Semantic Management in PESCaDO

Marco Rospocher, FBK
rospocher@fbk.eu :: http://dkm.fbk.eu/rospocher

Joint work with:
Emanuele Pianta, Luciano Serafini, Sara Tonelli

Overview

Part 1: Ontology-based Decision Support

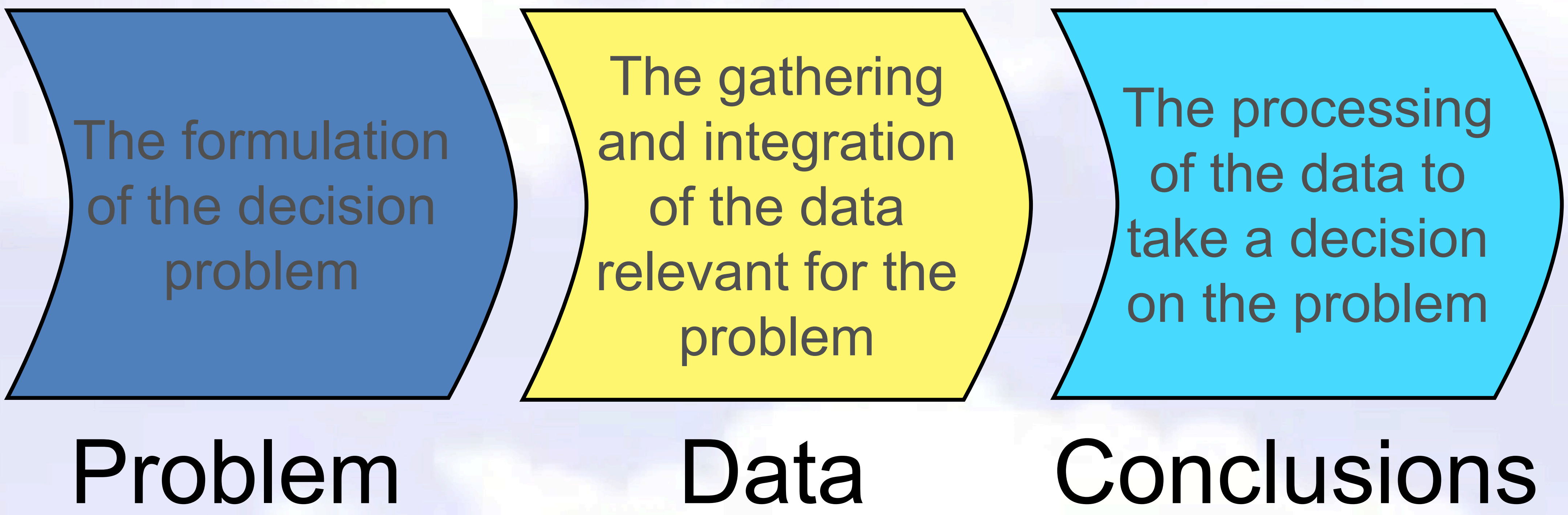
Part 2: Key-concept Extraction for Ontology Engineering

Part 1

ONTOLOGY-BASED DECISION SUPPORT

Decision Making

- The decision making process of a Decision Support System (DSS) typically consists of three phases:



PESCaDO Approach

- We propose to adopt an ontology-based knowledge base as the main (enhanced) data structure of the DSS:
  - T-Box: formally represents the content manipulated in the three decision-making phases (problem, data, conclusions)
  - A-Box: each request submitted to the system corresponds to a single incrementally-built A-Box (a "semantic request script")

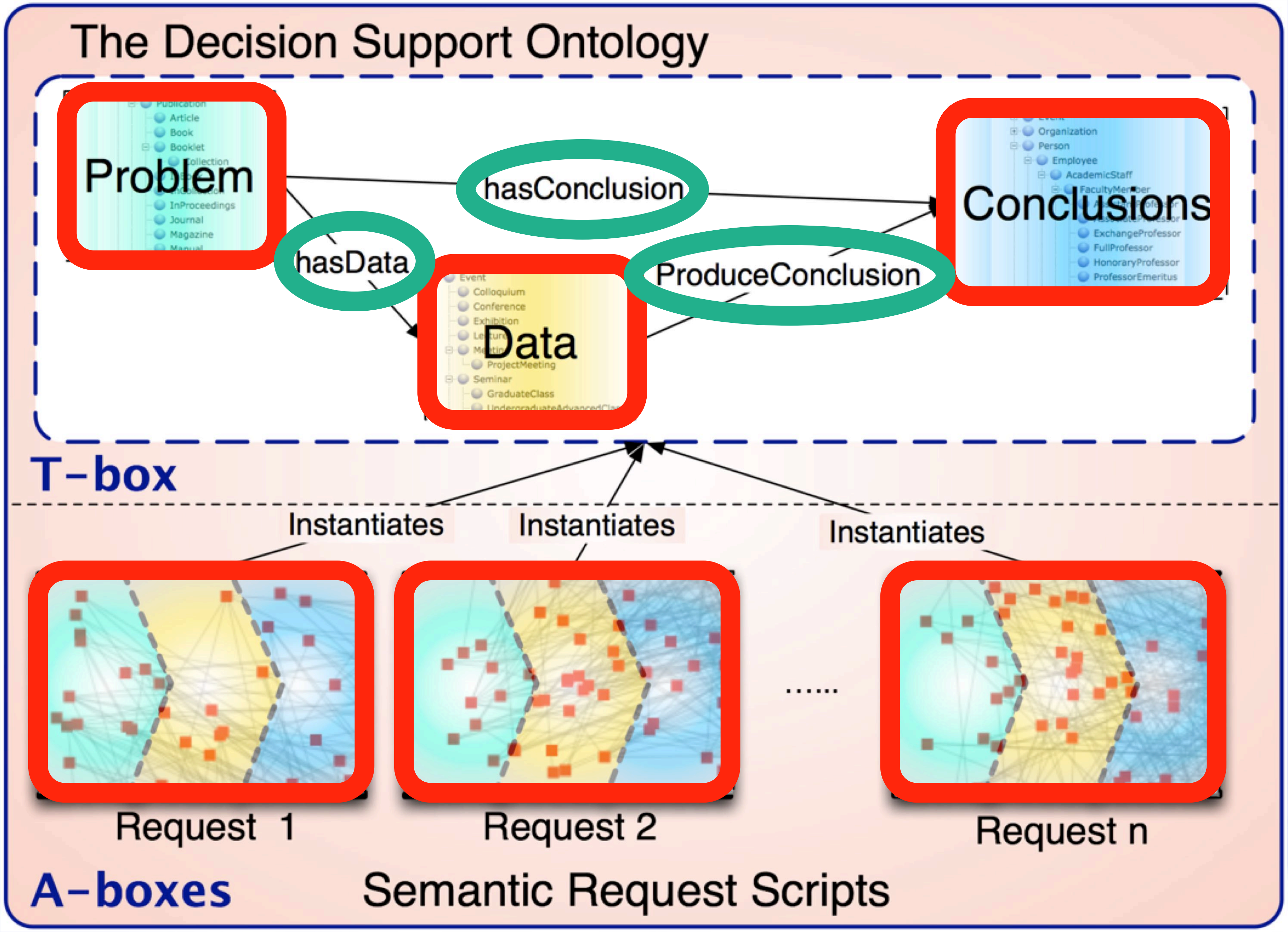
Advantages

- Facilitates the integration of heterogeneous knowledge and data sources
- Semantic exposure of DSS processing to other services
- Some of the inference steps of the DSS can be performed via state of the art logical reasoning services

Outline of First Part

- The Decision Support Knowledge base (DSKB)
  – Problem component
  – Data component
  – Conclusion component
  – Semantic Request Script (SRS)
- Incremental construction of a SRS
- Exploitation of SRSs
- On Engineering the DSKB
- Conclusions

The Decision Support Knowledge Base



The Problem Component

- Formally describes all the aspects of decision support problems that the user can submit to the DSS
- Examples of content:
  - taxonomy of the request types supported by the system
  - input parameters needed by the DSS to provide adequate decision support
  - users profile
  - ...
- May also be used to dynamically constrain the user input in the DSS User Interface

PESCaDO

SEVENTH FRAMEWORK PROGRAMME

The PESCaDO Problem Component

- Request
  - InstructionRequest
    - SuggestAdministrativePlan
  - ReportRequest
    - CheckAirQualityLimits
    - CheckBlackIceCondition
    - CompareAirQualityInMultipleRegions
    - ReportAirQualityForecast
  - WarningRequest
    - AnyHealthIssue
    - AnyRestrictionForPrivateTransport
    - WarningDueToEnvironmentalConditions

- Activity
  - AttendingOpenAirEvent
  - LongTermStaying
    - GoingOnHolidayLongTermStaying
    - LivingLongTermStaying
  - PhysicalOutdoorActivity
  - Travelling
    - BikeOrFeetTravelling
      - FeetTravelling
      - BikeTravelling
    - CarTravelling
    - PublicTransportTravelling

- User
  - AdministrativeUser
  - EndUser
    - AdultUser
    - ChildUser
    - ElderlyUser
    - InfantUser
    - PregnantFemaleUser
    - UserSensitiveToAirPollutant
    - UserSensitiveToPollen
      - UserSensitiveToAlderPollen
      - UserSensitiveToBirchPollen
      - UserSensitiveToGrassesPollen
      - UserSensitiveToMugwortPollen
    - UserSensitiveToWeather
    - UserSufferingOfAllergicRhinitis
    - UserSufferingOfCirculatoryDisease
    - UserSufferingOfNasalOrEyeAllergy
    - UserSufferingOfRespiratoryDisease
    - YoungUser
  - Expert

The Data Component

- Formally describes the data accessed and manipulated by the DSS
- An ontology to be used as data component may be already available in the web
- It favors the integration of (structured) data provided by heterogeneous sources (web-sites, LOD)

SEVENTH FRAMEWORK PROGRAMME

The PESCaDO Data Component

- It describes environmental related data:
  - meteorological data (e.g., temperature, wind speed)
  - pollen count data

```
EnvironmentalData
    EnvironmentalData SubClassOf hasFromDateTime some dateTime
    EnvironmentalData SubClassOf hasEnvironmentalDataNature exactly 1 EnvironmentalDataNature
    EnvironmentalData SubClassOf hasEnvironmentalDataEnvironmentalDataType exactly 1 EnvironmentalDataType
    EnvironmentalData SubClassOf hasToDateTime some dateTime

EnvironmentalNode
    EnvironmentalNode SubClassOf hasEnvironmentalNodeLocation max 1 Location
    EnvironmentalNode SubClassOf hasEnvironmentalNodeEnvironmentalNodeAreaType max 1 EnvironmentalNodeAreaType
    EnvironmentalNode SubClassOf hasEnvironmentalNodeName exactly 1 string
    EnvironmentalNode SubClassOf hasEnvironmentalNodeForm exactly 1 EnvironmentalNodeForm
    EnvironmentalNode SubClassOf hasEnvironmentalNodeEnvironmentalNodeType max 1 EnvironmentalNodeType
    EnvironmentalNode SubClassOf hasEnvironmentalNodeConfidenceValue max 1 double
    EnvironmentalNode SubClassOf hasEnvironmentalNodeEnvironmentalData only EnvironmentalData
    EnvironmentalNode SubClassOf
        hasEnvironmentalNodeEnvironmentalNodeSourceOfEmissionType max 1 EnvironmentalNodeSourceOfEmissionType
    EnvironmentalNode SubClassOf hasEnvironmentalNodeURL max 1 anyURI
    EnvironmentalNode SubClassOf hasEnvironmentalNodeEnvironmentalNodeLandUseType max 1 EnvironmentalNodeLandUseType
```

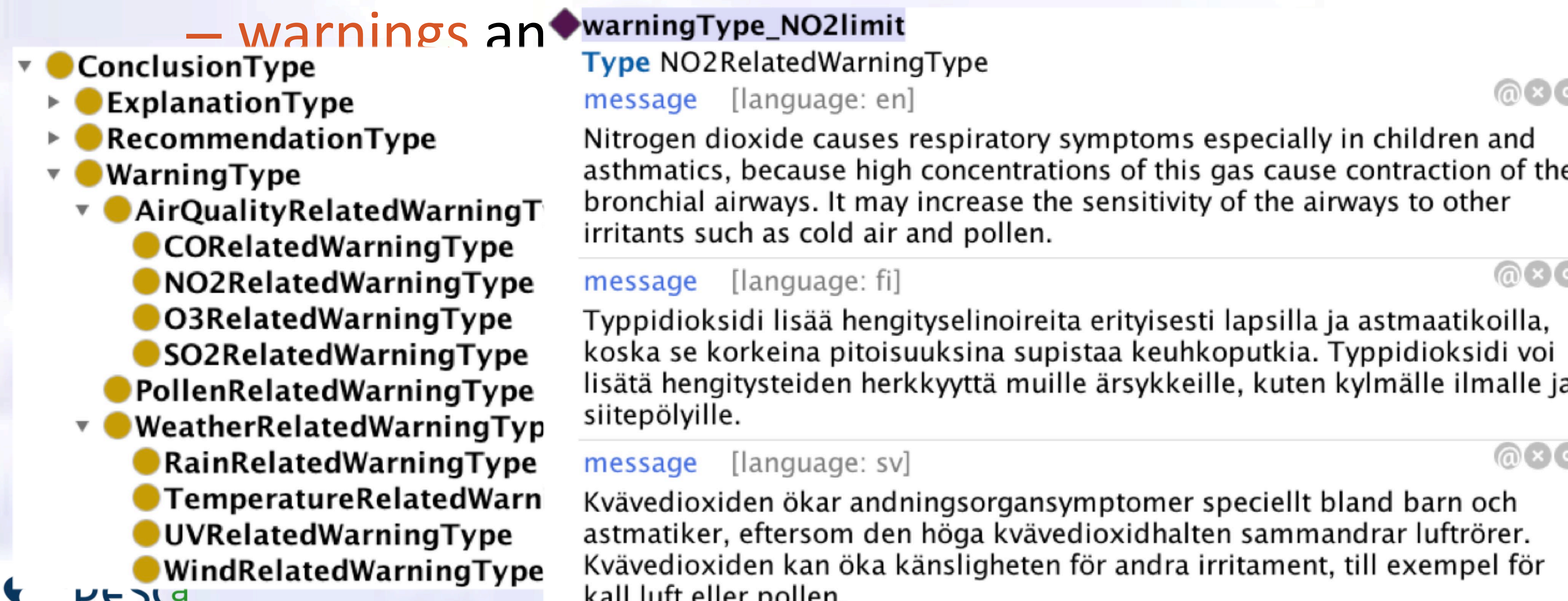
- It facilitated the integration of data obtained from heterogeneous sources, and with different techniques
  - e.g. content distillation from text and images

The Conclusion Component

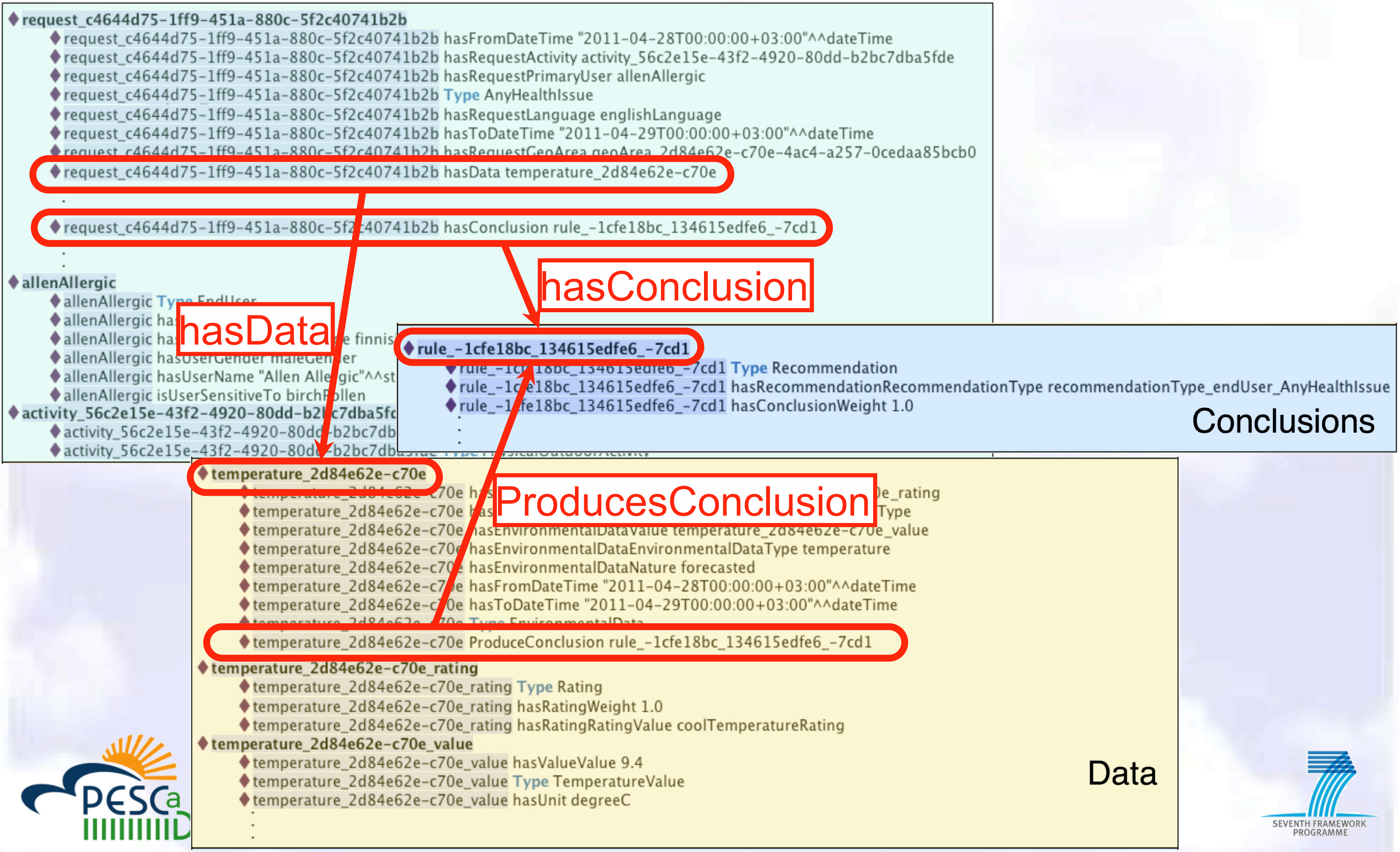
- Formally describes the output produced by the DSS by processing the problem description and the data available, e.g.
  - warnings/suggestions/instructions/decisions
  - data aggregations, data analysis results
- A weight (e.g. confidence, relevance) may be assigned to the conclusions produced
- Tracking of the data that triggered conclusions ("ProduceConclusion" object property)
- User feedback (degree of satisfaction) may also be included

The PESCaDO Conclusion Component

- It describes conclusion types like
  - exceedances of air pollutants limit values detected from data
  - warnings an



SRS: An A-Box of the DSKB



Incrementally building SRSs

Exploitation of Logical Reasoning

- Phase1: Instantiation of the problem
  - consistency check to verify that the user request is compliant with the problem supported by the DSS
- Phase2: Instantiation of the data
  - data relevant for the user problem may be determined via ontology reasoning
    - PESCaDO: using "owl:hasValue" restrictions
      - e.g. userSensitiveToBirchPollen subClassOf RelevantAspect value Rain
- Phase3: Instantiation of the conclusions
  - instantiation depends on the decision support techniques adopted by the DSS
    - PESCaDO: DL+RuleBased+Fuzzy reasoning

Exploitation of SRSs

A SRS provides a complete "semantic" snapshot of all the information processed and produced by the DSS for a request, with "explanations"

- A natural language report can be automatically generated from it
  - especially appreciated by laymen, media corporations, …
- SRSs could be archived in a semantic repository (e.g. Sesame, Virtuoso), incrementally fed
  - fine-tune the decision support strategies implemented in the DSS
  - expose to the world the DSS processing in LOD format, favoring its exploitation by other applications/web-services
  - easily compute relevant statistics

SEVENTH FRAMEWORK PROGRAMME

On Engineering the DSKB

- Checks on the DSKB
  - formal consistency check
  - correct instantiation with the usage in the DSS
- Assessment of the adequacy of the DSKB for the DSS
  - all decision support problems to be supported by the DSS are formally representable in the Problem component
  - all the data relevant for the DSS are characterized in the Data component
  - all the conclusions and explanations to be generated by the DSS are formalized in the Conclusions component
- In PESCaDO:
  - Problem: all the types of problems defined in the use cases can be represented
  - Data: environmental experts assessment (appropriateness: 94% - completeness: 92%)
  - Conclusions: environmental experts assessment (appropriateness: 90% - completeness: 87%)

Conclusions of First Part

- We proposed to adopt an ontology-based knowledge base as the main data structure in DSSs
- Each decision support request submitted to the DSS corresponds a semantic request script which describes
  - the request itself
  - the data relevant for the request
  - the conclusions/suggestions/decisions generated by DSSs
- Demonstrated the advantages in a concrete implementation for an environmental DSS (PESCaDO EU project)
  - integration of heterogeneous sources of data available in the web (e.g., web sites, web services)
  - tracking and exposure in a structured form of all the content processed and produced by the DSS for each request
  - exploitation of logical reasoning for several of the inference steps of the DSS decision-making process

SEVENTH FRAMEWORK PROGRAMME

Part 2

KEY-CONCEPT EXTRACTION FOR ONTOLOGY ENGINEERING

SEVENTH FRAMEWORK PROGRAMME

Automatic Concept Extraction

- Support ontology modeling by extracting concepts characterizing a domain from a reference text corpus.

- Automatic concepts extraction plays an important role in ontology modeling:
  - To boost the ontology construction/extension phase;
  - To "validate" an ontology against a domain corpus.

Our Contribution

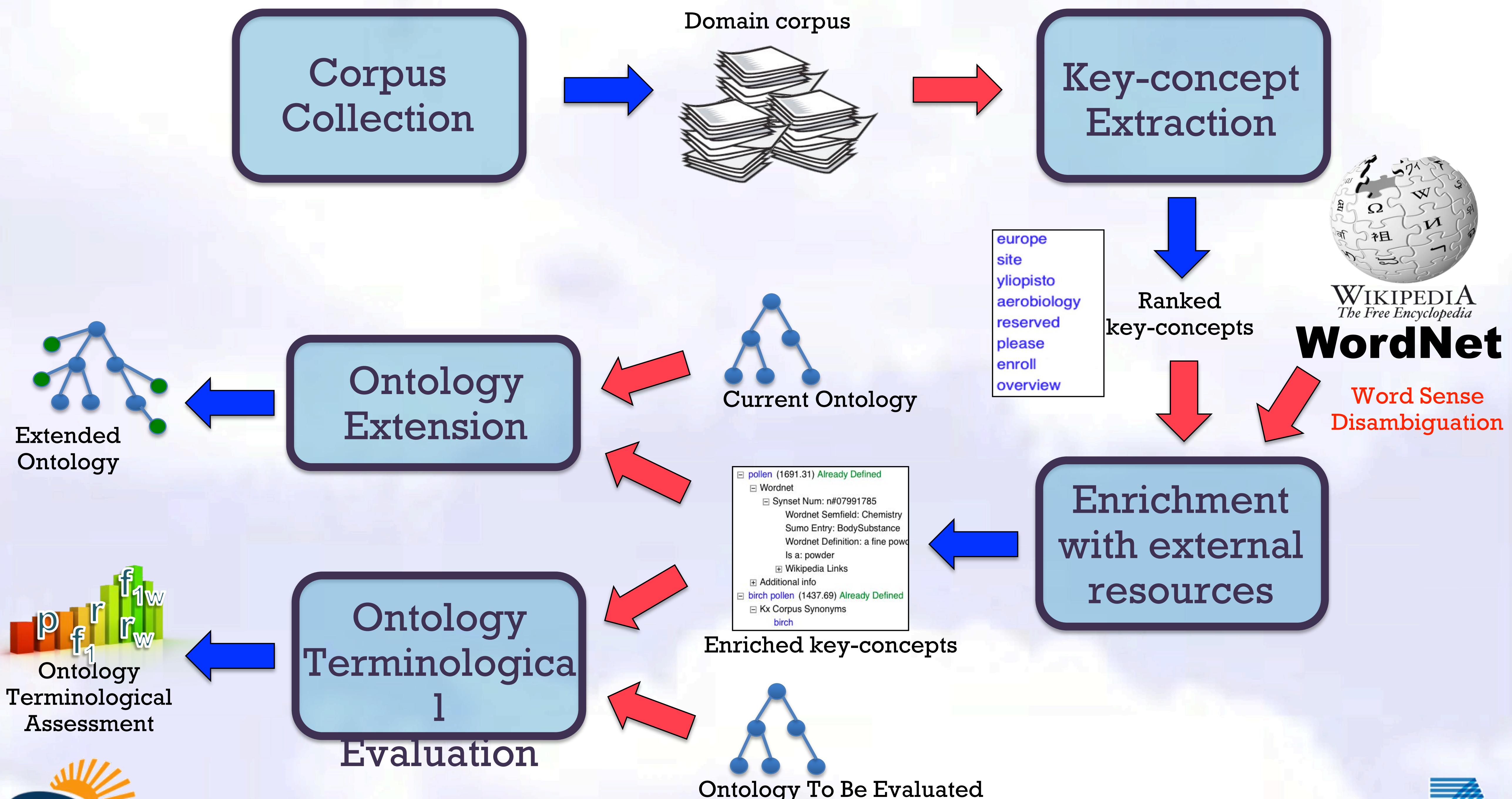
- A framework for supporting ontology engineering by automatic concept extraction from a reference text corpus

- A fully-working and publicly available implementation of the proposed framework

Outline of Second Part

- The Framework
- Implementation of the Framework
- Evaluation
- Usage in PESCaDO

Key-concept Extraction For Ontology Engineering



Corpus Collection

- The corpus can be manually or automatically collected (e.g. crawling web pages).

- Corpus could consist of:
  - (large) collection of documents
    - e.g. pollen bulletins crawled on-line
  - A single big document
    - e.g. the BPMN specification.

Key-concept extraction

- Performed by KX (Keyphrase eXtraction) tool.
  - exploits linguistic information and statistical measures to select a list of weighted keywords from documents;
  - handles multi-words;
  - flexible parameters configuration;
  - easily adaptable to new languages, available for English, Swedish, Finnish, French and Italian;
  - ranked 2nd (out of 20) at SemEval2010, task on "*Automatic Keyphrase Extraction from Scientific Articles*".

SEVENTH FRAMEWORK PROGRAMME

Enrichment with external resources

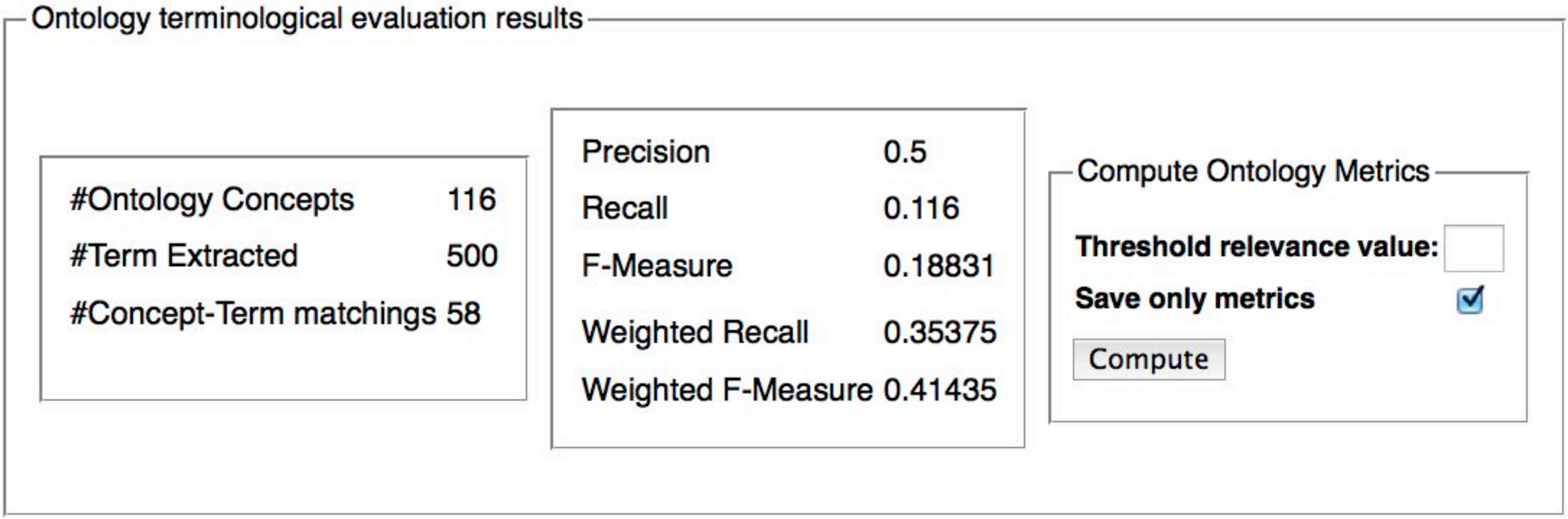
Concepts extracted (Ordered by Relevance)	Relevance	100% matching	Synonym 100% matching
▶ activity	1.00000	X	
▶ attribute	0.88020		
sequence flow	0.71714	X	
▶ business process modeling notation	0.70216		
▼ task	0.49418	X	
▼ Wordnet			
▼ Synset_#00795720			
Wordnet Definition: any piece of work that is undertaken or attempted			
Is a: work			
Sumo Entry: IntentionalProcess			
▼ Synonyms			
undertaking			
project			
labor			
Hyponims: cinch, breeze, picnic, snap1, duck soup, child's play, pushover, walkover, piece of cake, adventure, escapade, risky venture, dangerous undertaking, assignment, baby, enterprise, endeavor, endeavour, labor of love, labour of love, marathon, endurance contest, no-brainer, proposition, tall order, large order, venture, Manhattan Project			
▶ Wikipedia Links			
▶ mapping	0.48253		
▶ flow	0.47920		

Ontology Extension

- Enriched key-concepts list matched against the ontology under development (to detect already defined key-concepts);
- The user decides which of the extracted key-concepts to add to the ontology;
- The additional details provided in the enriched list may guide the formalization;
  - e.g. is-a related synsets, definitions, …

Ontology Terminological Evaluation

- Evaluation metrics are computed on the matching



  - F1 ≥ 0.15 or
  - weighted F1 ≥ 0.25

MoKi

the Modelling WiKi ---

- Collaborative wiki-based tool for modeling (integrated) ontologies and business processes;

- Supports an agile collaboration between domain experts and knowledge engineers via multi-mode knowledge access modalities;

- Offers several different functionalities:
  - Import/export of formal models;
  - Views on the is-a hierarchy and processes decomposition;
  - Graphical editing.

- **Available @ http://moki.fbk.eu**

PESCaDO

SEVENTH FRAMEWORK PROGRAMME

PESCaDO Ontology Construction

- Developed in PESCaDO to support the construction of an ontology describing the environmental domain.

- Corpus: plain text corpus composed of 390 pollen bulletins (541,000 tokens).

- The system outputted 91 key-concepts:
  - 26 pollen names (further validated against the Pollen Atlas);
  - 38 key-concepts enriched with additional information;
  - Extracted key-concepts having up to 4 tokens:
    - e.g. “oil seed rape pollen”.

Conclusions of Second Part

- We presented a framework for ontology building/validation based on automatic concept extraction;
- Fully-implemented in a working system;
- Approach evaluated in PESCaDO (environment) and other domains (e.g. business processes);
- Current/Future works:
  - Extend to consider other ontological knowledge (e.g. is-a relations defined in the corpus).

QUESTIONS?

Marco Rospocher, FBK
rospocher@fbk.eu :: http://dkm.fbk.eu/rospocher

PESCaDO

SEVENTH FRAMEWORK PROGRAMME