



**Technische Hochschule  
Brandenburg**  
University of  
Applied Sciences  
**Fachbereich  
Wirtschaft**



# Enterprise Knowledge Graphs & Related Technologies

## An Engineering Case Study around EduGraph

Prof. Dr. Vera G. Meister, Jonas Jetschni  
Dnipropetrovsk & Katowice, April 2017



# Agenda

- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph





## Problem Domain: Decision Support for Specific Study Programs

### (1) Decisions to be supported

- Choosing a proper university for getting a qualification for a favored job profile

### (2) Localization:

- Universities of Applied Sciences (UAS) in the DACH region (Germany, Austria and Switzerland)

### (3) Specification of Study Programs:

- Business and Information Systems Engineering (BISE) at UAS

### (4) Problem Owners:

- AKWI - Association for BISE Study Programs at UAS in the DACH region - organized under the umbrella of German Society of Computer Sciences
- TYPO3 Academic Alliance – an interest group of German universities using the popular Content Management System TYPO3



# (1) Decision Support for Choosing Study Programs

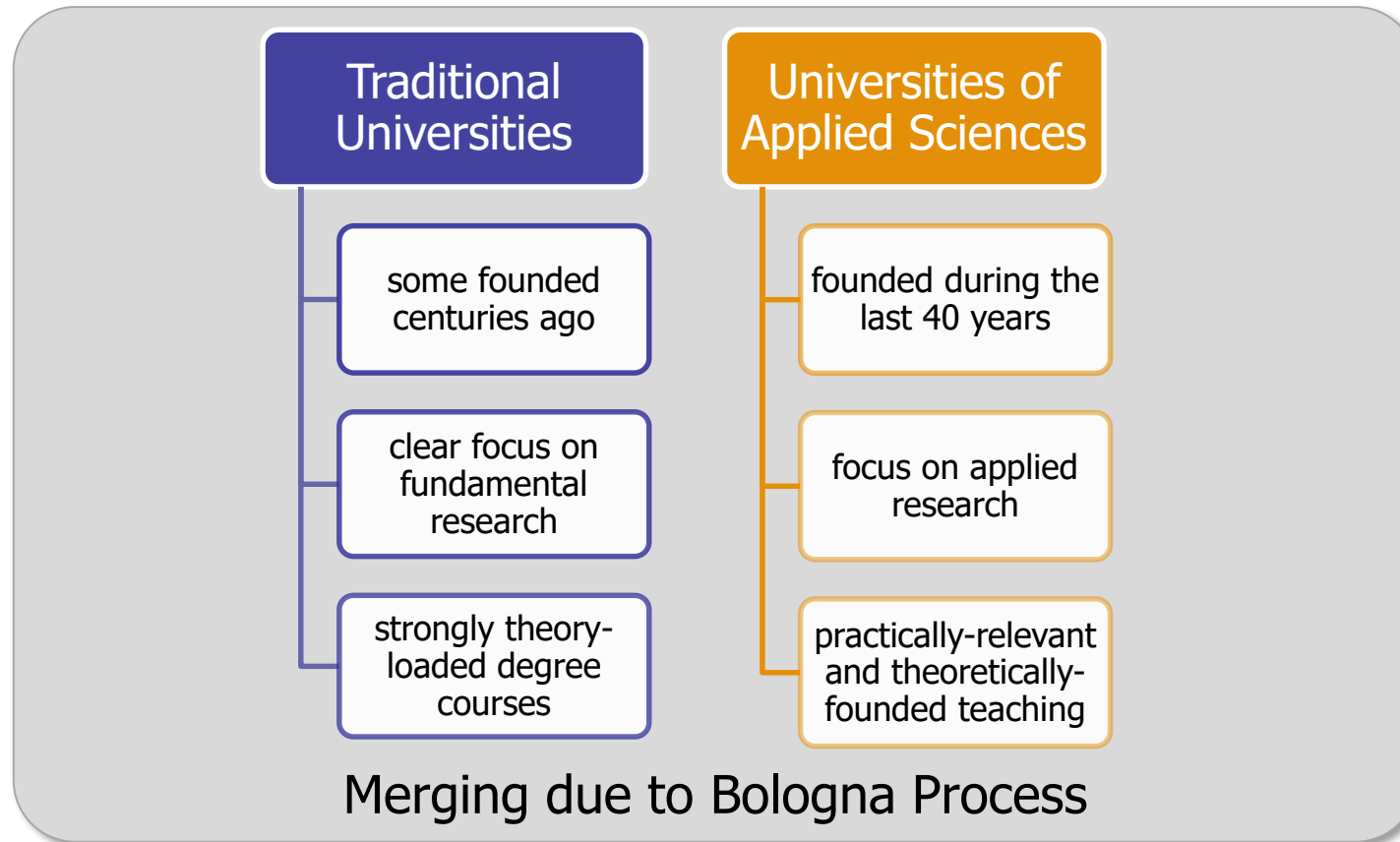
The screenshot shows the STUDYSEARCH website interface. On the left is a map of Europe with several blue location pins. On the right is a green sidebar titled 'Liste der Hochschulen'. It includes a 'Filtern nach' section with a list of job profiles: Administration, Consulting (checked), Informatik, IT-Management, and Softwareentwicklung. Below this is a 'Säulen' section with three categories: Betriebswirtschaftslehre (orange square), Informatik (blue square), and Wirtschaftsinformatik (red square). An orange arrow points from the 'Consulting' checkbox to the 'Säulen' section. The top navigation bar contains links for 'STARTSEITE', 'STUDIENGÄNGE', and 'IMPRESSUM'.





## (2) System of Higher Education in the DACH region

### Two Types of Institutions

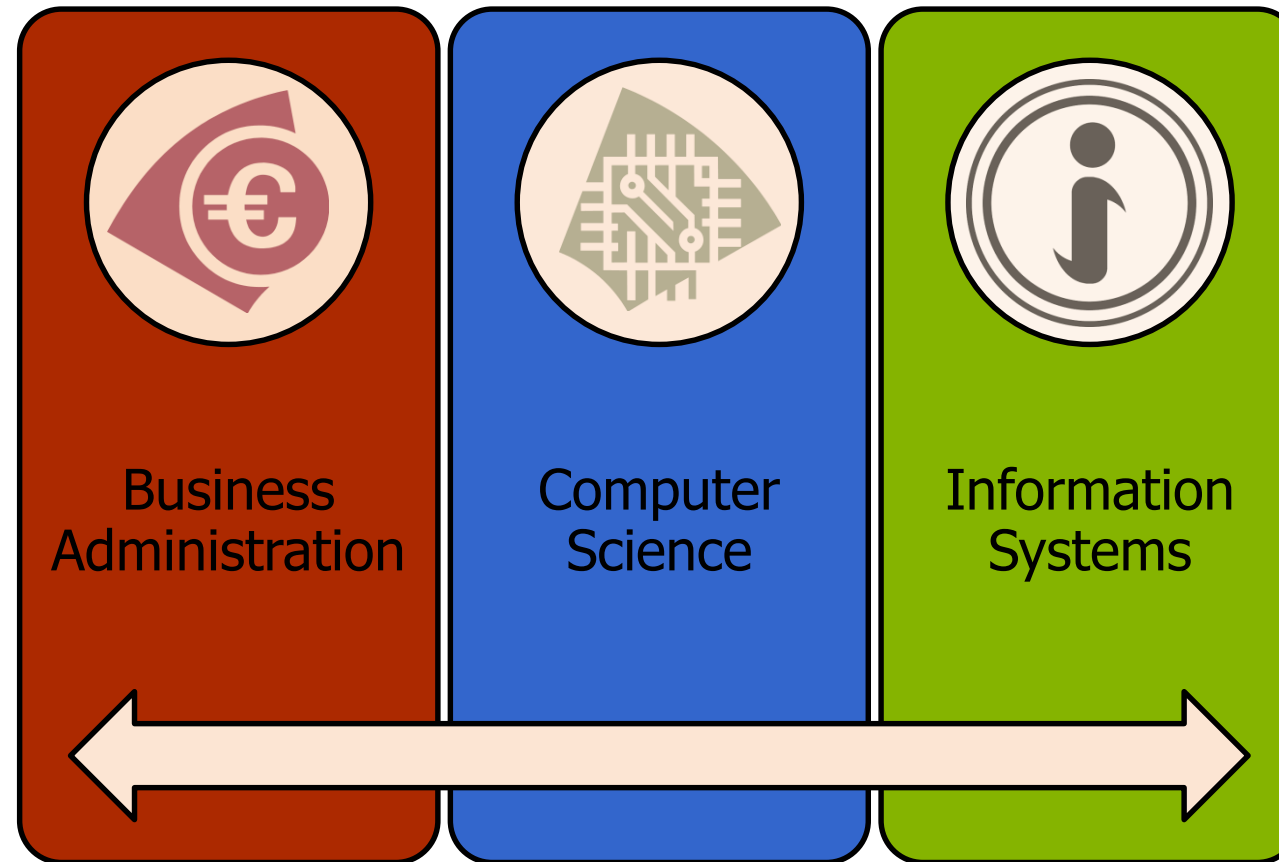




### (3) Specification of Study Programms



#### Business & Information Systems Engineering



#### Three Main Pillars ...

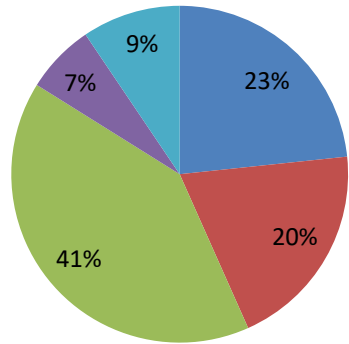
according to the recommendation of the BISE working group of the German Association of Computer Science (GI e. V.)

In addition, there should be a fourth pillar of general and methodical courses, like mathematics, languages, self-competencies etc.

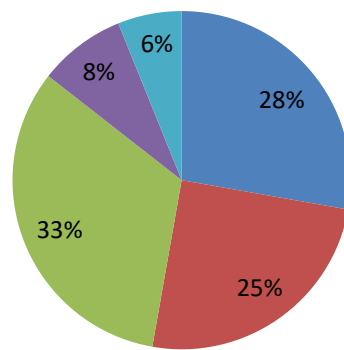


# Distribution of main BISE Pillars in Study Programs at different UAS

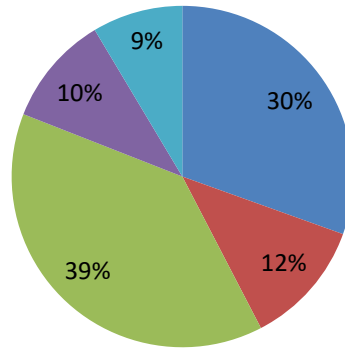
TH Brandenburg



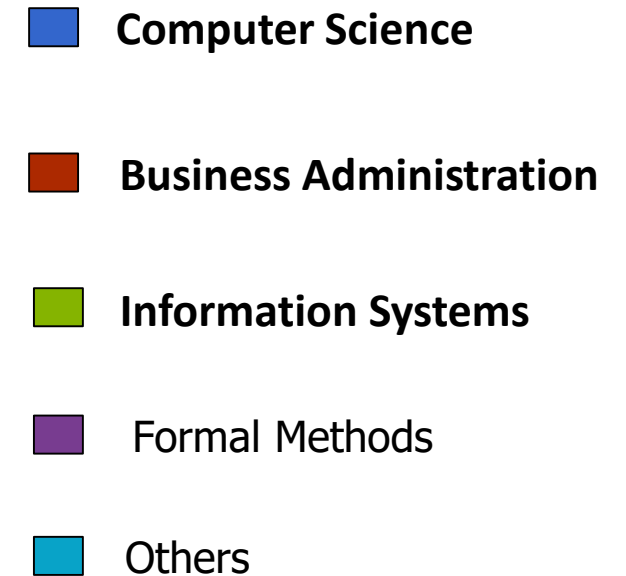
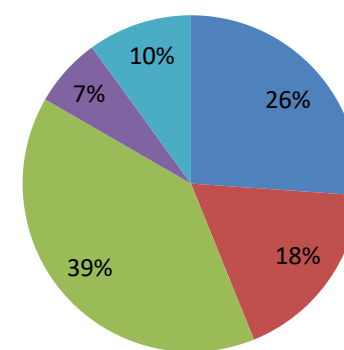
FH Flensburg



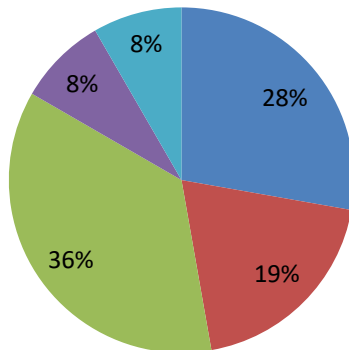
OTH Regensburg



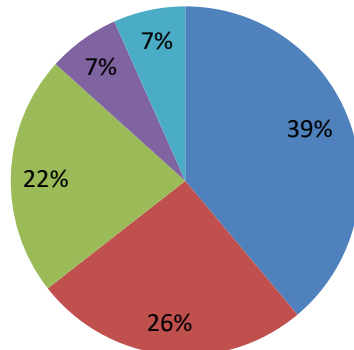
TH Köln



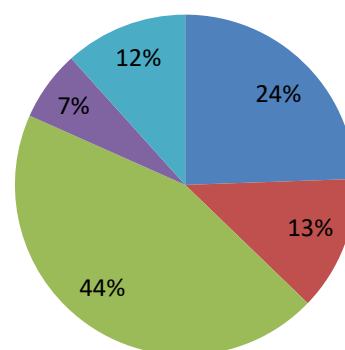
HS Fulda



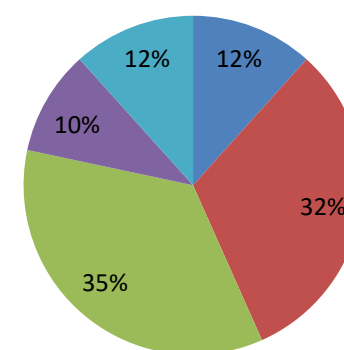
FOM



FH Technikum Wien



ZHAW Zürich



Source: Vera Meister: Wirtschaftsinformatik an Fachhochschulen - Aufbau von Bachelor-Studiengängen, Leitbilder und Status.  
Talk at the annual AKWI conference 2014, Regensburg (Results of a preliminary manual analysis)



## (4) Problem Owners and Authorities

### Relevant Professional Associations and State Institutions

- AKWI - Association for BISE Degree Courses at UAS in the DACH region  
*has primarily identified the need for decision support*
- Subdivision of the German Society of Computer Sciences  
*specifies the main requirements to BISE study programs*
- German Agency for Labor  
*categorizes job profiles for graduates of Computer Sciences and similar study programs*
- TYPO3 Academic Association  
*aims at enhancing the quality and usability of university CMS by achieving a better domain fit*







# Job Profiles for Computer Scientists



## ❖ Informatics

to design, build, install, supervise, or to investigate hardware and software solutions or complex IT systems

## ❖ Consulting

to analyze IT systems; to advise users and customers; to distribute IT products and services

## ❖ Administration

to set up and maintain IT networks; to coordinate and organize the IT of enterprises and organizations; to administer IT systems and Web applications; to configure and administer databases

## ❖ SW Engineering

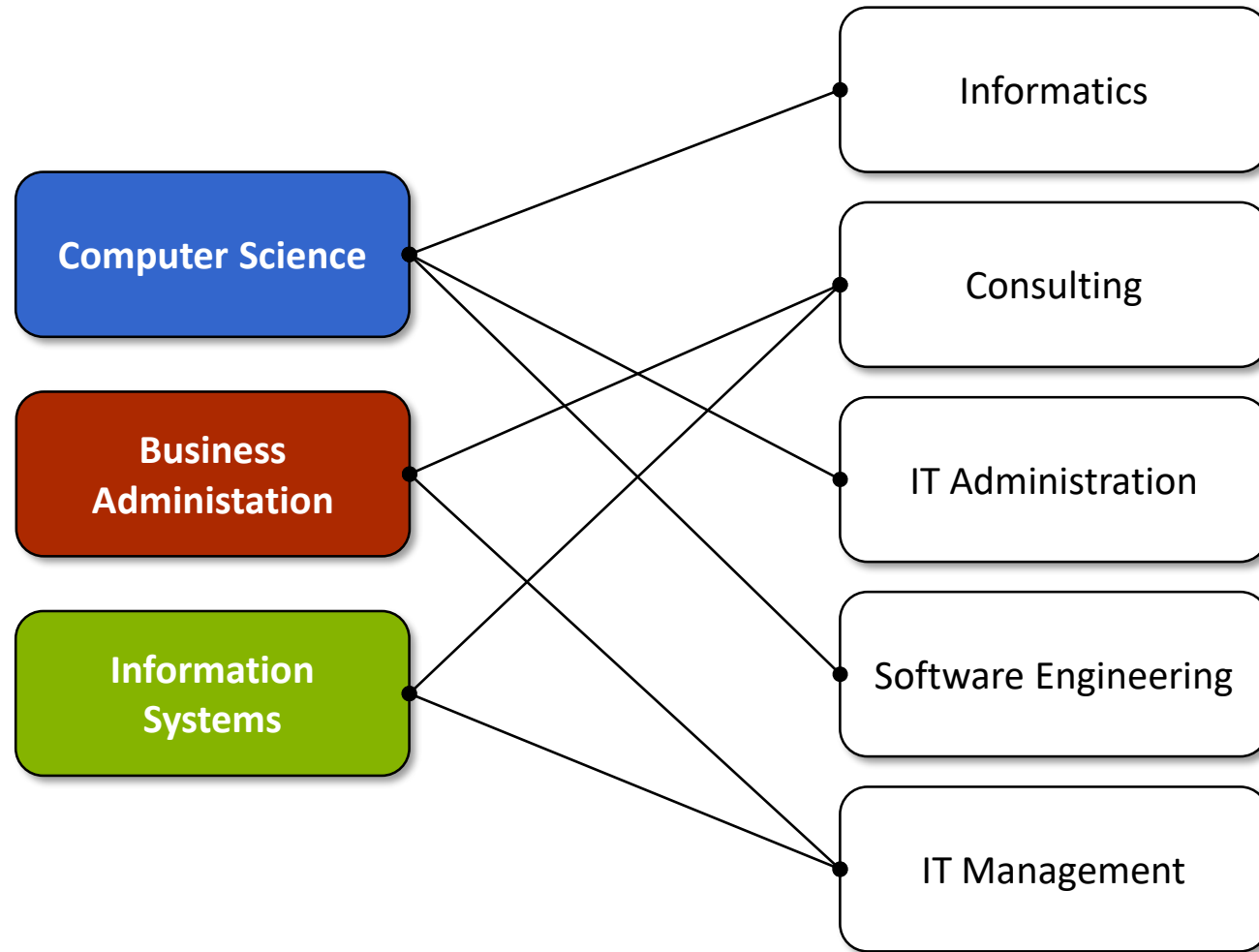
to design, develop and program software

## ❖ IT Management

to direct IT projects, IT departments, or IT workgroups; to develop and implement IT strategies; to be responsible for IT governance



# Matching between Pillars and Job Profiles



**For Graduates in Business &  
Information Systems Engineering**



## Problem Definition: Stakeholder Needs



**BISE degree course:** We want our course offers be found and to be comparable with other degree courses with respect to their structure and content.



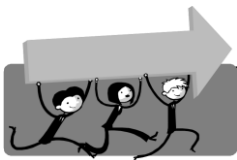
**Prospective BISE student:** I want to study BISE and I want to know which universities of applied sciences offer such courses and what are their focuses.



**Faculty member:** I want to contact a colleague from a university of applied sciences in the federal state XY who is also teaching in a degree course of BISE.



**Enterprise/Organization:** For the reinforcement of our project team we are looking for interns / graduates BISE with particular qualifications.



**Project sponsor/Centre of knowledge transfer:** In the context of a tender for innovation transfer to regional business we are looking for universities of applied sciences with competences in BISE.



# From where to get the Data?

- Study Program Websites in CMS
  - data are stored in silo databases
  - databases use proprietary schemas
  - text crawling produces ambiguous data
- Existing Portals
  - comprises mostly standard formal data
  - provide data via REST APIs (not for free)
  - shall be maintained by universities



- Collect manually in a Business Process
  - costly implementation and maintenance
  - shall be organizationally implemented
  - wide range for individual interpretation





## What do we need?

1. A common schema for the educational domain to annotate and integrate distributed data.
2. A technical solution for automatic schema-based publication of structured data from CMS content.
3. Managed processes for ...
  - a. the extraction and integration of structured data from CMS websites,
  - b. the analysis of unstructured data from texts and the transformation into structured data,
  - c. the enrichment of structured data by harvesting Linked Open Data sources,
  - d. the validation of extracted, integrated and enriched data according to schema constraints.
4. A persistent storage for the structured data including all processual metadata.
5. Technical solutions for providing external knowledge services the access to the database.

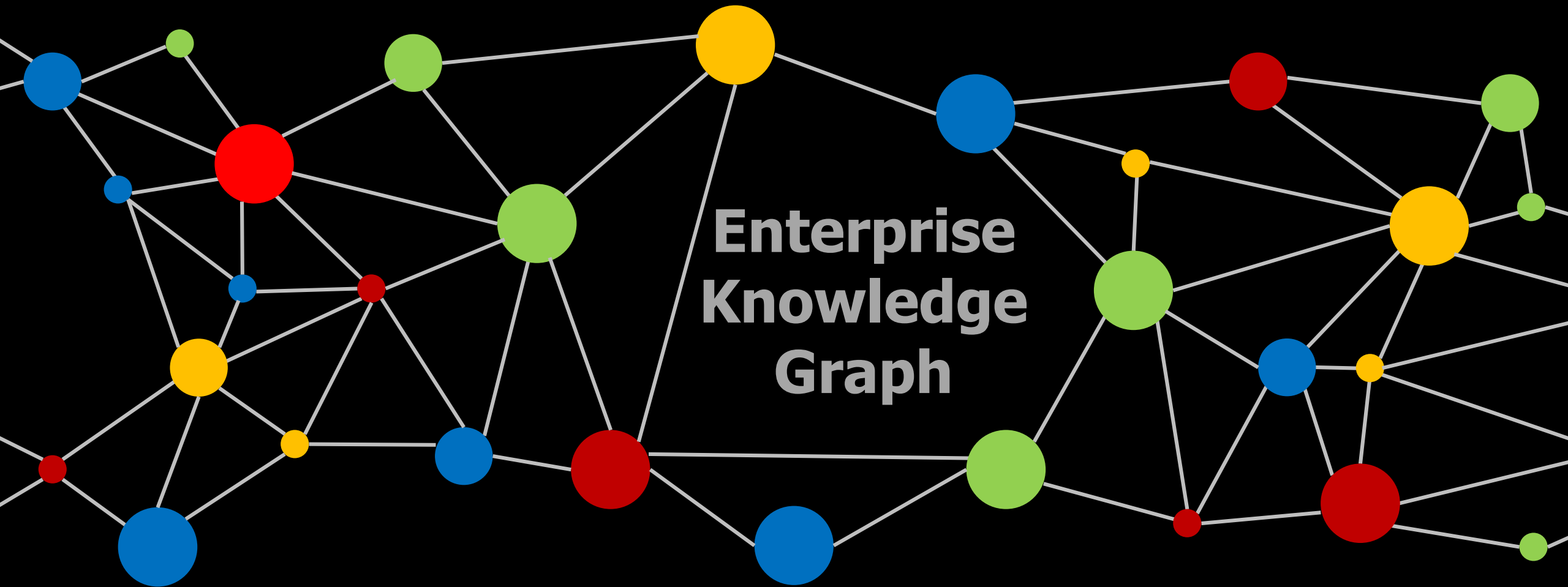
**What we need is an Enterprise Knowledge Graph for the educational domain!**



# Agenda

- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph



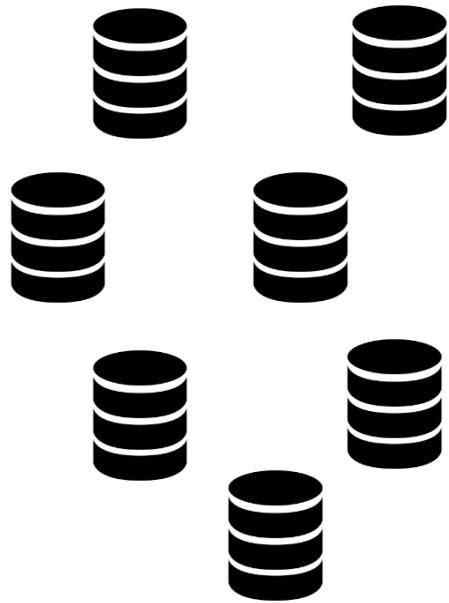


**Enterprise  
Knowledge  
Graph**



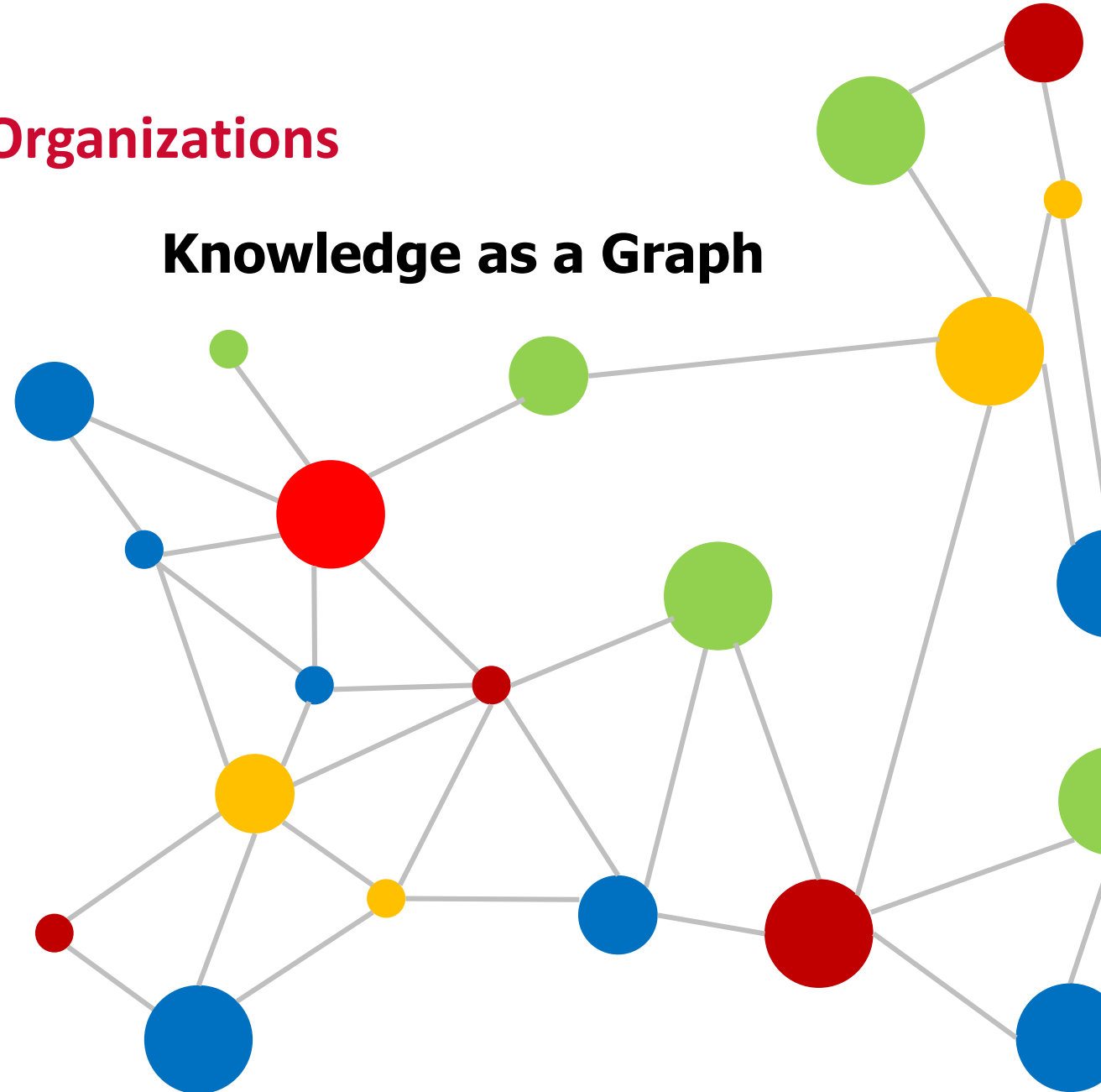
# Relevance for Enterprises and Organizations

## Data silos



Paradigm  
Shift

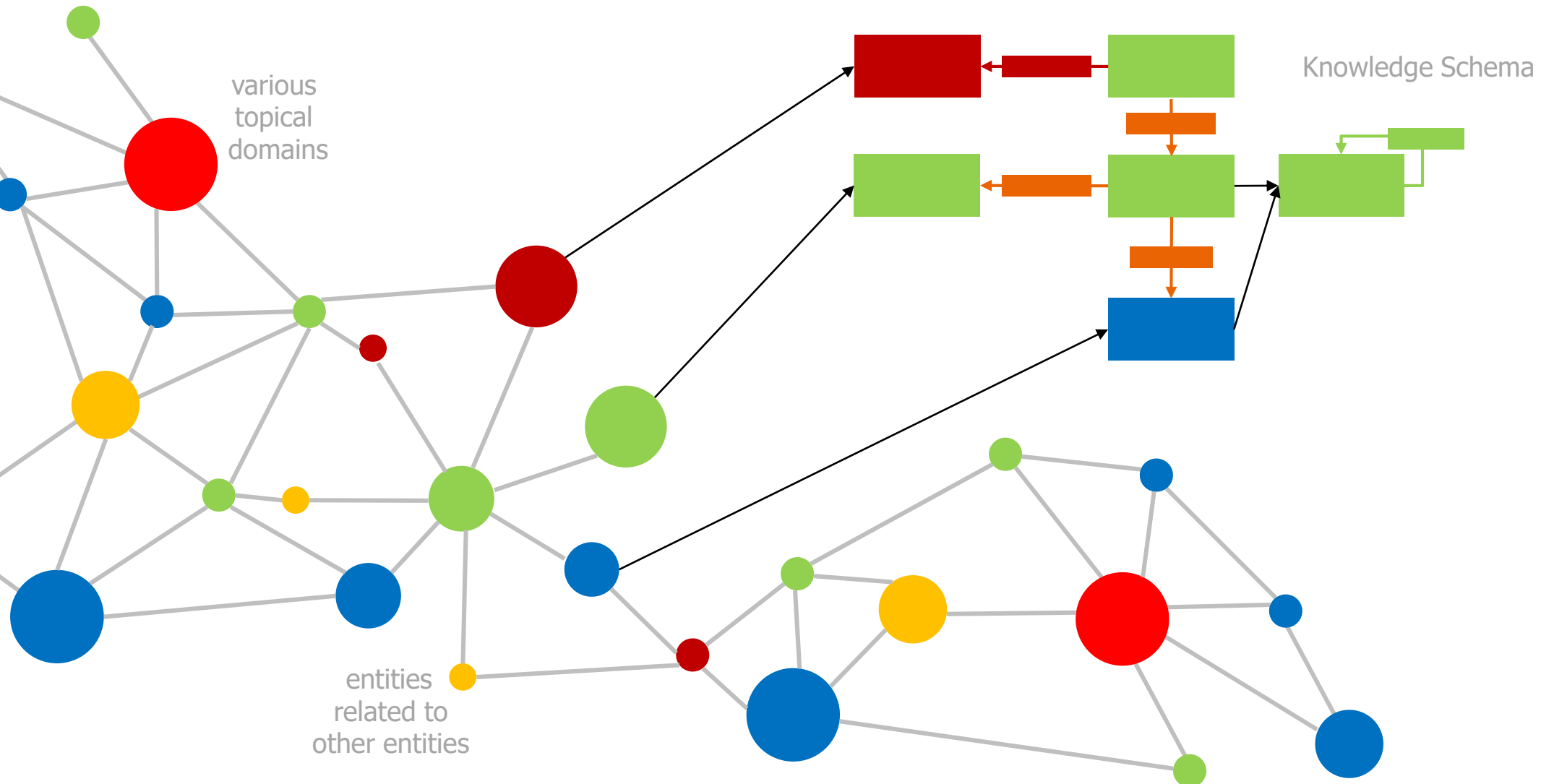
## Knowledge as a Graph





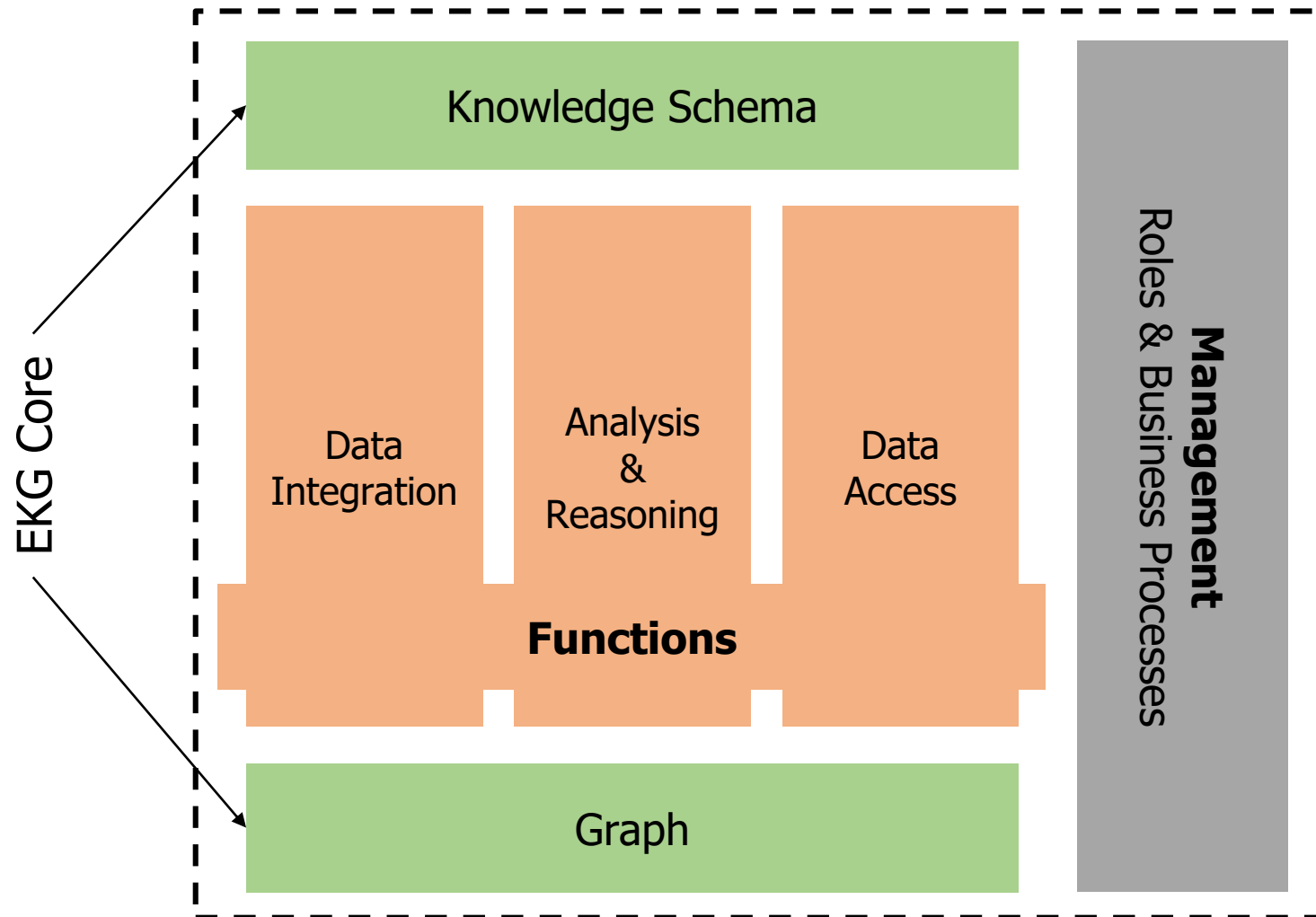


# Base Principles of a Knowledge Graph





# Model of an Enterprise Knowledge Graph





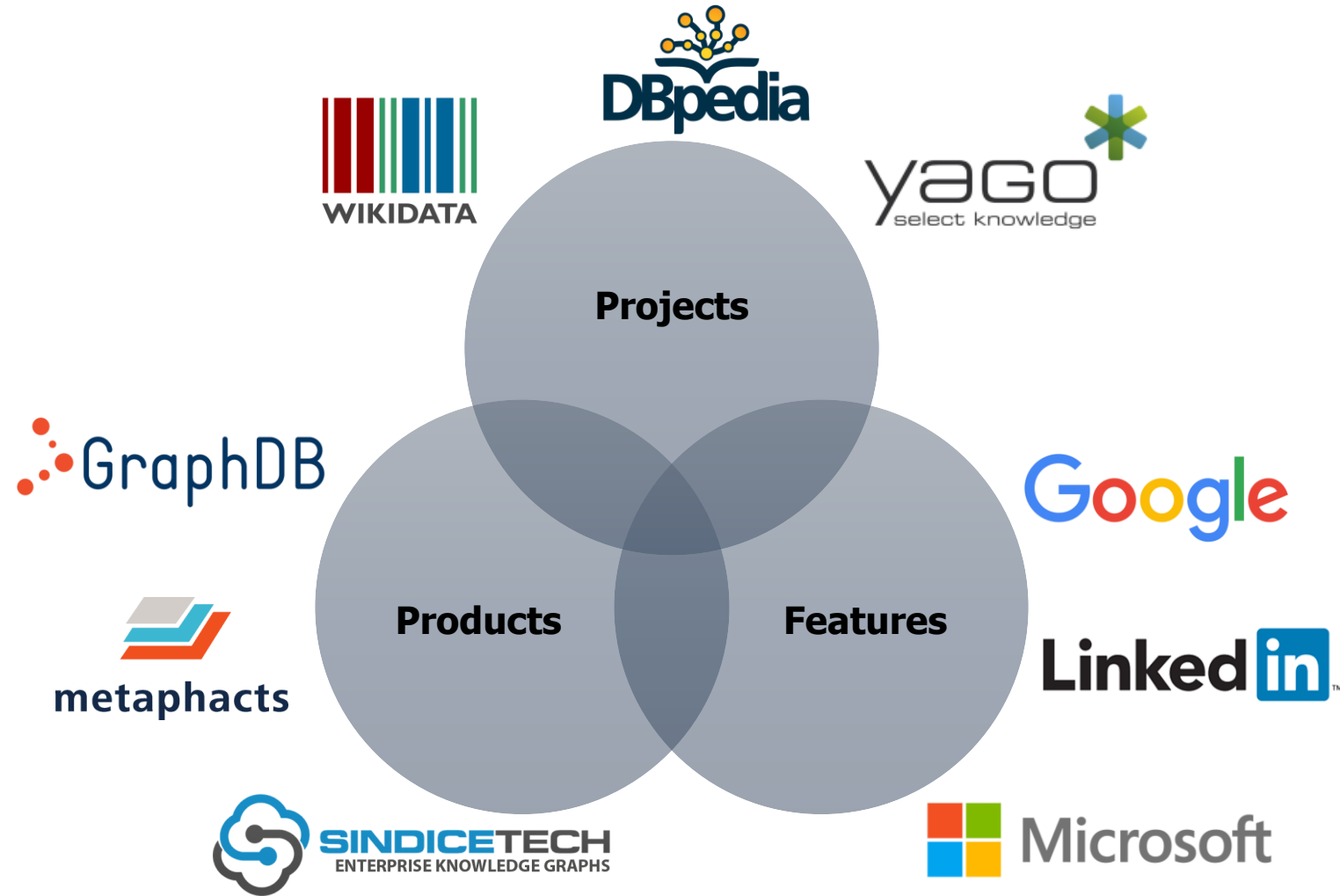
# Agenda

- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph





# Knowledge Graph Technologies in Industry



**B**

**B**

**C**

Deloitte.



ELSEVIER







# Agenda

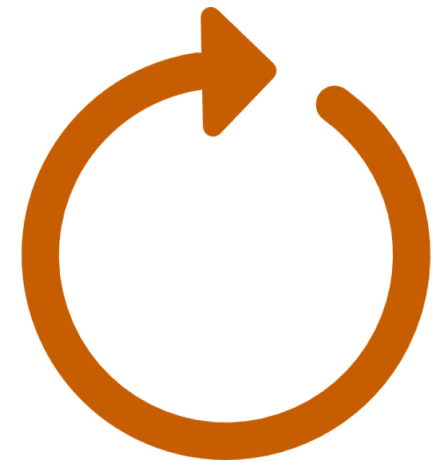
- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph





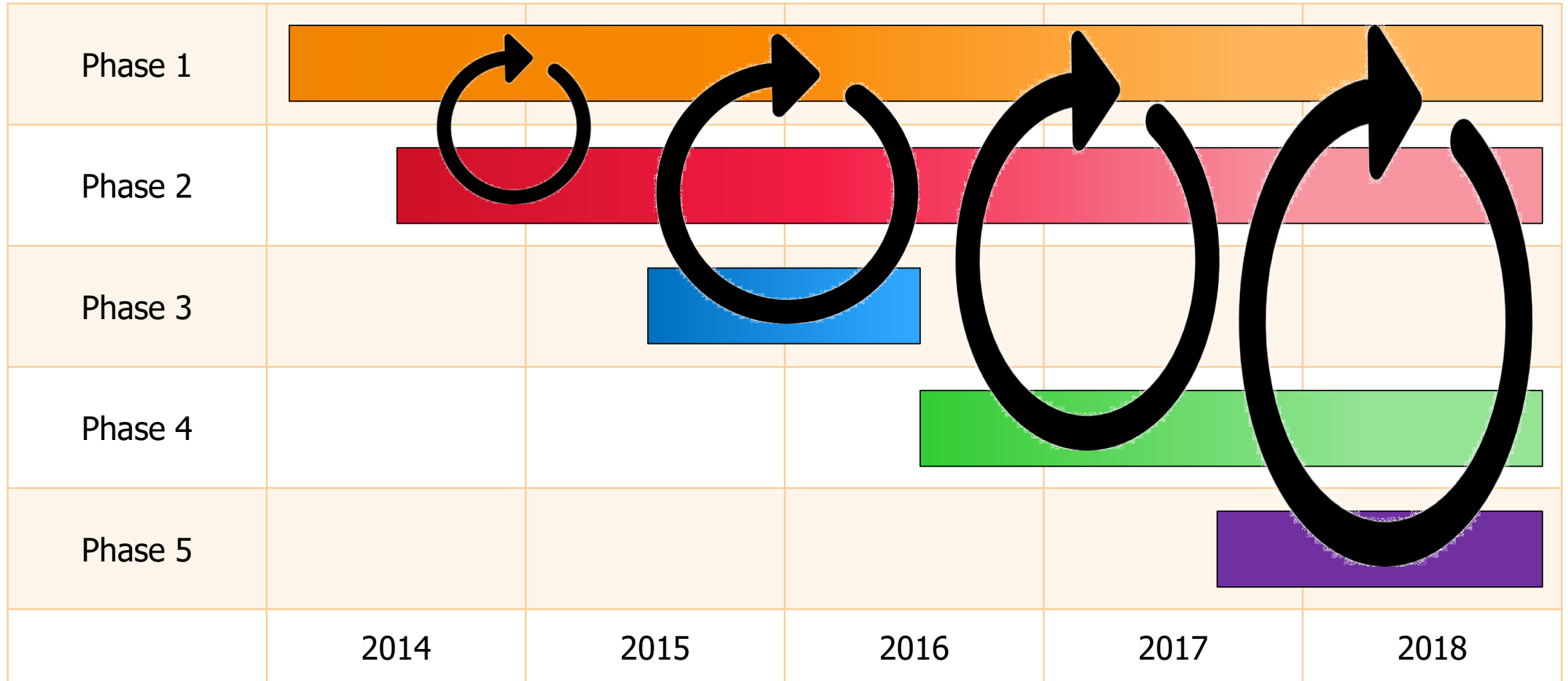
## EduGraph – Iterative Problem Solving

- Phase 1 – Exploration of the domain since 2014
- Phase 2 – Schema engineering and initial design since 2014
- Phase 3 – Architectural draft and spike solution 2015 - 2016
- Phase 4 – Prototypical implementation since 2016
- Phase 5 – Productive implementation will start in 2017





# Iterative Problem Solving





# Base Problem Statement & Vision of a Solution

1

## CASE

Study guides promise assistance with the study search. The data must be maintained by the universities in about 50 systems. The university sites retain this data anyway. A unified annotation could lead to significant savings.

## SOLUTION

Prospective students can access the federated data of courses in a professional domain by an application. The data is maintained locally by those responsible at the universities. Areas of specialization and qualification profiles become transparent.





## Initial Situation at Universities

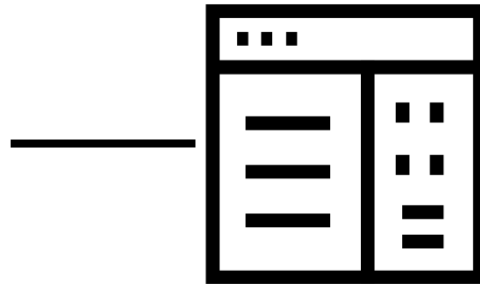
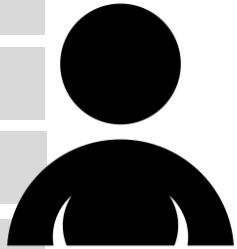
①

Study Programs

Courses

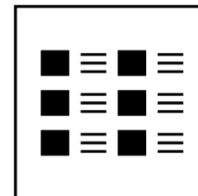
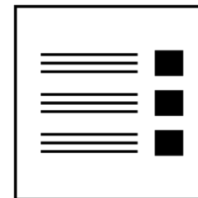
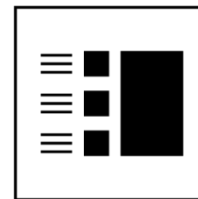
Experts

Publications



# Data Maintenance in x Systems

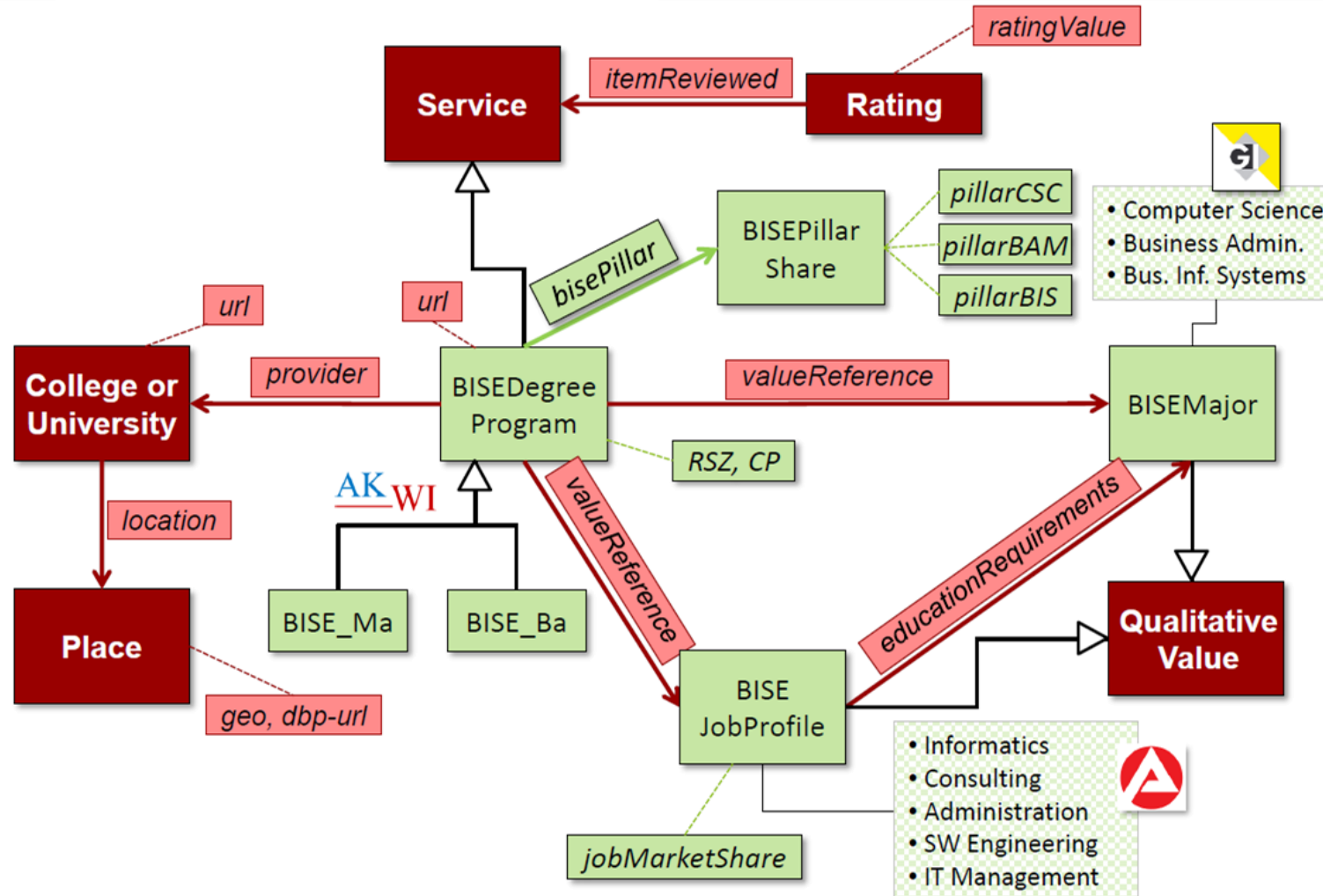
- Effort
- Inconsistencies
- Inhomogeneity





# First Draft of an EKG Schema

2



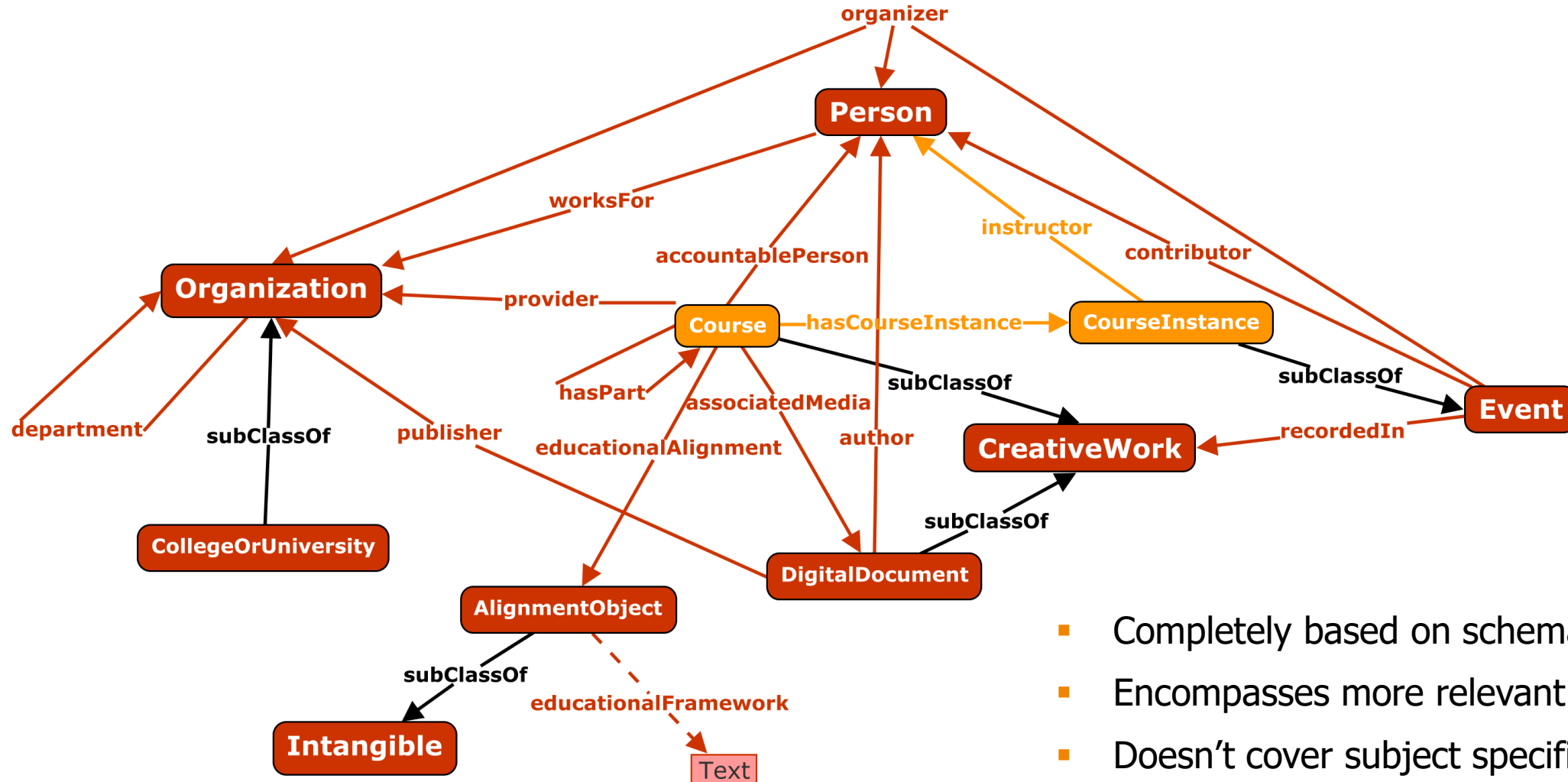
schema.org

proprietary schema elements



# Refinement and Enhancement of the Schema

2

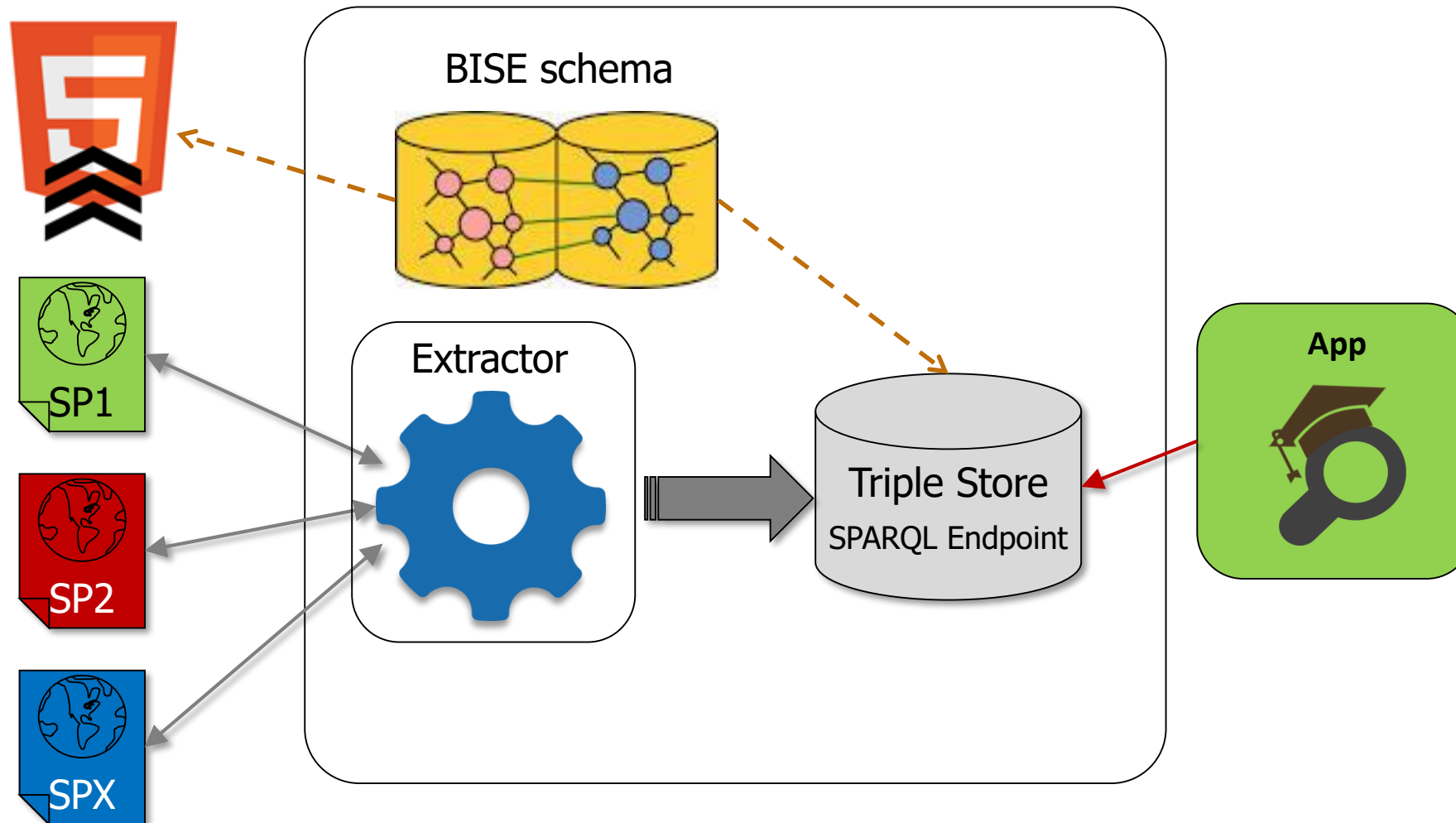


- Completely based on schema.org
- Encompasses more relevant entities
- Doesn't cover subject specific elements



# Architectural Draft for Spike Solution

3

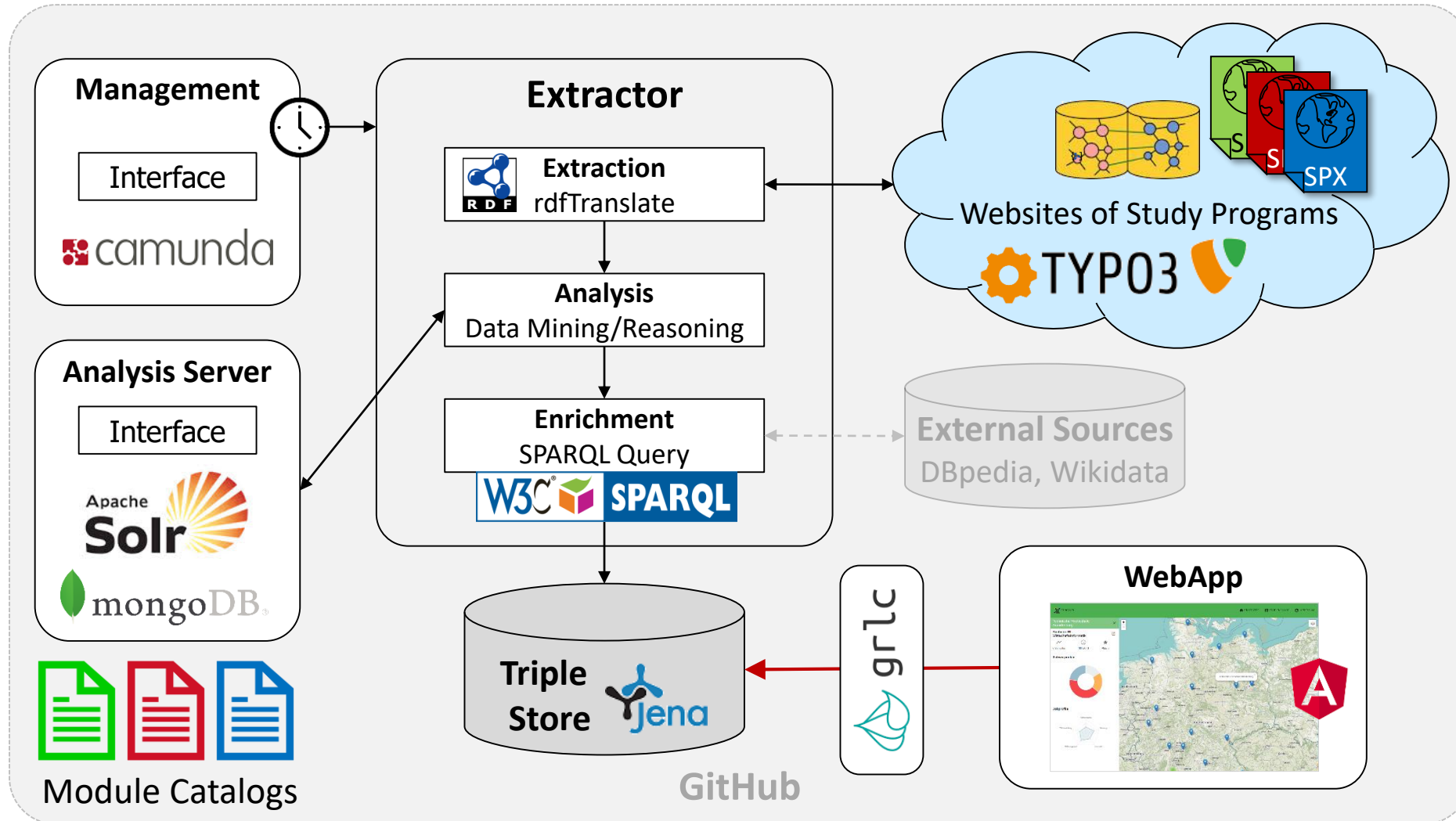






# Prototypical Implementation – Architecture and Technologies

4

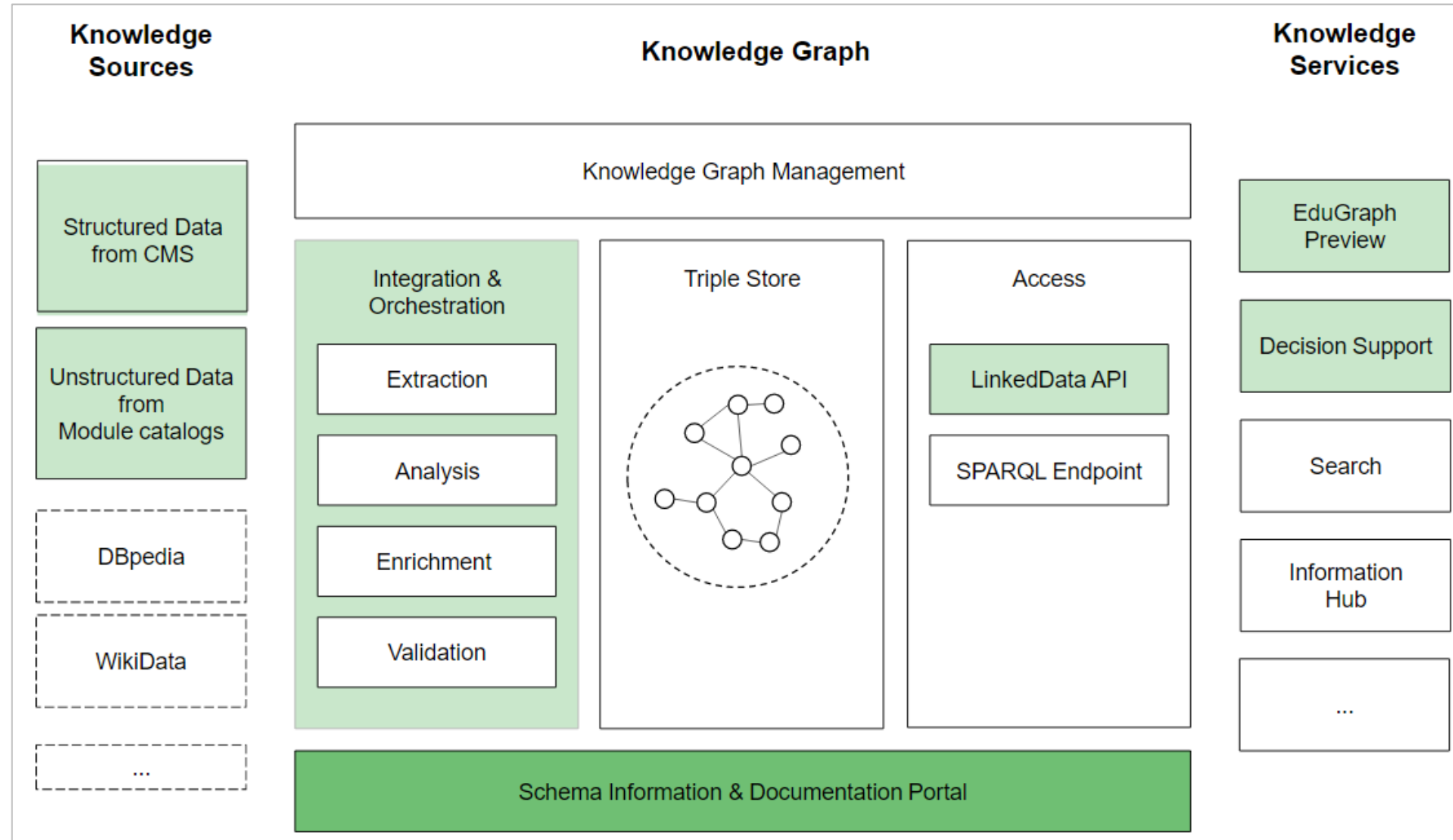




# Agenda

- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph







# Agenda

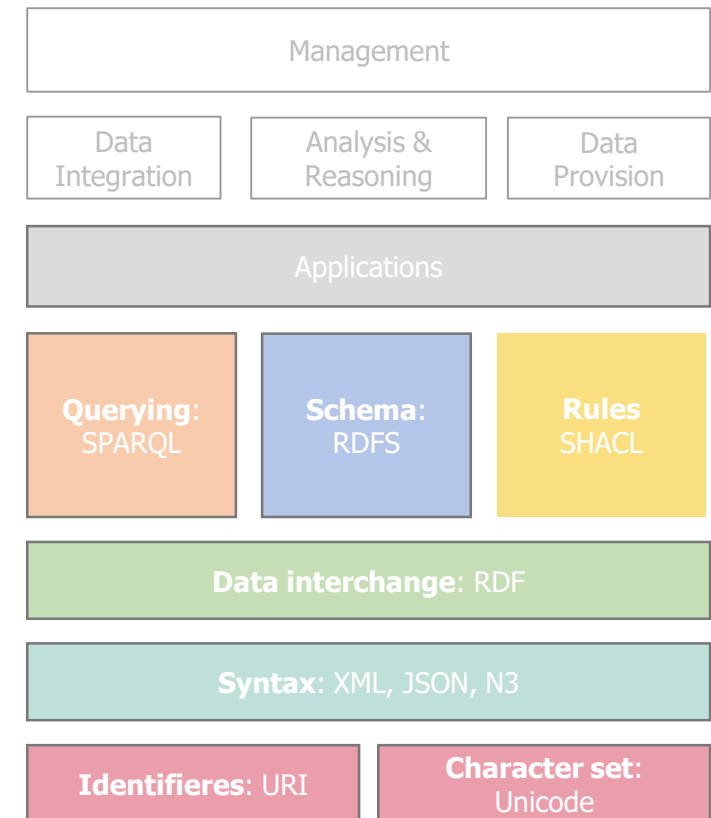
- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph





# Adapted Semantic Web Stack

- Well known standards of the W3C
- Standards build on each other
- Also known as Semantic Web layer cake
- A lot of different versions are published
- Here: adapted version with regard to EduGraph





# Adapted Semantic Web Stack

`http://akwi.de/ns/bise#WIB-THB`

**Identifiers:** URI

**Character set:**  
Unicode



# Adapted Semantic Web Stack

**N3/Turtle** `@prefix bise: <http://akwi.de/ns/bise#> .`  
`@prefix schema: <http://schema.org/> .`

`bise:WIB-THB a bise:BISEBachelor;`  
`schema:name "Bachelor WI - Wirtschaftsinformatik".`

**JSON-LD** `{`  
 `"@id": "http://akwi.de/ns/bise#WIB-THB",`  
 `"@type": "http://akwi.de/ns/bise#BISEBachelor",`  
 `"schema:name": "Bachelor WI - Wirtschaftsinformatik",`  
`}`

**Data interchange:** RDF

**Syntax:** XML, JSON, N3

**Identifiers:** URI

**Character set:**  
Unicode



# Adapted Semantic Web Stack

@prefix schema: <http://schema.org/> .  
@prefix bise: <http://akwi.de/ns/bise#> .

## Schema

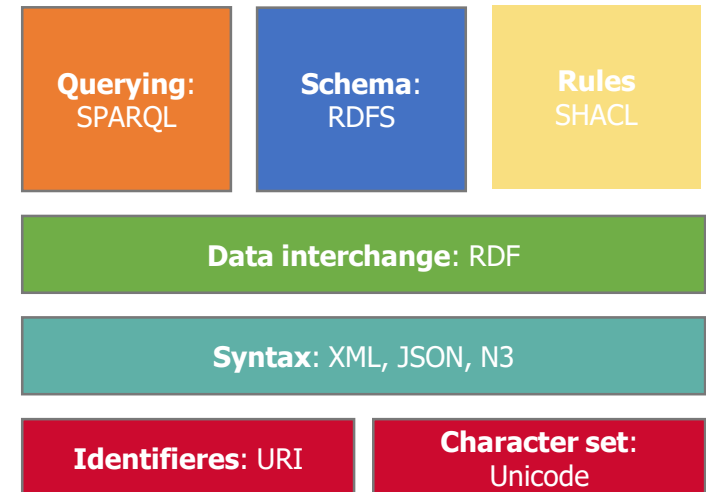
```
bise:BISEBachelor a owl:Class ;  
  rdfs:subClassOf bise:DegreeProgram ;  
  rdfs:comment "Ein Bachelor-Studiengang WI an einer FH ..."@de ;  
  rdfs:label "BISEBachelor" ;  
  skos:definition "Ein Bachelor-Studiengang WI an einer FH ."@de ;  
  skos:prefLabel "Bachelor Wirtschaftsinformatik"@de  
  skos:prefLabel "Bachelor in Information Systems"@en .
```

## Data

```
bise:WIB-THB a bise:BISEBachelor;  
  schema:name "Bachelor WI - Wirtschaftsinformatik".
```

## SPARQL

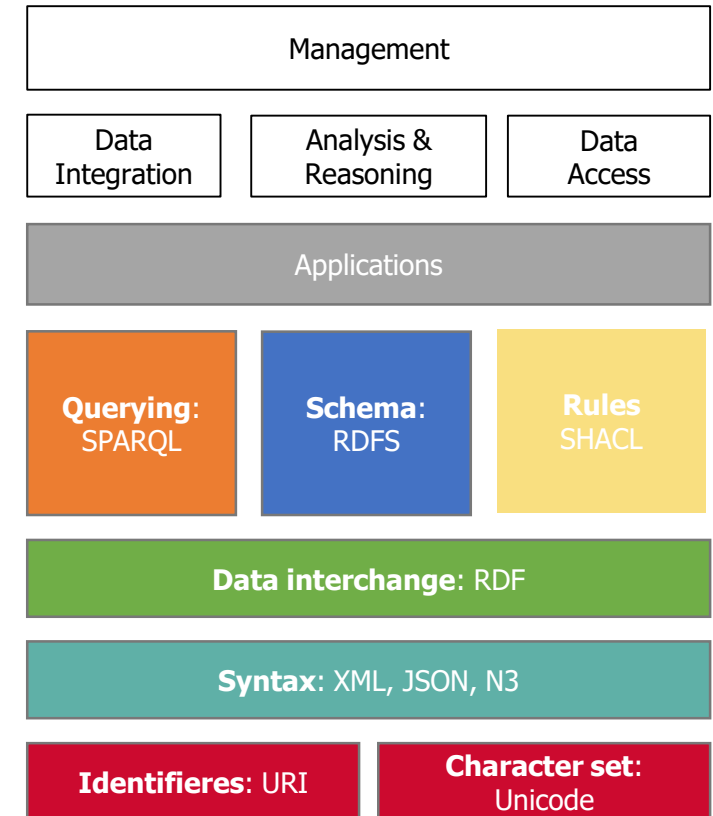
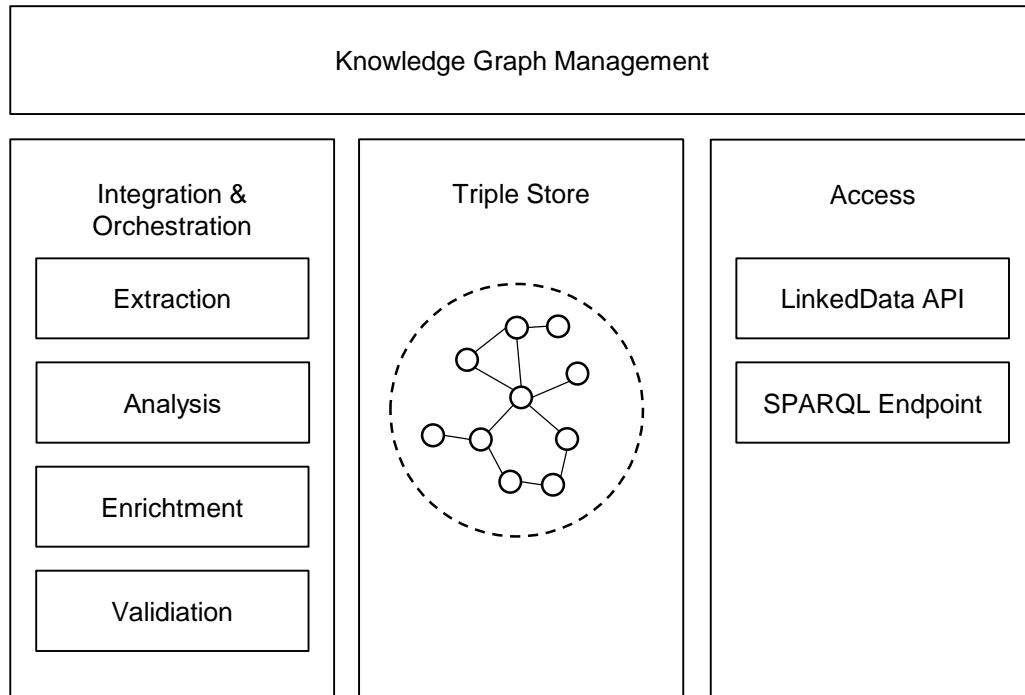
```
SELECT ?name  
WHERE {  
  ?degreeProgram a bise:BISEBachelor;  
  ?schema:name ?degreeProgramName.  
}
```







# Adapted Semantic Web Stack





# Agenda

- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph





# Exercise: Orientation in EduGraph Technologies

General remark: Use the links at the EduGraph demo site (see link on slide 36).

## 1. Source code of CMS websites containing structured data in JSON-LD format

- Open the block **Structured Data from CMS** and click the **Demo site** link,
- Open the source view of the site, detect the JSON-LD script tag and copy its content,
- Use <http://json-ld.org/playground/> for the validation and visualization of the data,
- Compare the data structure with the schema presented at slide 32.

## 2. Web application for decision support for prospective students

- Open the block **EduGraph Preview**, click the **Demo** link and start the preview,
- Compare the visualized data with the two schemas presented at slides 31 and 32.

## 3. Data access via REST API using the grlc service

- Open the block **LinkedAPI** and click the **API doc** link,
- Explore the three different resources provided by the API and try them out,
- Interpret the server responses.



# Agenda

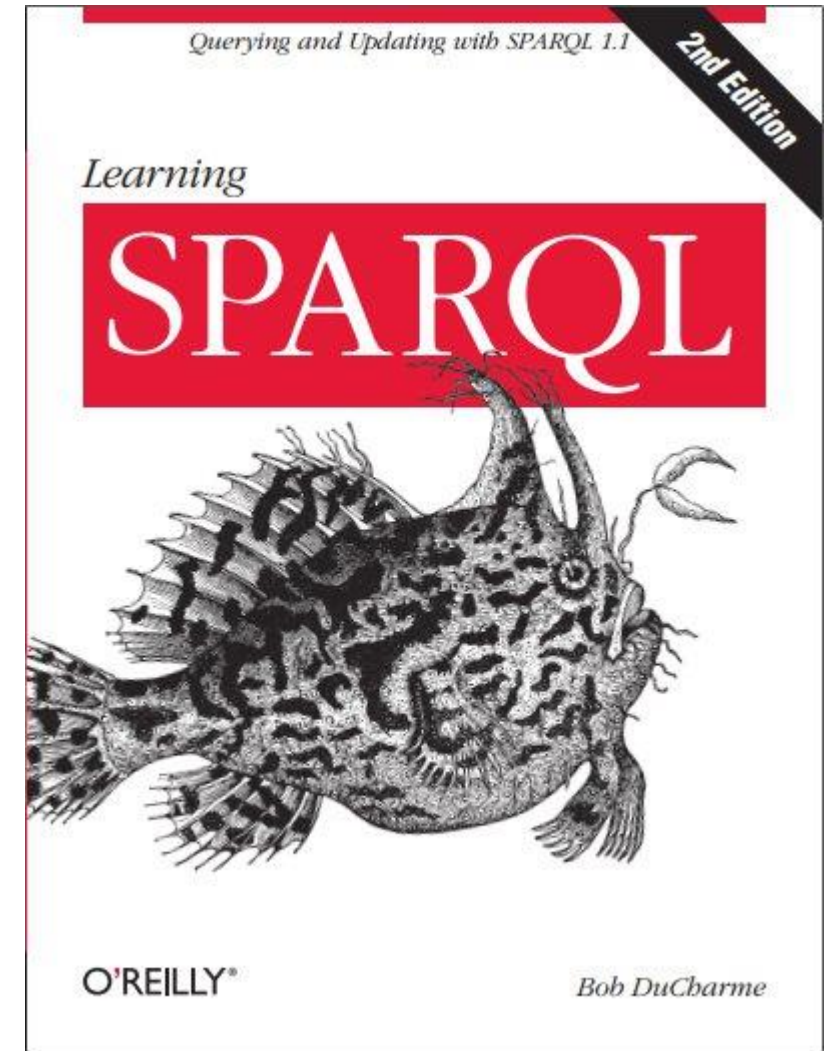
- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph





## Bob DuCharme: Learning SPARQL

- Watch the YouTube video: SPARQL in 11 minutes:  
<https://www.youtube.com/watch?v=FvGndkpa4K0>
- Make notices to discuss the following terms:
  - ✓ RDF • URI • triple • graph
  - ✓ Turtle • @prefix • vocabularies
  - ✓ Representing table data as triples
  - ✓ Where clause • triple patterns • variables
  - ✓ Select clause • results representation
  - ✓ Filter patterns • Optional clause
  - ✓ Further key words and functions





## SPARQL Orientation

- Examine the following github sources of the EduGraph project for SPARQL queries:
  - **Source code for the web application**  
<https://github.com/EduGraph/StudySearch-WebApp/blob/master/app/assets/js/services.js>
  - **Basic queries for the LinkedData API**  
<https://github.com/EduGraph/EduGraph-Queries>
  - **Process for orchestration of extraction and enrichment services**  
<https://github.com/EduGraph/EduGraph-Integration/blob/master/src/main/resources/edu-graph.bpmn>
- Analyze and document the used patterns and functions.



## SPARQL Exercises

Explore the **EduGraph** test database and answer the following questions:

1. How many different predicates (relations between graph nodes) are used in the database?
2. Which subject nodes are connected to other nodes with the predicate `<http://akwi.de/ns/bise#jobMarketShare>` ?
3. To what kind of object nodes the subjects found in 2. are connected?
4. To how many nodes is the following subject resource connected?  
`<http://de.dbpedia.org/resource/Fachhochschule_Brandenburg>`
5. To what object is the subject node mentioned in 4. connected with the predicate `<http://schema.org/geo>` ? Of what type is this object?
6. Which study program is provided by the resource mentioned in 4. ?  
(use the predicate `<http://schema.org/provider>` )
7. In how many triples the study program found in 6. is the subject?  
Of what type are the objects in that triples?



# Agenda

- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph







## Final Assignments

General remark: use in addition to GraphDB the DBpedia Endpoint via <https://dbpedia.org/sparql>

1. The EduGraph test database contains a number of UAS providing study courses in Information Systems. First, find out how many UAS are in the database.
2. To provide further data about the DBpedia resource link to the city where the UAS is located is connected to each UAS via the predicate `<http://schema.org/location>`.  
Check that for each UAS found in 1. there is given a location in the described way.
3. Explore by way of example whether the DBpedia provides data in Russian/Polish language for university locations collected in the EduGraph database.
4. Develop a SPARQL query for enriching the EduGraph database by this language-specific data.
5. Save your query in a text file, name the file as follows: `ekgFA_lastName.txt` and send this file via eMail to the lecturer: [vera.meister@th-brandenburg.de](mailto:vera.meister@th-brandenburg.de) .



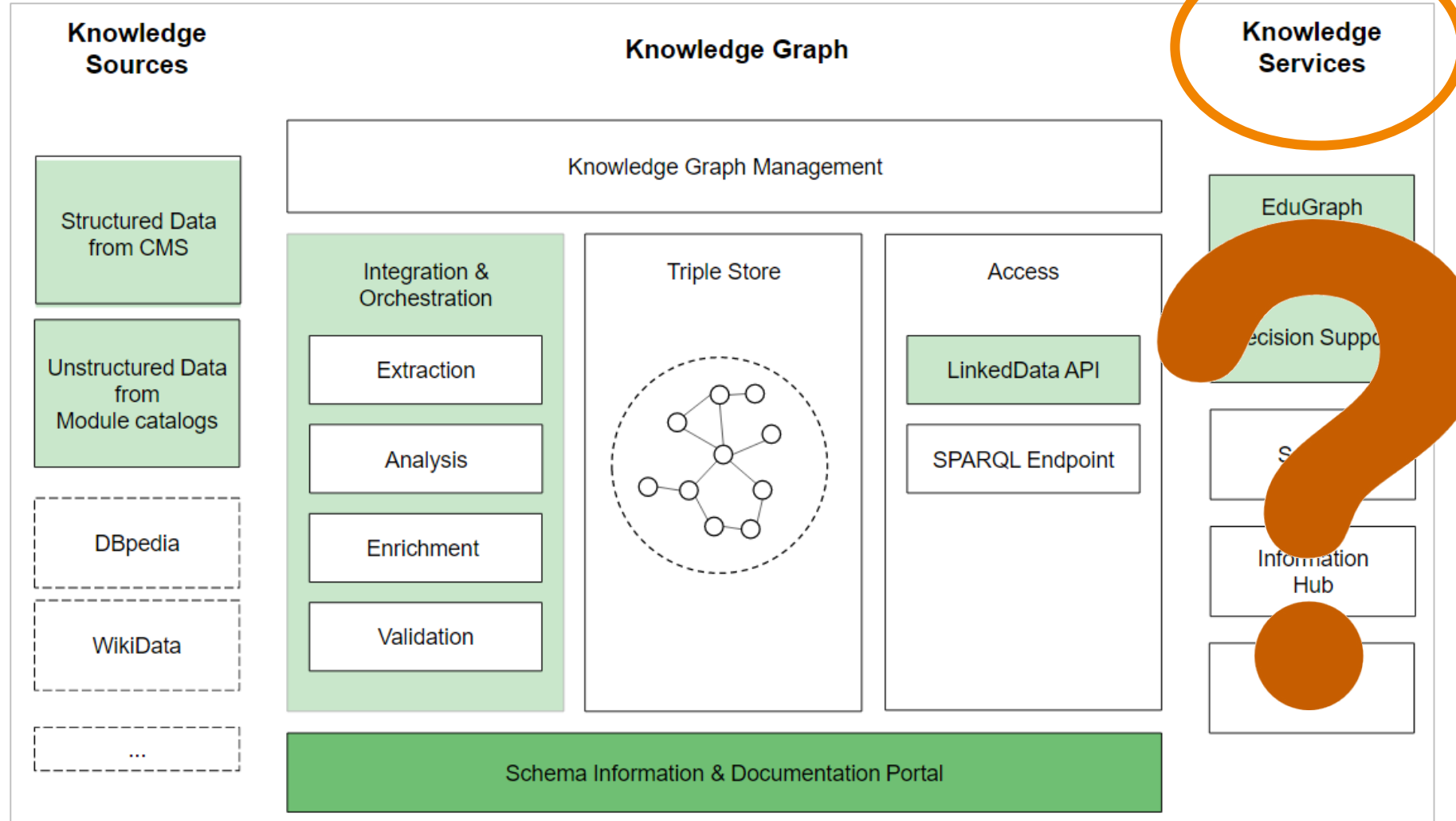
# Agenda

- Problem and Domain Description
- Introduction: Enterprise Knowledge Graph
- EKG – Industry Implementations
- EduGraph – Iterative Problem Solving
- EduGraph – Architecture and Technologies
- EKG Technologies and the Semantic Web Stack
- Exercise: Orientation in EduGraph Technologies
- Exkurs & Exercises: SPARQL Basics
- Final Assignment: Further Enrichment of EduGraph Data
- Brain Storming: New Knowledge Services related to EduGraph?





# EduGraph Demo Site





## New Knowledge Services related to EduGraph

- ❖ Collect ideas for new knowledge services related to the EduGraph infrastructure.
  - What may be new target groups for the collected data?
  - What information needs you can observe or anticipate for that target group?
  - Which kind of data/information access fits to the target group?
  - What is to be expected the biggest obstacles or barriers for the service provision?





**Technische Hochschule  
Brandenburg**

University of  
Applied Sciences

**Fachbereich  
Wirtschaft**

Thank you for your attention!

More Information at:      <http://bmake.th-brandenburg.de>  
                                 <http://edugraph.github.io/ESWC2017/>

Prof. Dr. Vera G. Meister • [vera.meister@th-brandenburg.de](mailto:vera.meister@th-brandenburg.de) • +49-175-5634180  
Research Group Business Modeling and Knowledge Engineering (BMaKE)