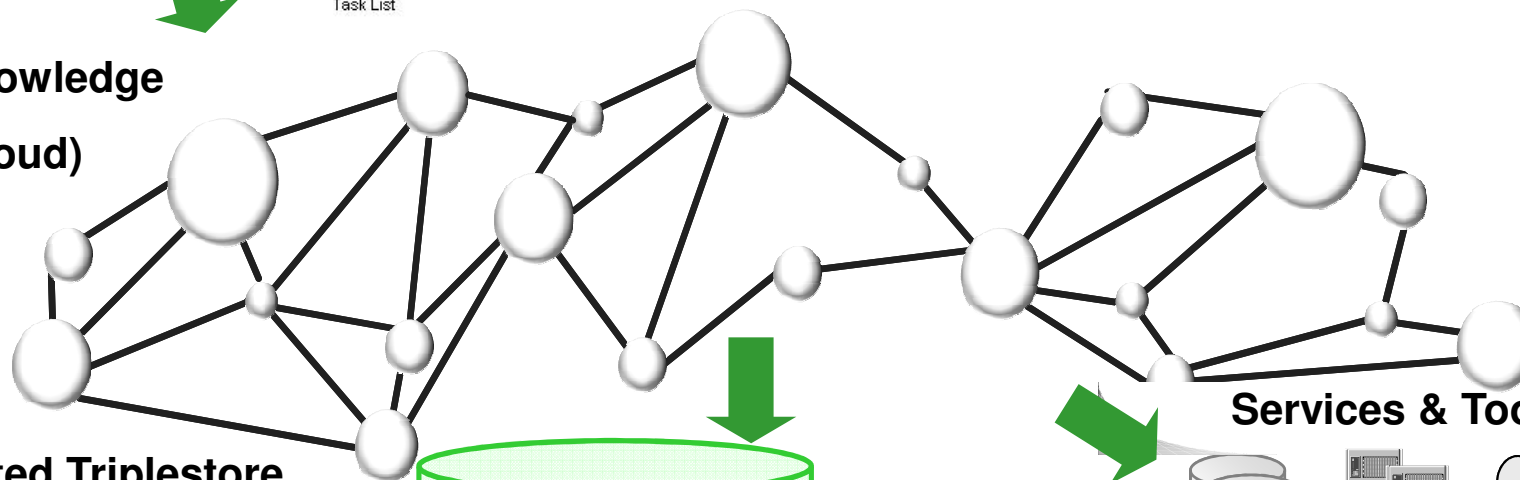


Services, User Interface

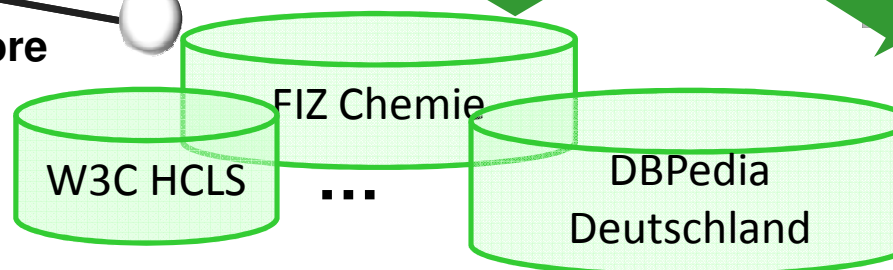


Scientists

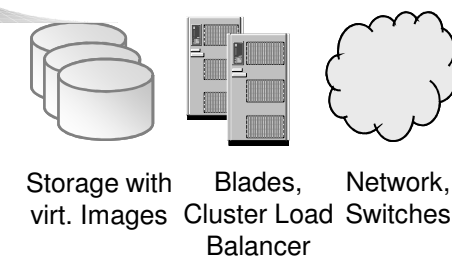
Semantic Knowledge (ChemCloud)



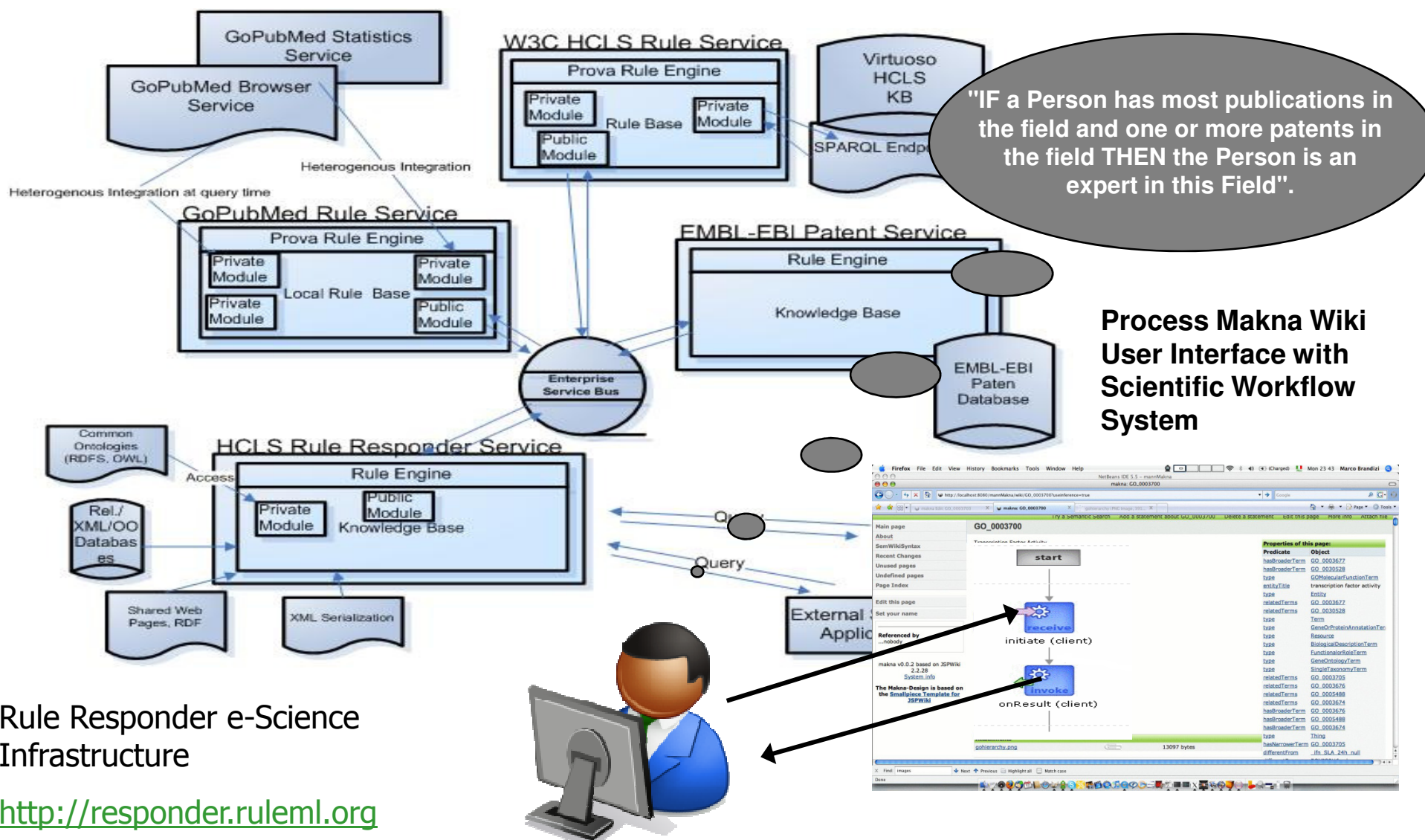
Federated Triplestore (AllegroGraph)



Services & Tools



Example: Rule Responder e-Science Infrastructure



Rule Responder e-Science
Infrastructure

<http://responder.ruleml.org>

Agenda

- ChemCloud – Chemical e-Science Information Cloud
- Semantic Model of the ChemCloud
- Exemplary Use Cases
- **Conclusion**

Summary

■ Goals

- **Improve Data Quality**
 - Synonyms, identifier, multi-lingual, complements, ...
- **Improve Search and Knowledge Management**
 - Semantic context, multi-lingual, faceted search, ...
- **Enabling new Kinds of Queries and Applications**
 - Cross database, cross domain, reasoning

■ Current Implementation

1. Step 1 Conversion

- Reuse and create ontologies for the ChemCloud Semantic Model
- Convert relational and XML data into RDF

2. Step 2 Data Integration

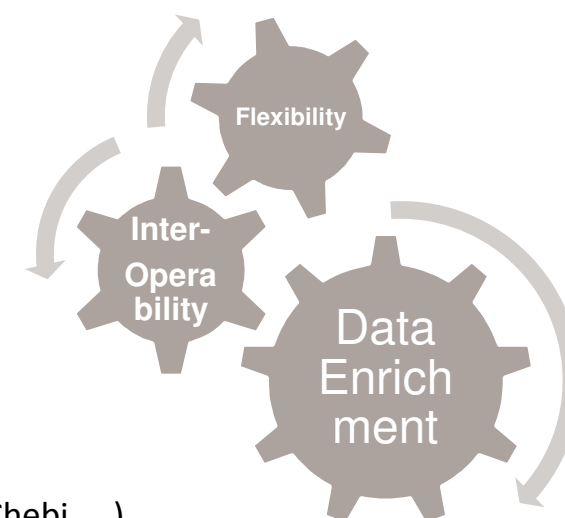
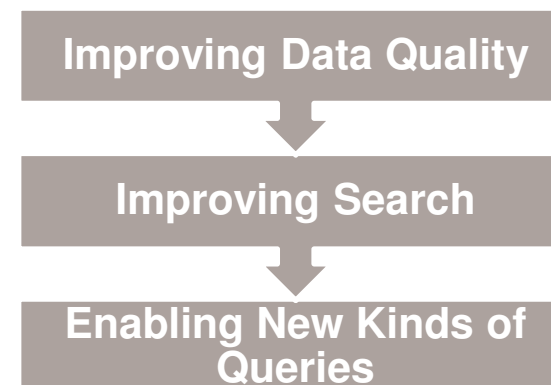
- Linked Data Integration (Chebi, Pubchem)
- Ontology Integration (Bio2RDF, Chem2Bio2RDF)

3. Step 3 Applications

- Semantic Faceted Search (Semantic Browser; Drupal and Exhibit)
- Semantic e-Book Recommender and e-Learning
- Semantic e-Science Scientific Workflows (Rule Responder)

■ Future Steps

- More data and improved data quality in ChemCloud
- Better integration of existing data and ontologies from others (ChemAxiom, Chebi, ...)
- Integration into Linked (Open) Data Cloud and tools/applications



Questions



<http://www.chemistry.de/>

<http://www.corporate-semantic-web.de>