FAKULTÄT FÜR !NFORMATIK

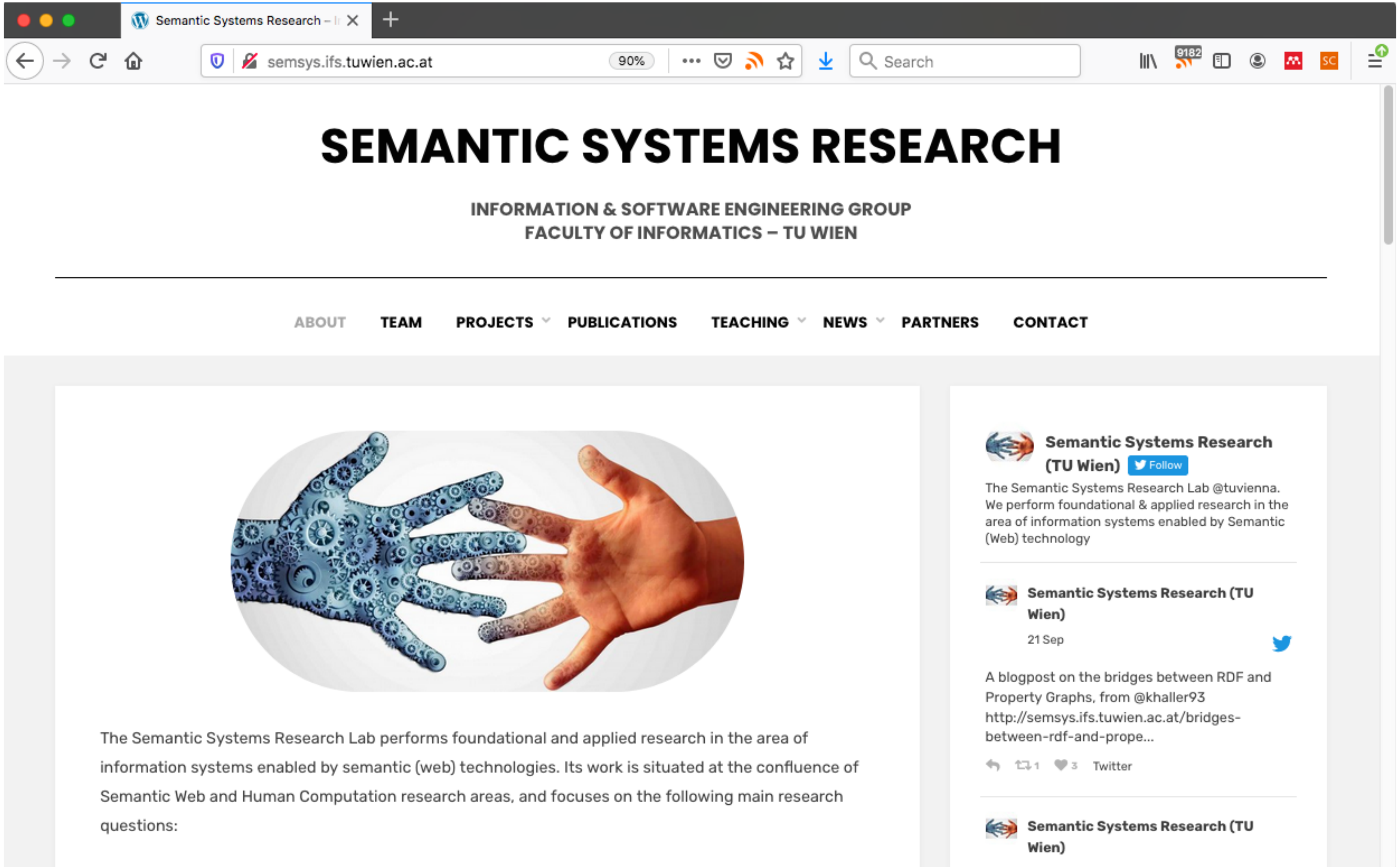
Faculty of Informatics

Semantic Systems Research Lab -Projects-

Marta Sabou
Fajar J. Ekaputra
Peb R. Aryan
Laura Waltersdorfer
Kevin Haller
Stefani Tsaneva
Gregor Käfer
Ayu Irsyam

Institute of Software Technology and Interactive Systems
Information and Software Engineering Group
Vienna University of Technology

Favoritenstraße 9-11/177,
1040 Vienna, Austria


Semantic Systems Research –
semsys.ifs.tuwien.ac.at
90%
Search
SEMANTIC SYSTEMS RESEARCH
INFORMATION & SOFTWARE ENGINEERING GROUP
FACULTY OF INFORMATICS – TU WIEN
ABOUT
TEAM
PROJECTS
PUBLICATIONS
TEACHING
NEWS
PARTNERS
CONTACT
The Semantic Systems Research Lab performs foundational and applied research in the area of information systems enabled by semantic (web) technologies. Its work is situated at the confluence of Semantic Web and Human Computation research areas, and focuses on the following main research questions:
Semantic Systems Research (TU Wien)
Follow
The Semantic Systems Research Lab @tuvienna. We perform foundational & applied research in the area of information systems enabled by Semantic (Web) technology
Semantic Systems Research (TU Wien)
21 Sep
A blogpost on the bridges between RDF and Property Graphs, from @khaller93 http://semsys.ifs.tuwien.ac.at/bridges-between-rdf-and-prope...
Twitter
Semantic Systems Research (TU Wien)

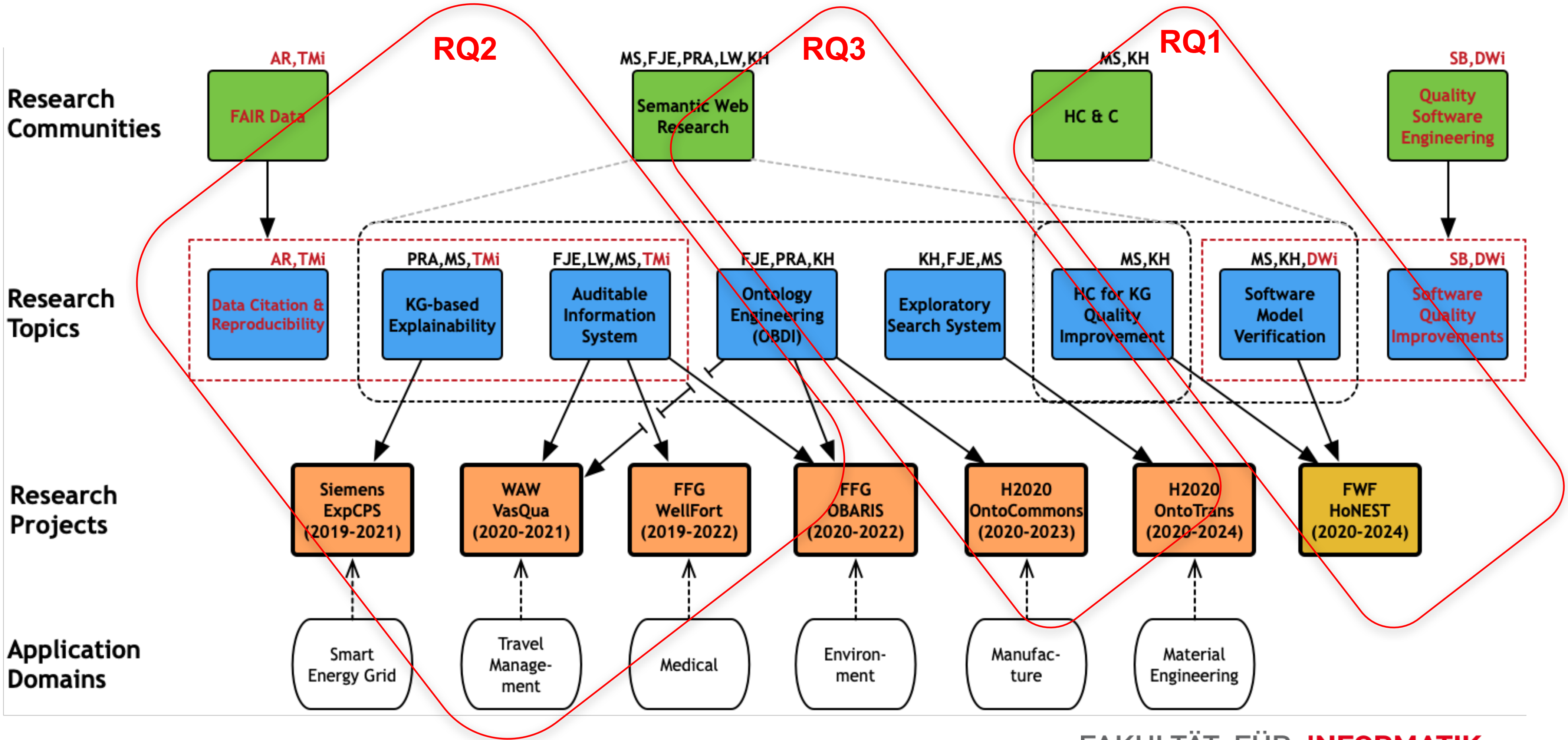
(Broad) Research Questions

(1) How to involve people in (semantic) system design both for information and software engineering? We investigate the use of Human Computation methods for acquiring and verifying knowledge structures such as ontologies and knowledge graphs (HOnEst). We also use Human Computation methods for verifying software engineering models.

(2) How to ensure that (semantic) information systems operate in reliable (trustworthy) and ethically acceptable ways? We perform work on using semantic technologies to make information systems auditable (OBARIS, WellFort, VasQua). Additionally, we extend this research to cyber-physical systems (CPS) where we aim to enable their explainability (ExpCPS).

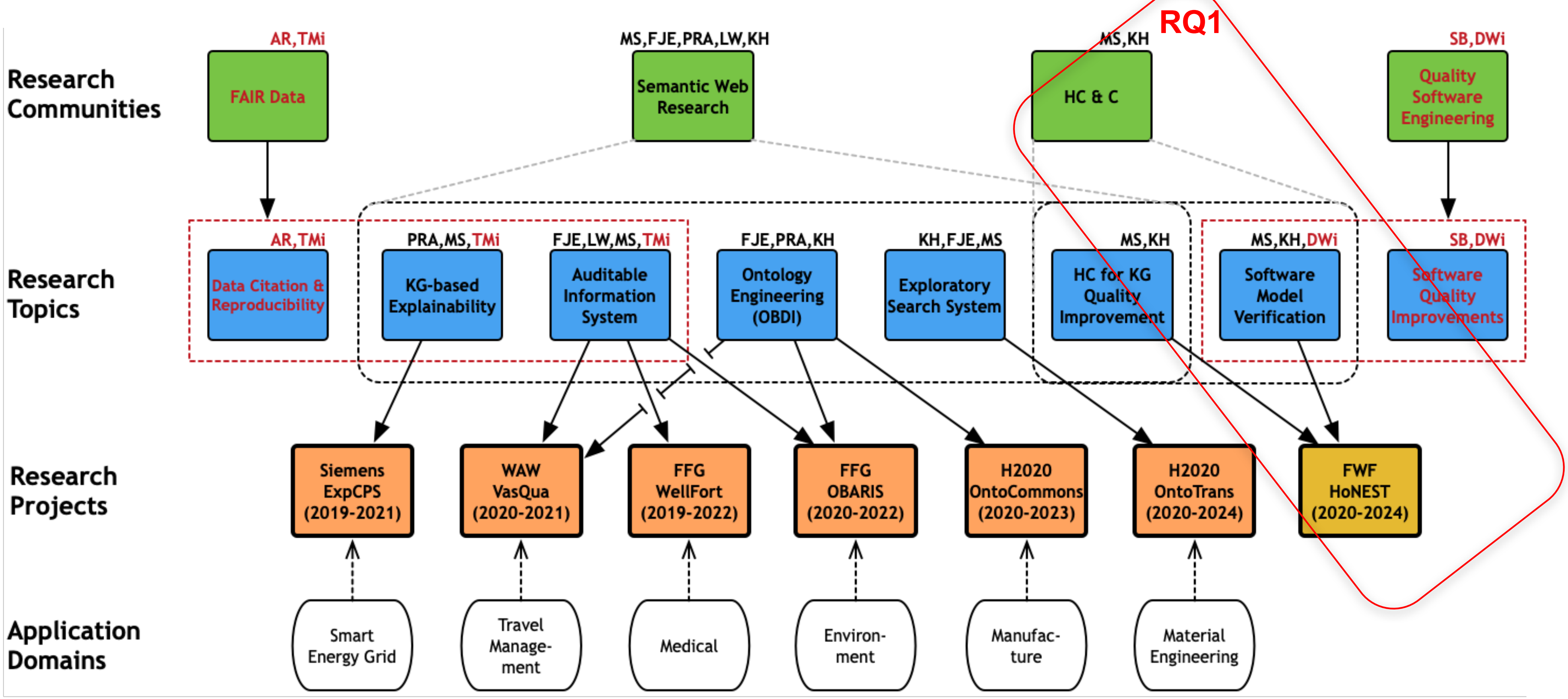
(3) How can semantic (web) technologies improve information systems? We investigate in particular the benefits of semantic methods from the area of data integration (CitySPIN, CDL-Flex) and exploratory search (OntoTrans, STAR) in information systems from a broad range of domains including Smart City, Industrie 4.0, material engineering, and medical science.

Research Areas and Projects

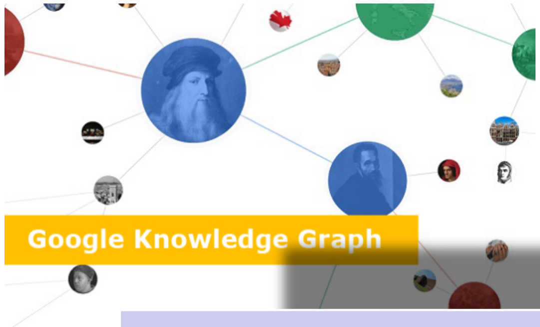


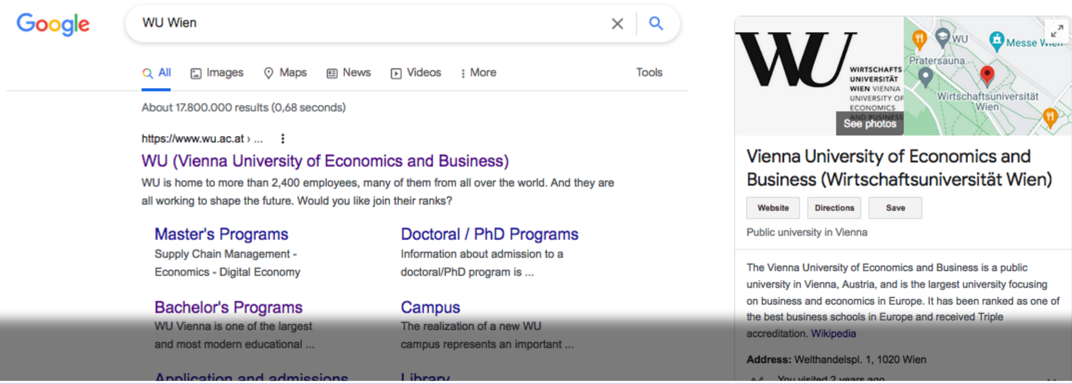
Research Areas and Projects

(1) How to involve people in (semantic) system design both for information and software engineering?



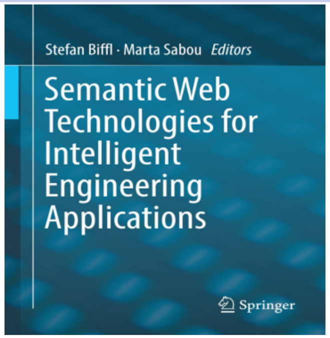
Ontologies and Knowledge Graphs



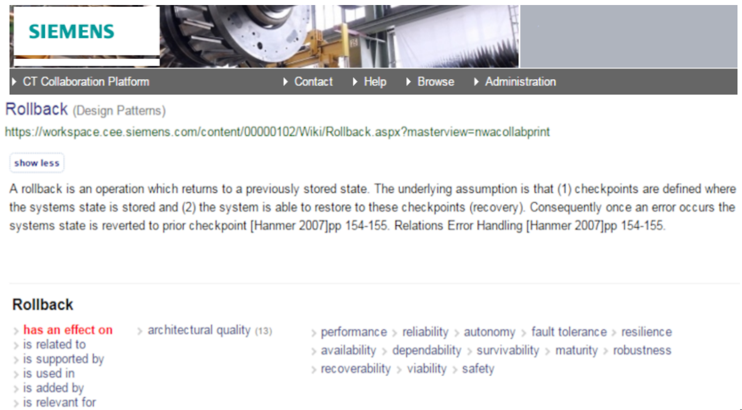


Ontology defects hamper the system’s ability to provide factually correct and unbiased answers.

- often human involvement is needed to ensure high quality domain models

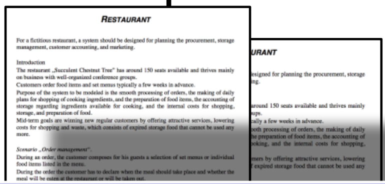


CDL-Flex
Industrie4.0



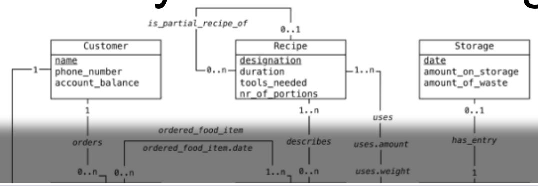
System EER Diagrams

System Specification



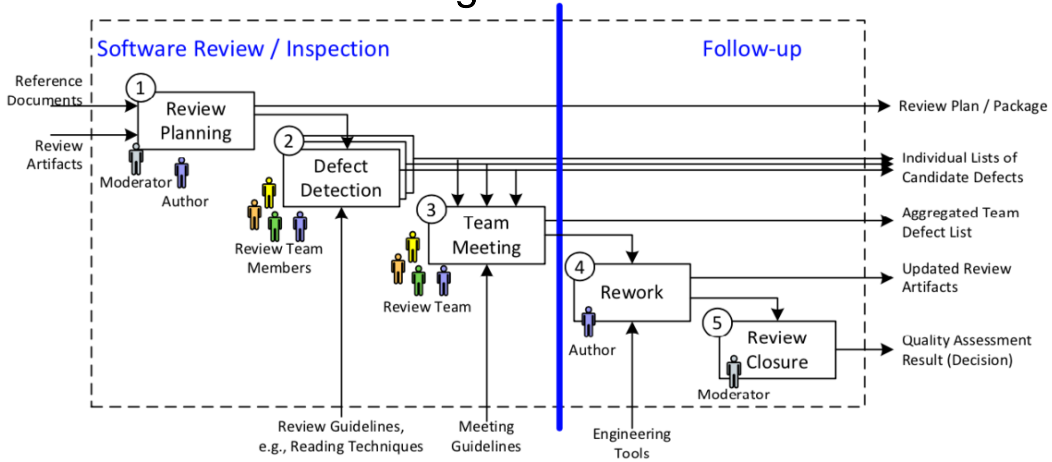
Does the model **completely** and **correctly** represent the system specification?

System EER Diagram

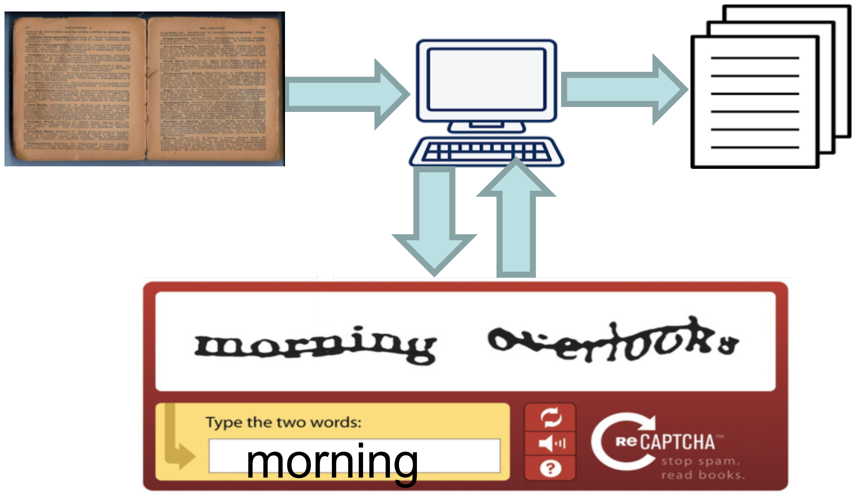


Incorrect system diagrams lead to faults in the follow-up artifacts:

- High costs
- Software model inspection
  - on-site verification processes:
  - not scalable to larger models
  - not suitable for virtual organisations

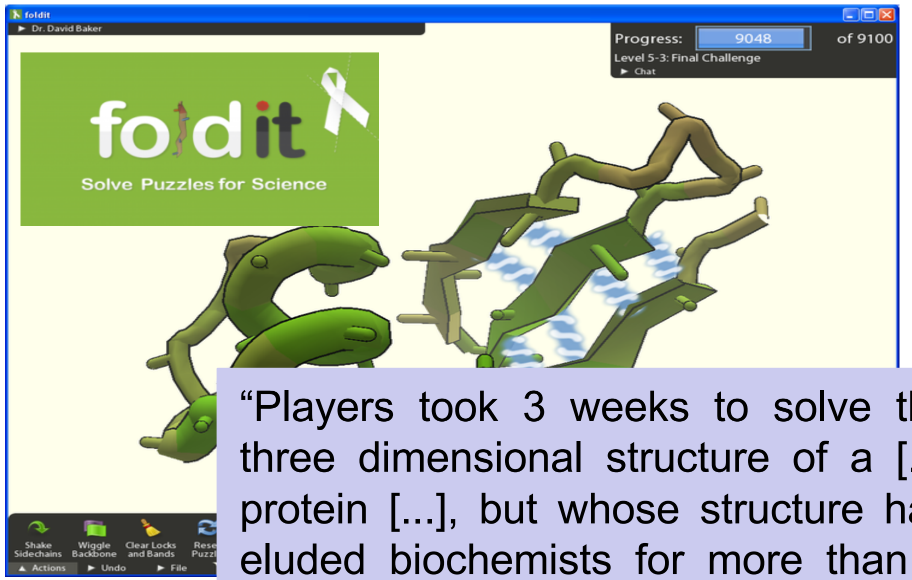


Human Computation harnesses Human Contributions on Large Scale



100 million/ day

Digitization of:
13 million NYT articles;
2 million books per year;



"Players took 3 weeks to solve the three dimensional structure of a [...] protein [...], but whose structure had eluded biochemists for more than a decade."
http://blogs.nature.com/spoonful/2012/04/foldit-games-next-play-crowdsourcing-better-drug-design.html

S. Cooper, [other authors], and **Foldit players**: Predicting protein structures with a multiplayer online game. Nature, 466(7307):756-760, 2010.

Source: L.vonAhn,B.Maurer,C.McMillen,D.Abraham,M.Blum, Science 321, 1465 (2008).

Motivation

Semantic Web

Software Engineering

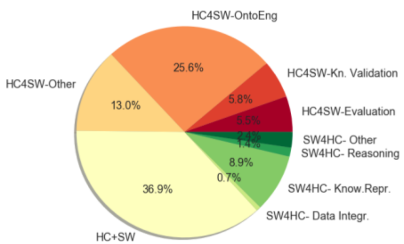
Ontologies, taxonomies, Knowledge Graphs need manual verification

EER diagrams manually checked for completeness and correctness with respect to system specifications

How to propose an HC based solution to the problem of conceptual model verification that is applicable across research areas?

Contributions:

- *Problem formalization*
- *HC Approach*
- *Evaluation in the Software and Knowledge Eng. Areas*

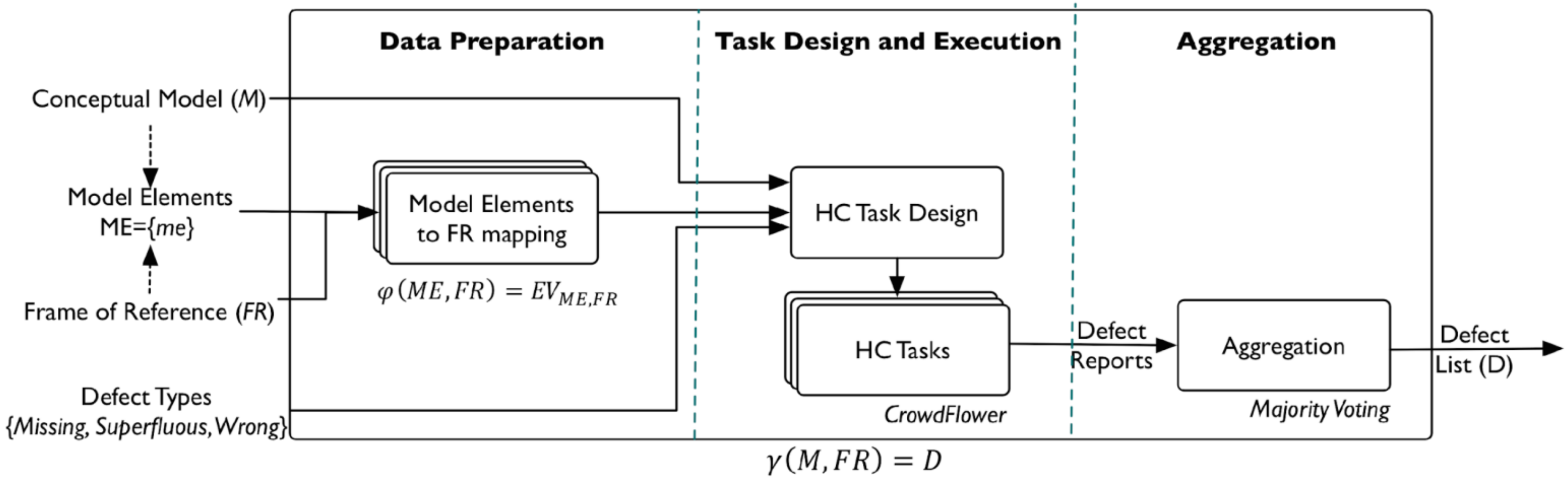


M. Sabou, L. Aroyo, K. Bontcheva, A. Bozzon, R. K. Qarout. Semantic Web and Human Computation: the Status of an Emerging Field. Sem. Web. J. 9(3): 291-302 , 2018.

Human Computation

HC techniques primarily used to address cognitively easy tasks (e.g., labeling, object recognition)

Problem Formalization & Generic Approach for Verifying Conceptual Models (VeriCoM)



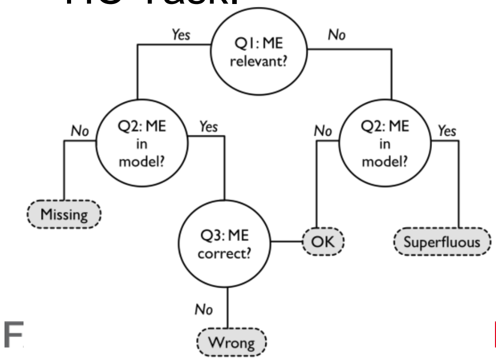
P1: ME-based.
MEs guide the verification.

$$M = \cup_n ME_n$$

P2: Evidence-based.
Evidence for *me* from FR is used as context in the HC Task.

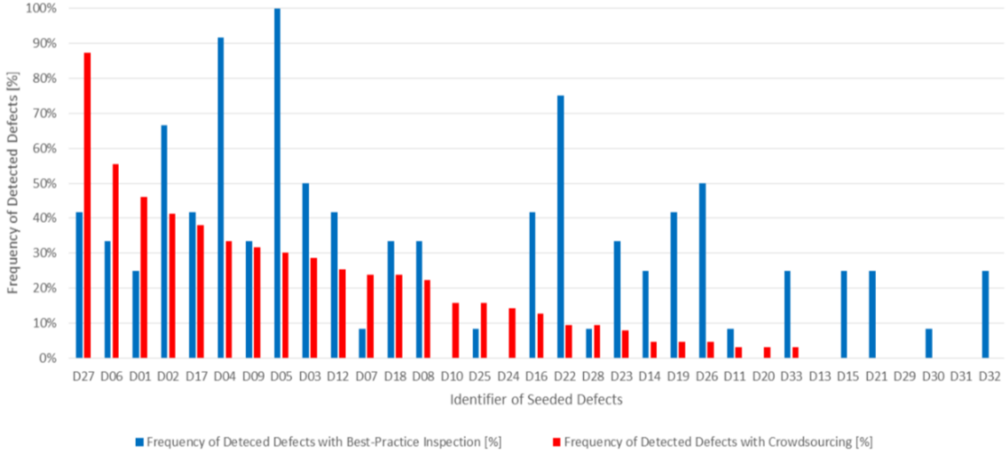
P3: Model as context.
Model fragment used as context in the HC Task.

P4. Defect types.
Defect types used to design HC Task.

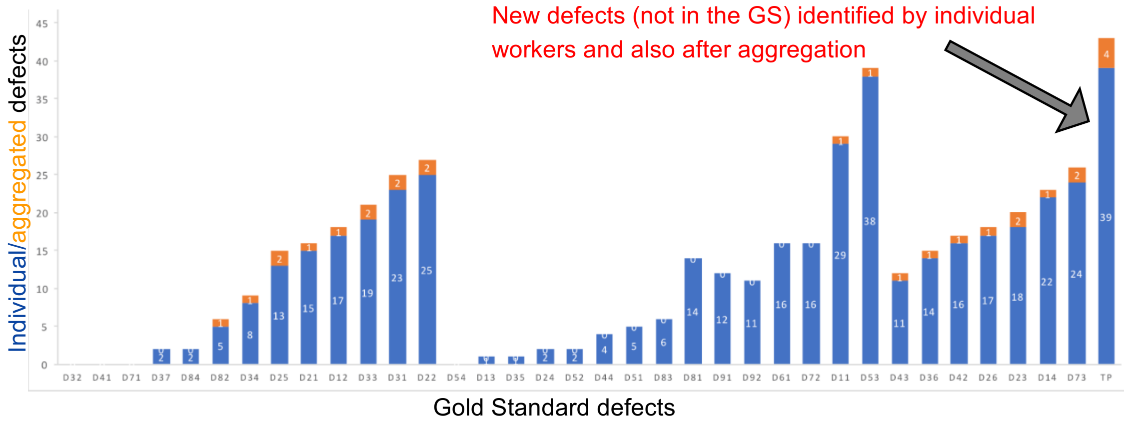


D. Winkler, P. Penzerstadler, S. Biffl: Verifying Conceptual Domain Models with Human Computation: A Case Study in Software Engineering. AAAI HCOMP 2018: 164-173

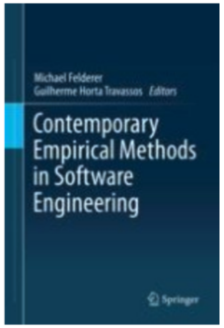
In a Nutshell: Further Results



Crowd-based inspection comparable to traditional inspection [1]



Open ended task design helps harness subtle insights and creativity [2]



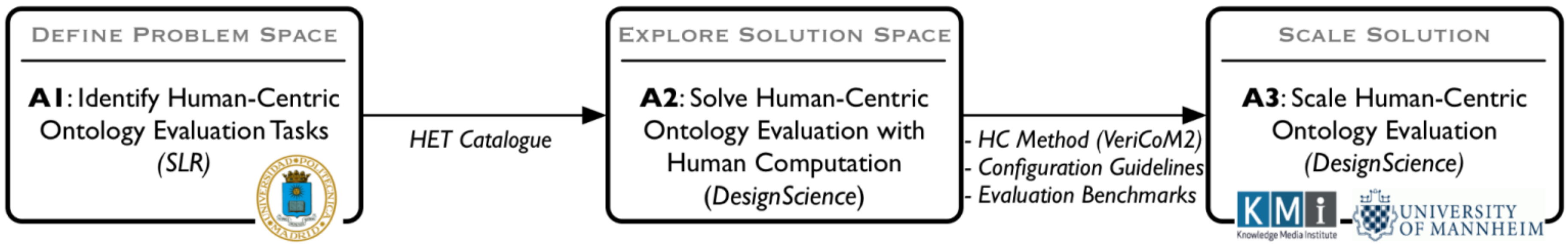
Human Computation technique as a novel method for empirical software engineering experimentation [3].

[1] D. Winkler, M. Kalinowski, M. Sabou, S. Petrovic, S. Biffl: Investigating a Distributed and Scalable Model Review Process. CLEI Electron. J. 21(1) (2018)
[2] M. Sabou, K. Käsznar, M. Zlabinger, S. Biffl, D. Winkler: Verifying Extended Entity Relationship Diagrams with Open Tasks. AAAI HComp. 2020:132-140.
[3] M. Sabou, D. Winkler, S. Biffl: Empirical Software Engineering Experimentation with Human Computation. M. Felderer et al., Contemporary Empirical Methods in Software Engineering 2020:173-215

Ongoing and Future Work

HOnEst: Human-centric Ontology Evaluation (FWF)
AIM: “scalable human-centric ontology evaluation”

- FWF Elise Richter
- 2020-2024
- UPM, KMi, UMannheim



Systematic literature review [1]:
What semantic resources are evaluated with human-involvement, **why** and **how** is the evaluation performed?

Task design, **empirical studies**, support ontology engineer courses during **distance learning [2]**.

Hybrid human-machine system including reasoners and machine learning systems [3].

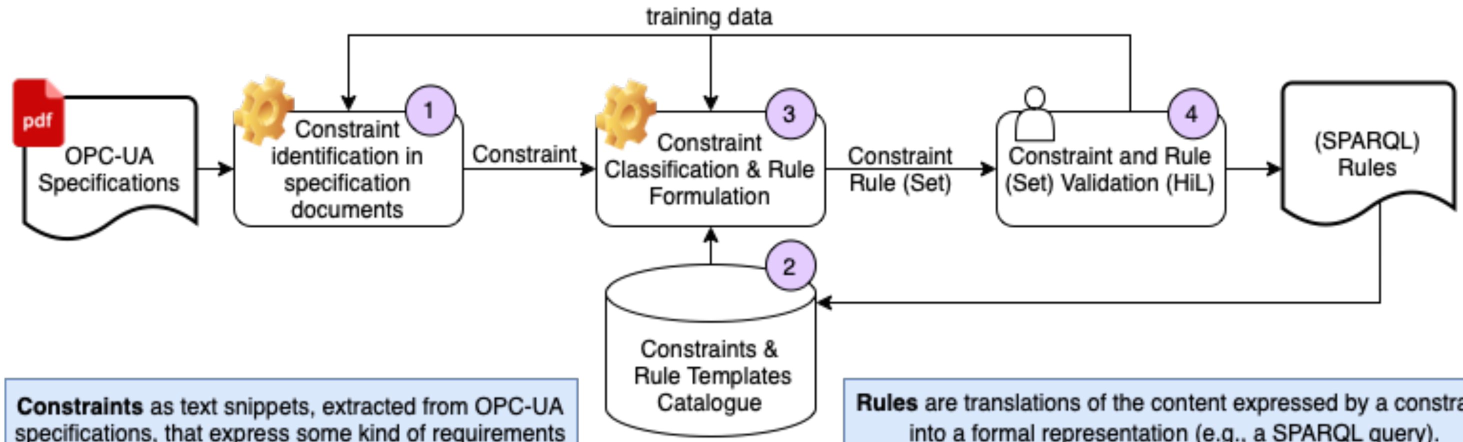
Industry Project: SIEMENS OPC UA

[1] M. Sabou, M. Fernandez, M. Poveda-Villalon, M. C. Suarez-Figueroa, S. Tsaneva. Human-centric Evaluation of Semantic Resources: A Systematic Mapping Study. Prepared for ACM Comp. Surveys.
[2] S. Tsaneva, M. Sabou. A Human Computation Approach for Ontology Restrictions Verification. AAAI HComp 2021.
[3] A. Prock. Hybrid Human-Machine Ontology Verification Identifying Common Errors in Ontologies by Integrating Human Computation with Ontology Reasoners. Master Thesis, TU Wien, 2021.

ORE: OPC-UA Rule Editor (ORE) Project

- Siemens DE
- 2021
- (2022 extension)

Goal: supporting OPC-UA experts in identifying **constraints** in pdf documents of Companion Specifications and formalizing **rules** to validate if the constraints are satisfied in OPC-UA files.



Similarities to BRISE?

Constraints as text snippets, extracted from OPC-UA specifications, that express some kind of requirements that the OPC-UA file conforming to those specifications should fulfil.

EXAMPLE

Each FiniteStateMachine can have at most one State of type InitialStateType, but a FiniteStateMachine does not have to have a State of this type.

Rules are translations of the content expressed by a constraint into a formal representation (e.g., a SPARQL query). In this survey, rules are represented in a structured natural language so that they can be understood also by OPC-UA experts that are not familiar with the syntax of the formal rule language. Complex constraints can be translated in several rules that we refer to as a **Rule Sets**.

EXAMPLE

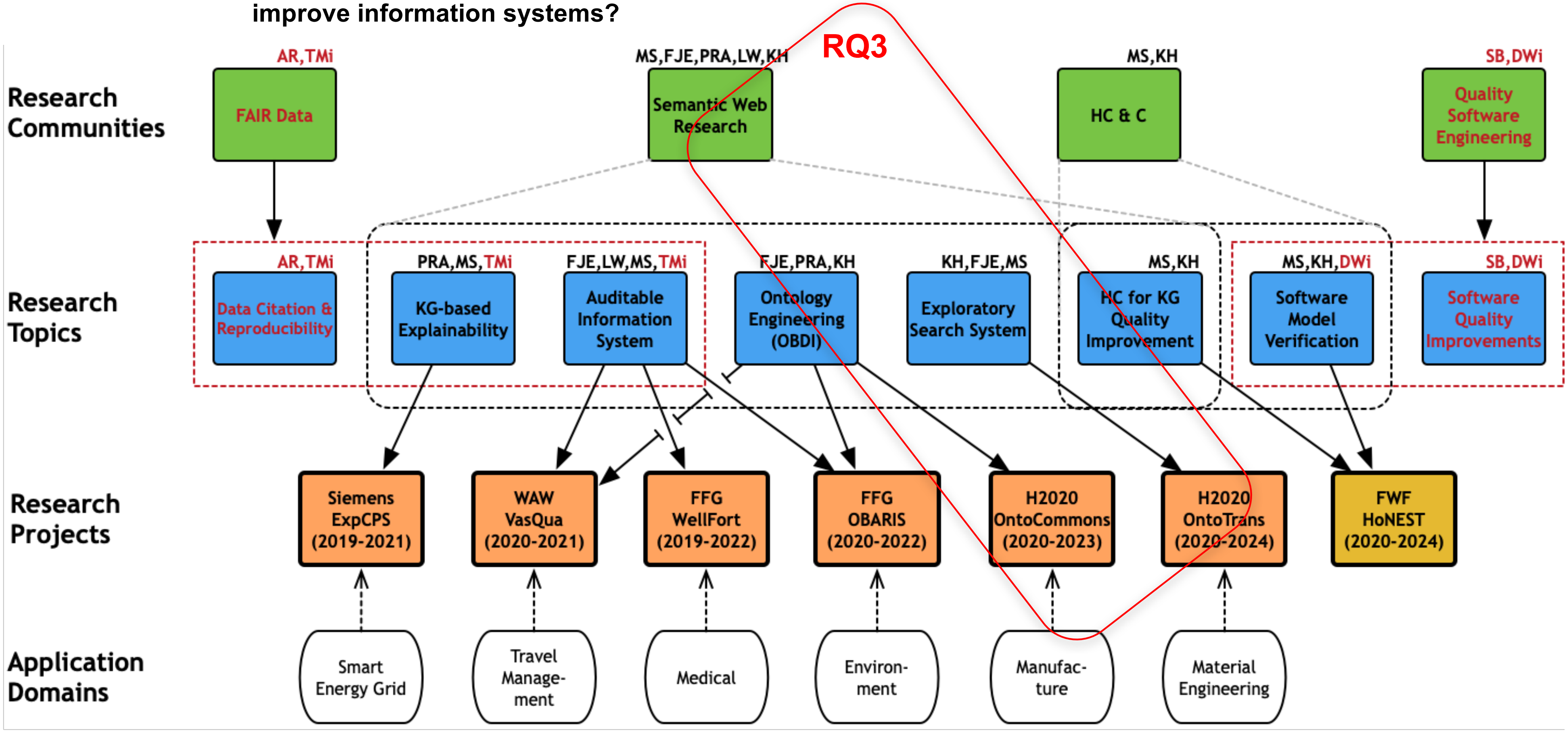
1. The node FiniteStateMachine exists.
2. The node InitialStateType exists.
3. There should not exist a FiniteStateMachine with more than one State of type InitialStateType

SIEMENS

FAKULTÄT FÜR !NFORMATIK

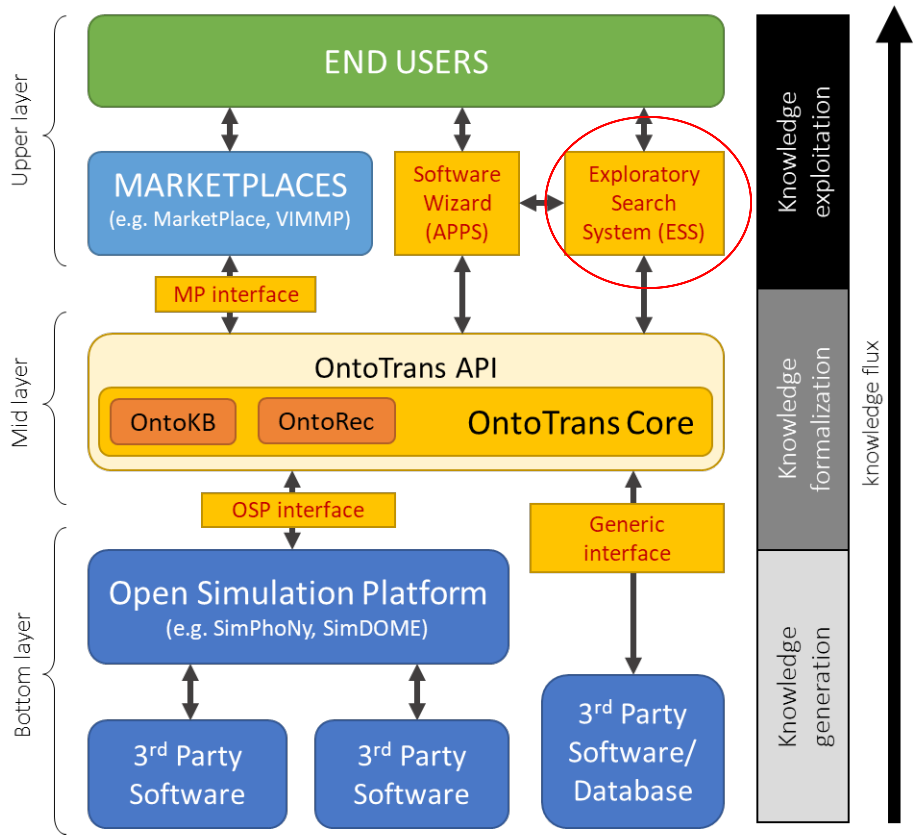
Research Areas and Projects

(3) How can semantic (web) technologies improve information systems?



OntoTrans

- EU - H2020
- 2020 - 2024



(A) representing manufacturing process challenges in a standard ontological form (User Case)

(A) connecting User Cases with existing appropriate information sources i.e. available data and materials modelling solutions

(A) recommending consistent materials modelling workflow options

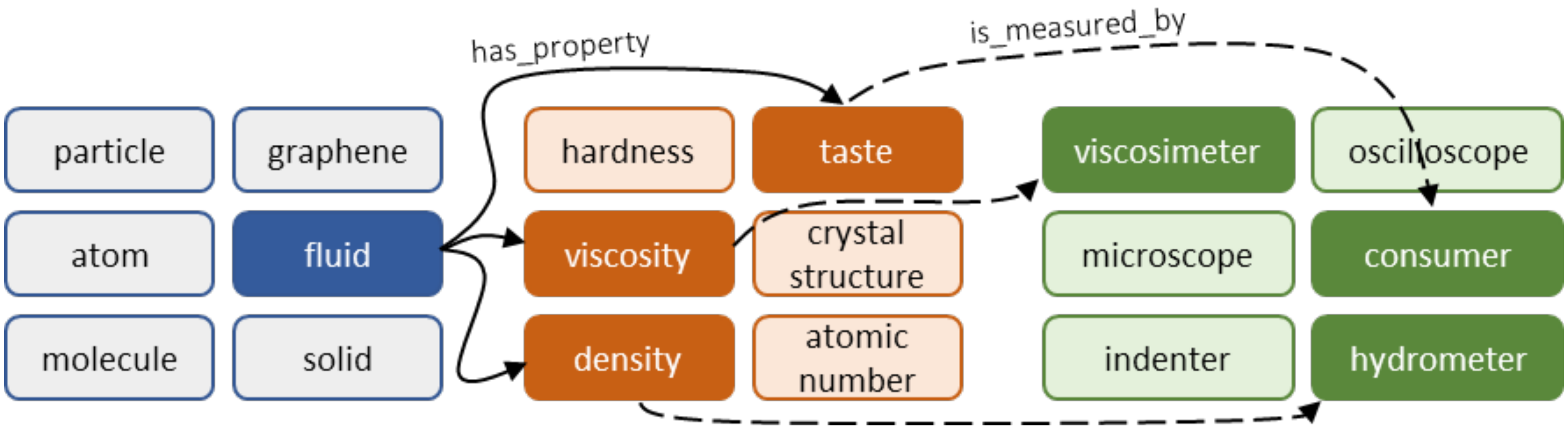
aim: support the development of dedicated Apps delivering a smart guidance for materials producers and product manufacturers

Industrial Partners:

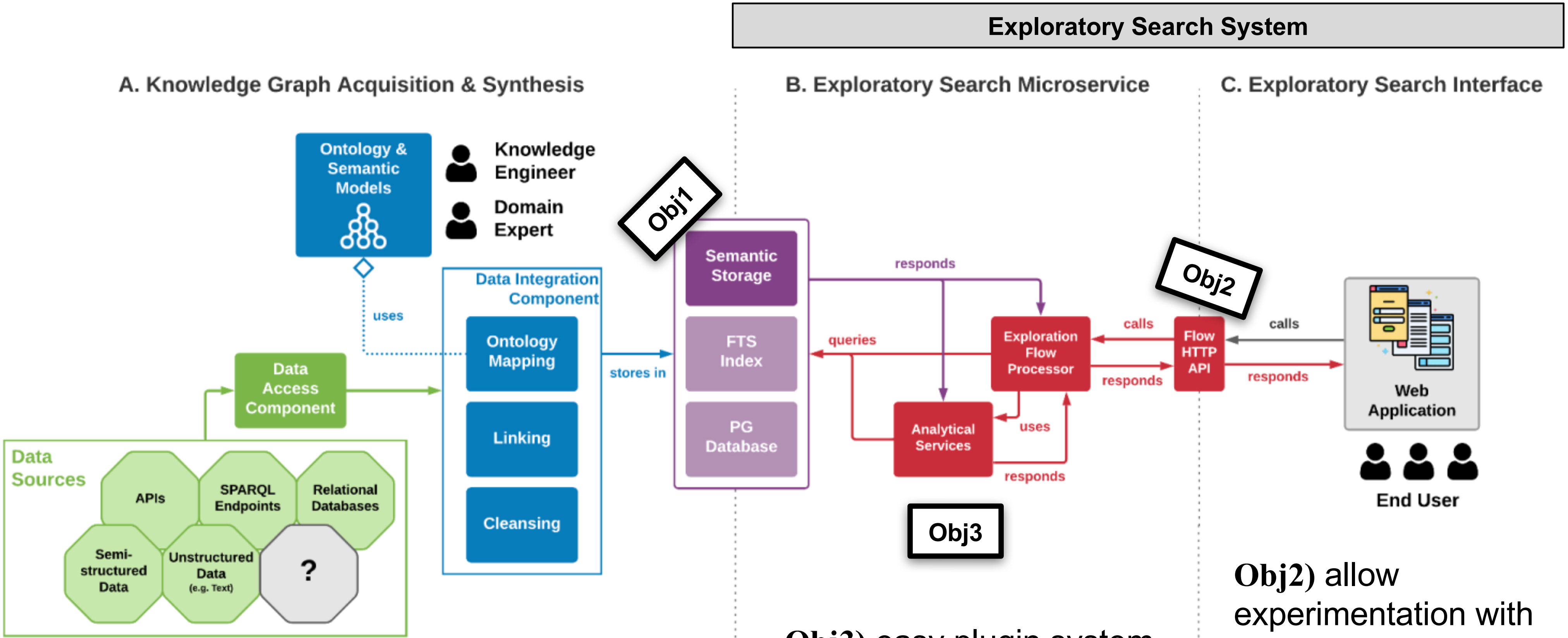
P&G Composites Evolution ArcelorMittal

Exploratory Search System [ESS] OntoTrans

End users and translators may not have a clear view of the information they need but rather require support to discover new information and learn about the domain. OntoTrans will support them with an ESS which will rely on data science algorithms to identify emerging patterns in the available semantic information.



Exploratory Search Platform



Obj1) adaptable to different use cases and legacy environments (without costly changes)

Obj3) easy plugin system for new algorithms for knowledge graph analytics

Obj2) allow experimentation with different interface paradigms (without costly changes)

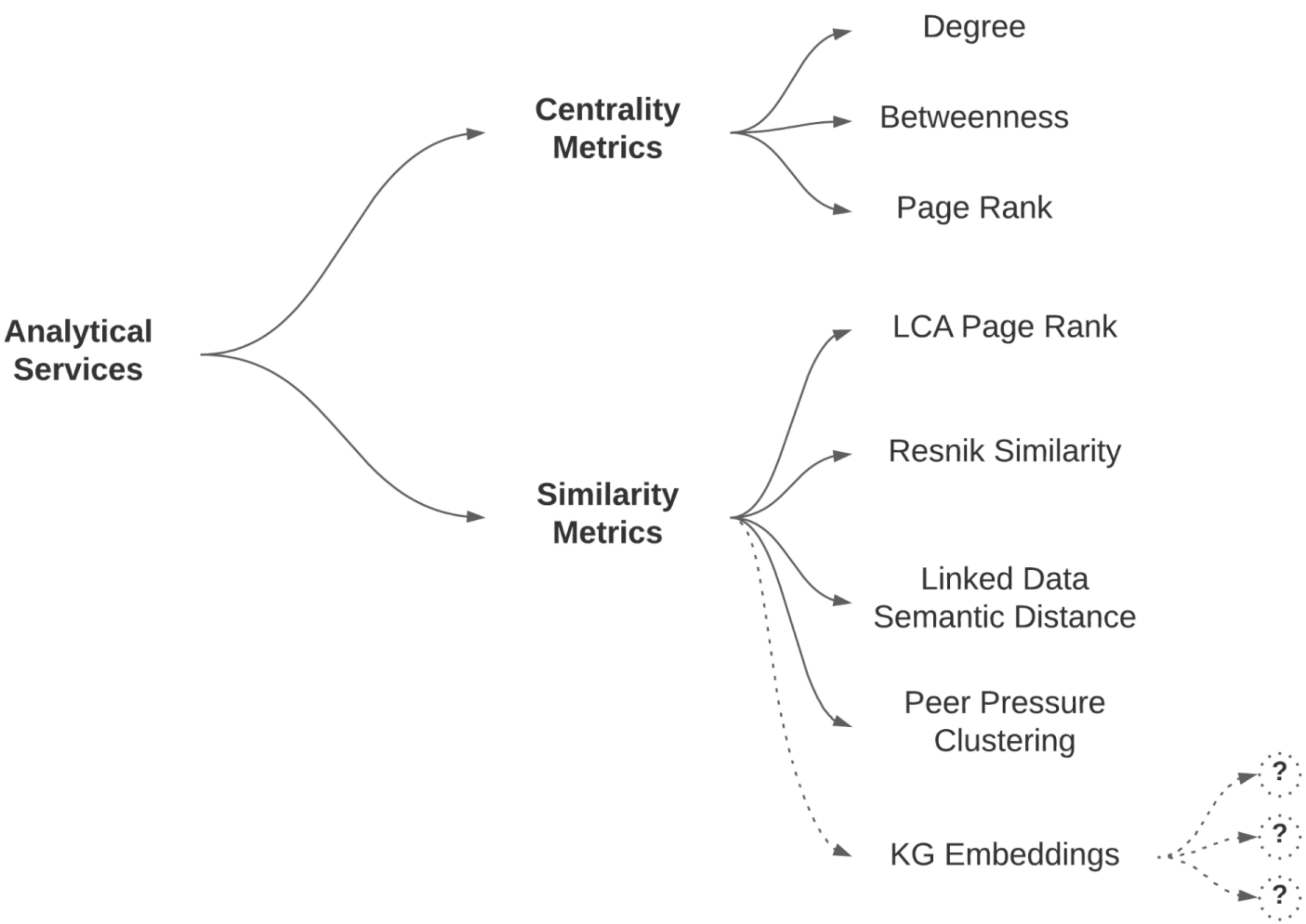
Obj3) Analytical Pipeline

- **Analytical Services** analyze the (semantic) structure of a knowledge graph
  - can make use of:
    1. SPARQL
    2. Gremlin
    3. Full-Text Search Index
    4. Other Analytical Services
- new services can be plugged into the application

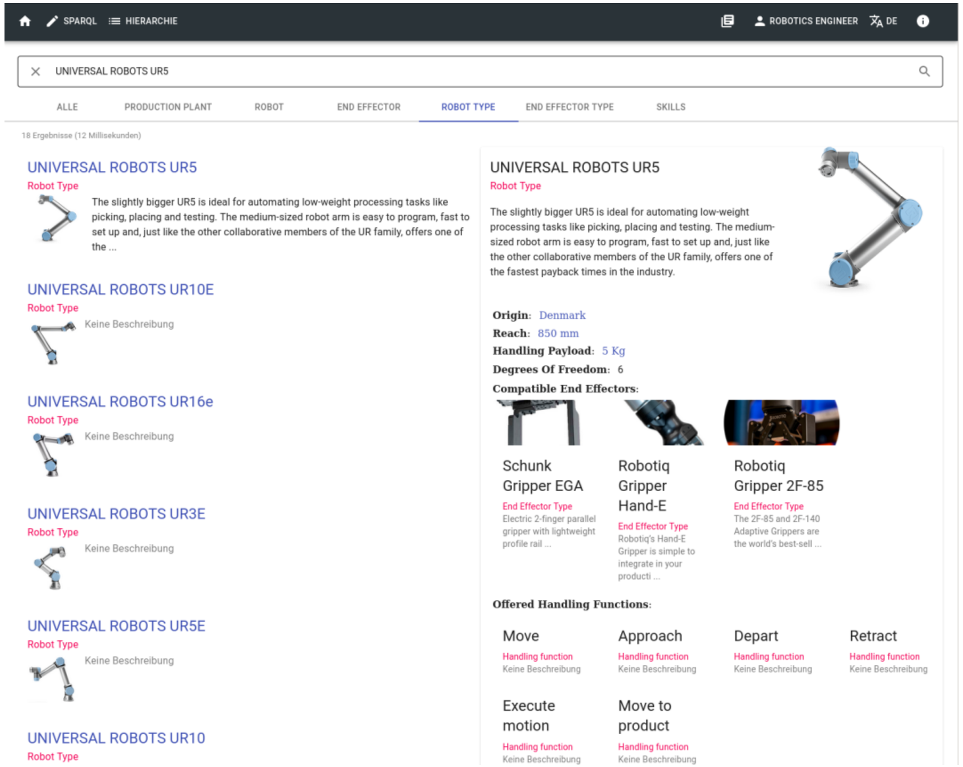
Centrality metrics assign an "importance" value to entities in the knowledge graph.

Similarity metrics assign a "similarity" value to a pair of entities in the knowledge graph.

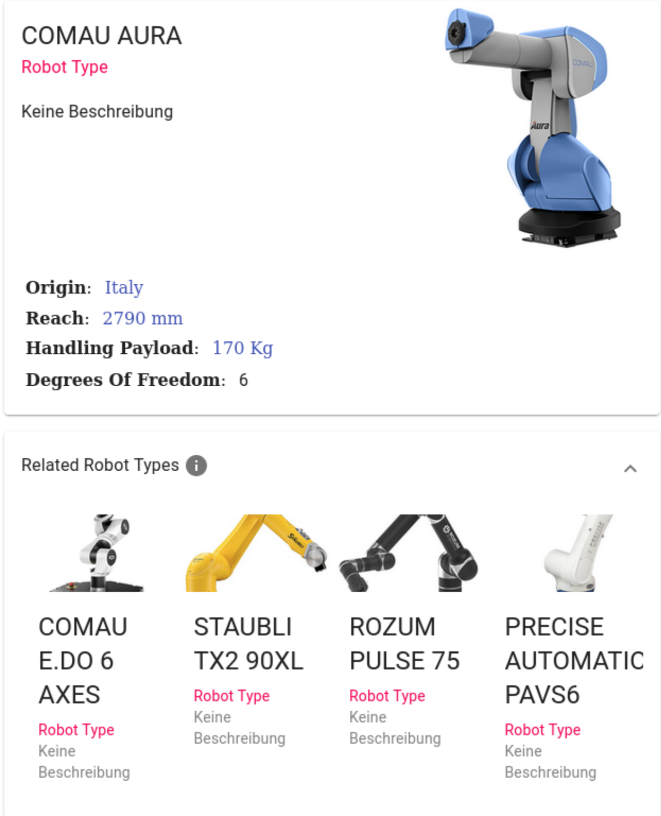
Analytical Services



Exploratory Search Interface (UC1)



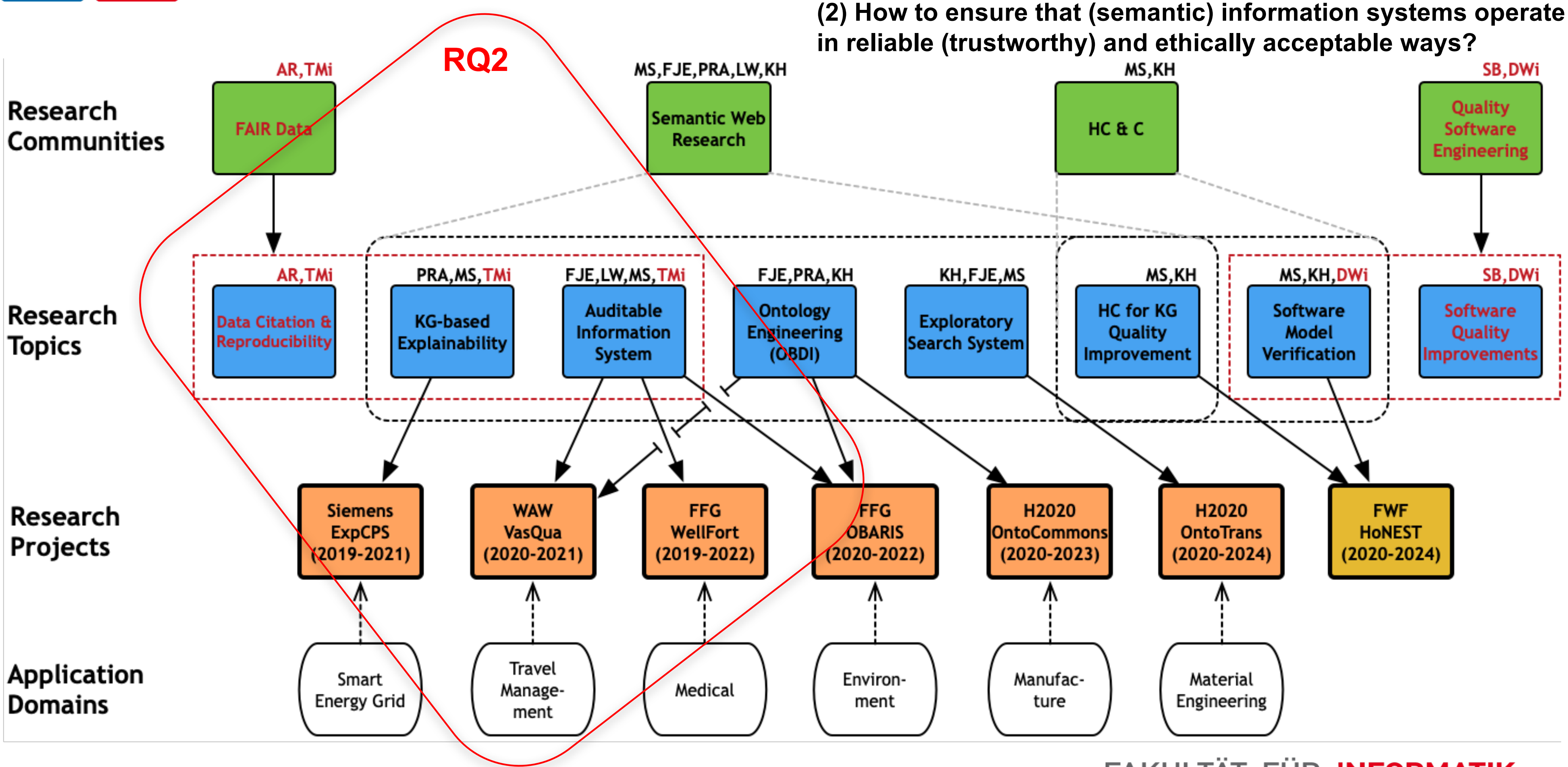
Keyword Search



Recommendations/ Infobox

Research Areas and Projects

(2) How to ensure that (semantic) information systems operate in reliable (trustworthy) and ethically acceptable ways?



Ontology-Based Artificial Intelligence Systems in environmental Sector [OBARIS]

- FFG IKT der Zukunft
- 2020-2022

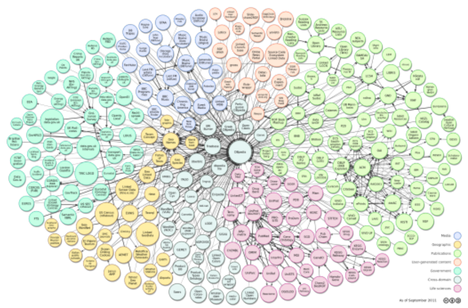
How to design and implement Auditable Semantic AI Systems (SAIS)?

- **G1: Provide a taxonomy and characterization of SAIS**
- G2: Establish a technology stack for SAIS
- G3: Provide data integration & management methods for SAIS
- **G4: Create an auditability framework for SAIS**
- G5: Introduce and evaluate SAIS stack in concrete environmental use cases

Context: 3rd Wave of AI

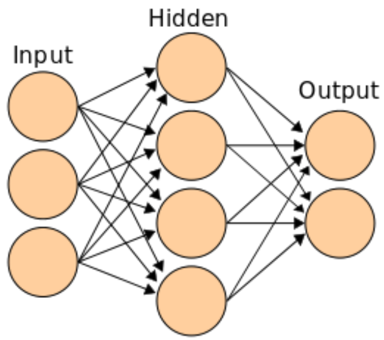
G1: Provide a taxonomy and characterization of SAIS

1. Semantic Web



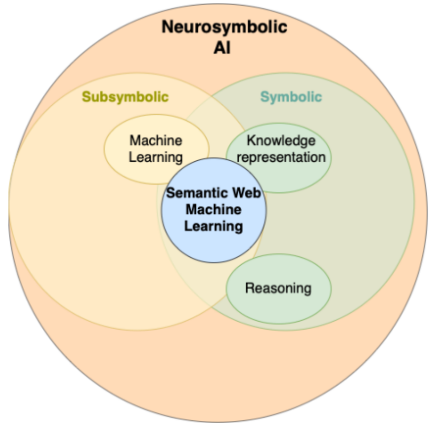
+ symbolic knowledge
+ explainability
- vulnerable to noisy data
- knowledge acquisition bottleneck

2. Machine Learning



+ knowledge from sparse data
+ broad applicability
- intransparency
- often lots of training data required

3. Semantic Web Machine Learning



Focus on **semantic web** resources and m**achine learning** approaches

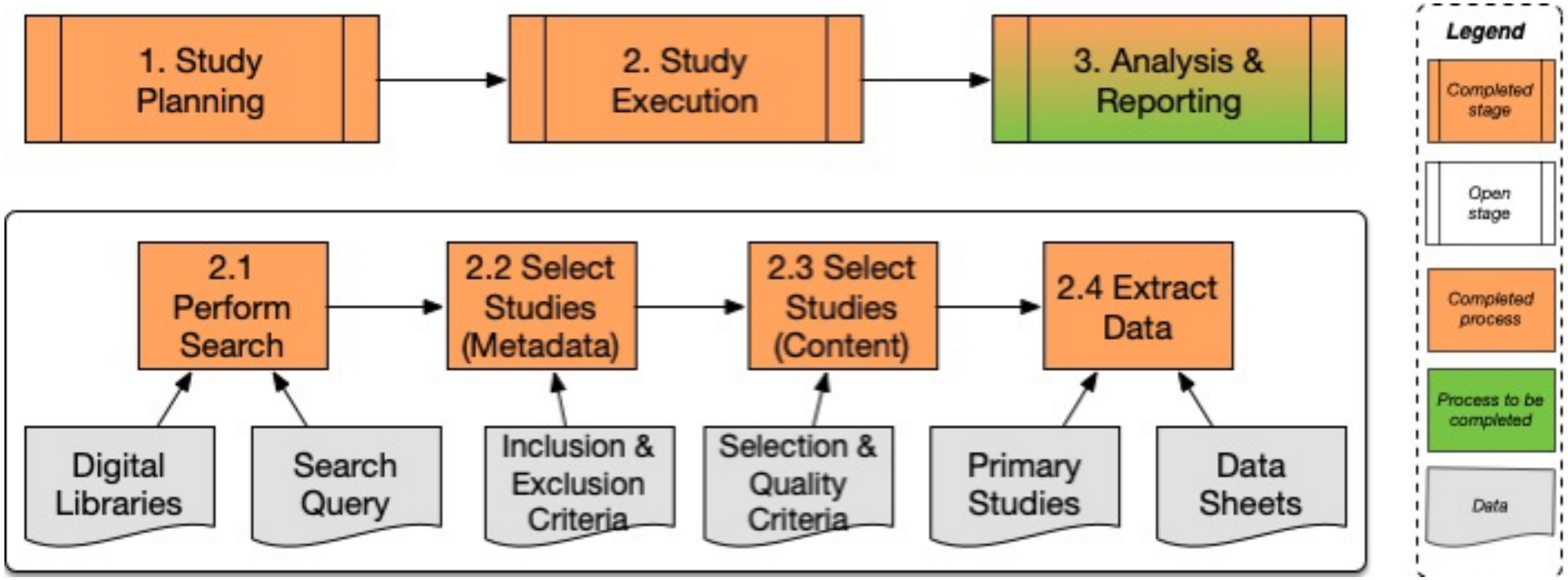
= SWeMLS, subset of neurosymbolic AI

Systematic Mapping Study: Methodology

G1: Provide a taxonomy and characterization of SAIS

Query Date: October 2020

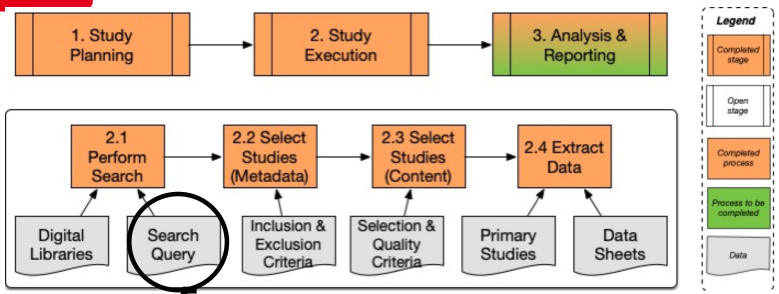
Research Team: 10 participants



Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.

Search Queries

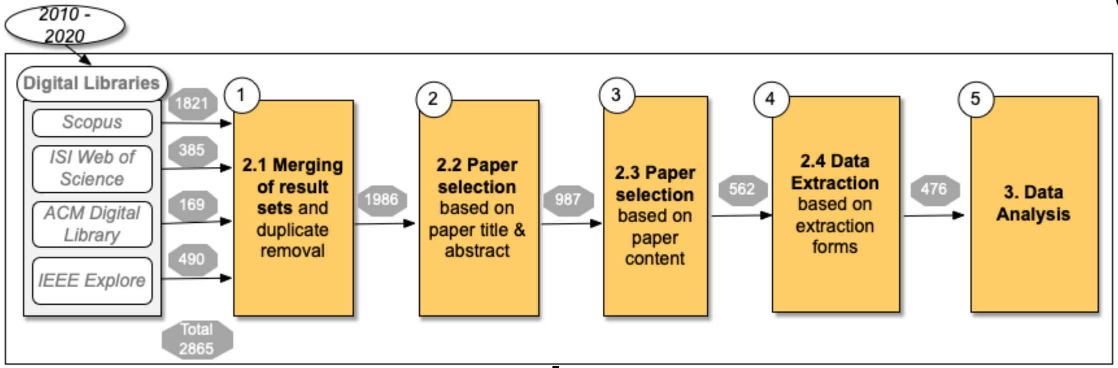
G1: Provide a taxonomy and characterization of SAIS



Sub-Query	Used Search Keywords
Q1 (SW module)	knowledge graph, linked data, semantic web, ontolog*, RDF, OWL, SPARQL, SHACL
Q2 (ML module)	deep learning, neural network, embedding, representation learning, feature learning, language model, language representation model, rule mining, rule learning, rule induction, genetic programming, genetic algorithm, kernel method
Q3 (System)	Natural Language Processing, Computer Vision, Information Retrieval, Data Mining, Information Integration, Knowledge Management, Pattern Recognition, Speech Recognition

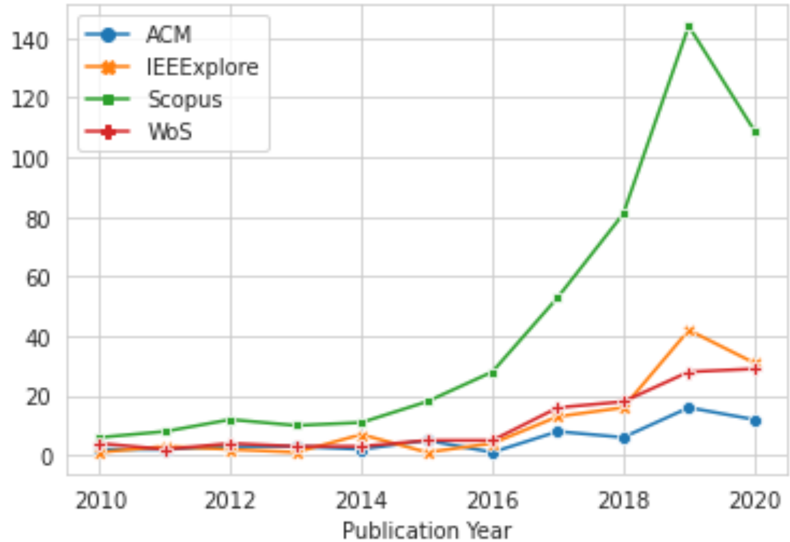
Systematic Mapping Study: Result Set

G1: Provide a taxonomy and characterization of SAIS



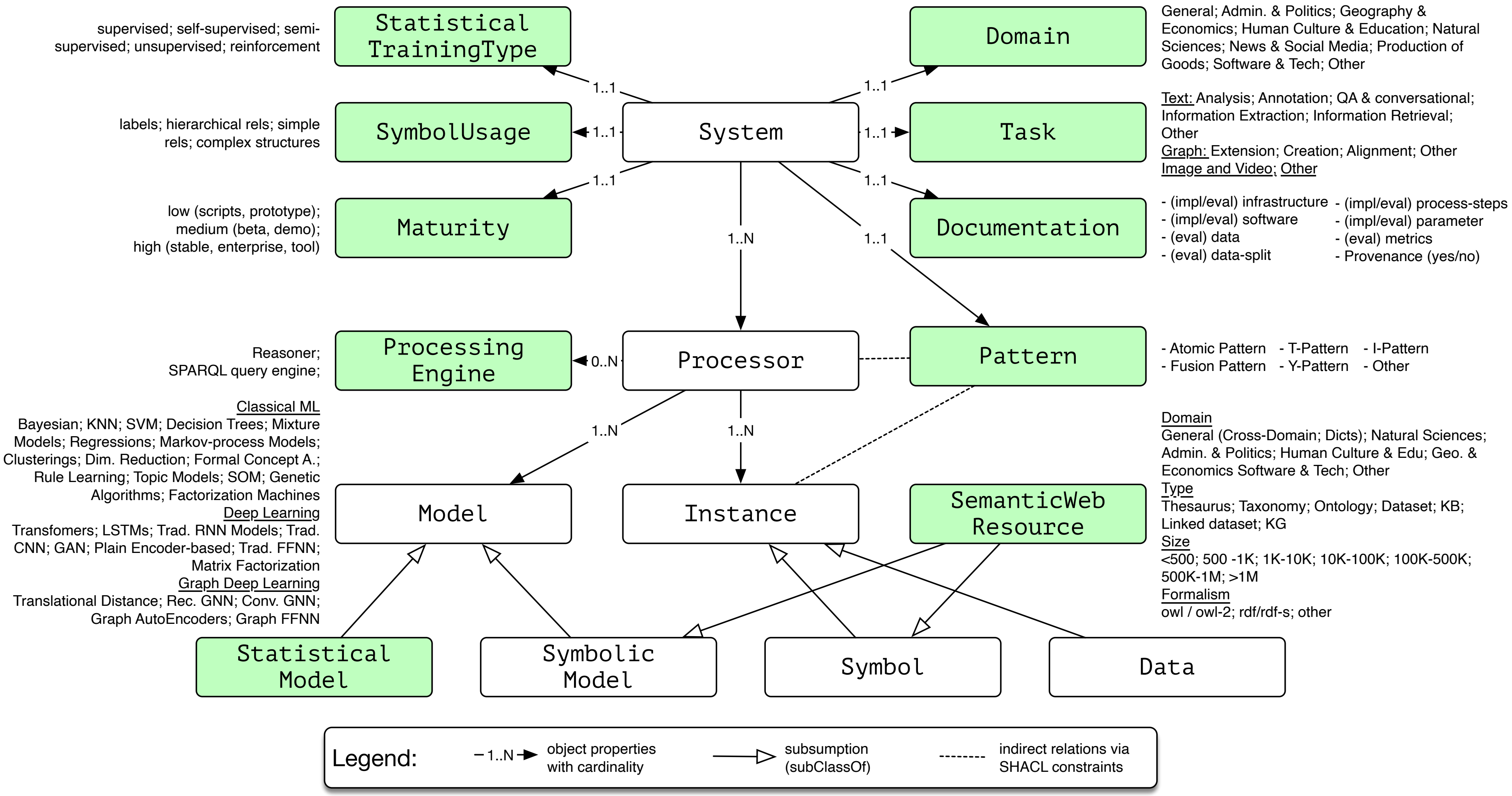
Bibliographic analysis:

1) Scopus accounts for largest portion
2) Strong increase of papers over last years



SWeML System Classification

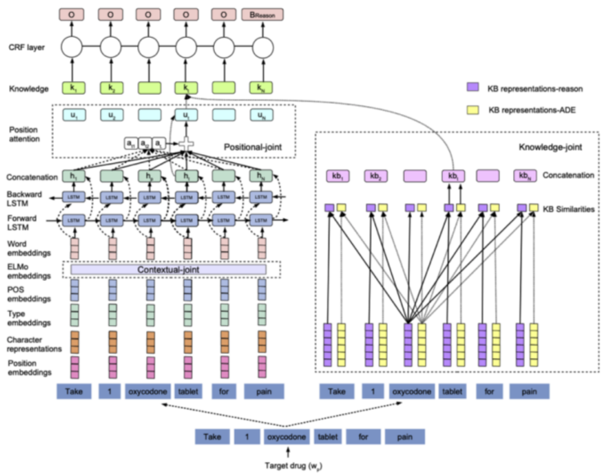
G1: Provide a taxonomy and characterization of SAIS

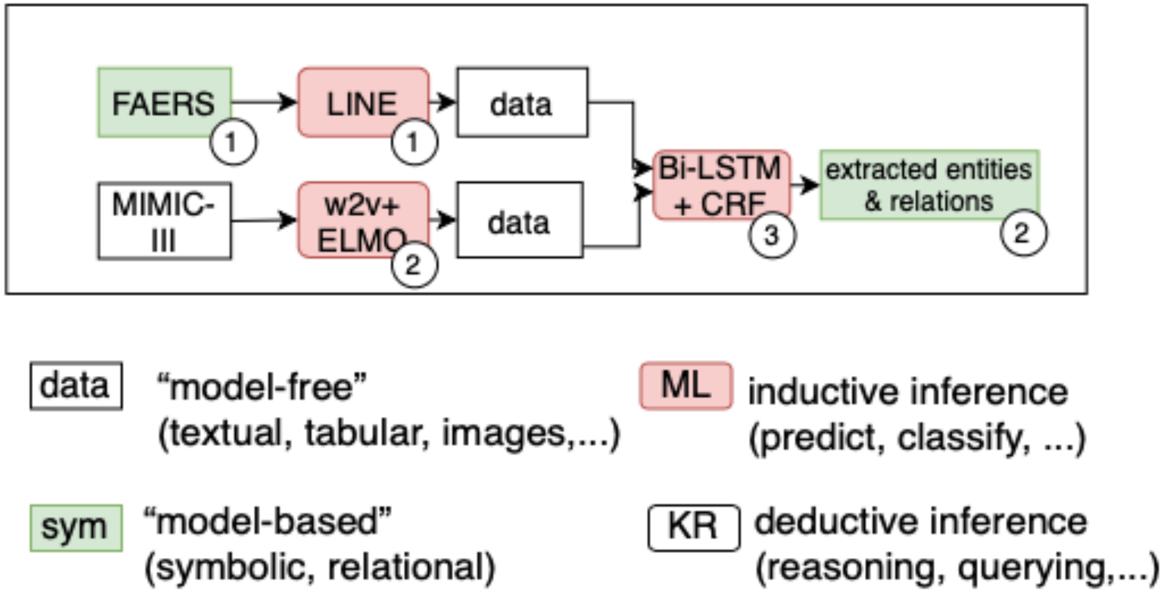


SWeML System Design Patterns

G1: Provide a taxonomy and characterization of SAIS

- Based on [1], compositional design patterns for hybrid systems
- SMS retrieved papers beyond these patterns, e.g. [2]



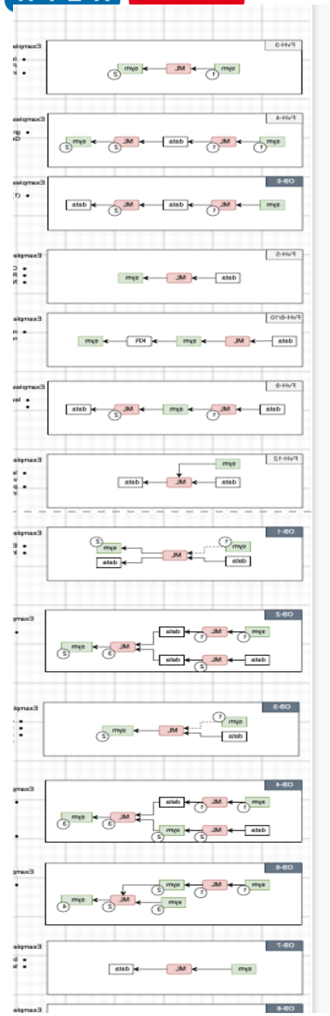


[1] Van Harmelen, F., & Teije, A. T. (2019). A boxology of design patterns for hybrid learning and reasoning systems*: Journal of Web Engineering, Vol 18, pp. 97-124, 2019.*

[2] Dandala, B., Joopudi, V., Tsou, C. H., Liang, J. J., & Suryanarayanan, P. (2020). Extraction of Information Related to Drug Safety Surveillance From Electronic Health Record Notes: Joint Modeling of Entities and Relations Using Knowledge-Aware Neural Attentive Models. *JMIR medical informatics, 8*(7), e18417.

Outlook

G1: Provide a taxonomy and characterization of SAIS

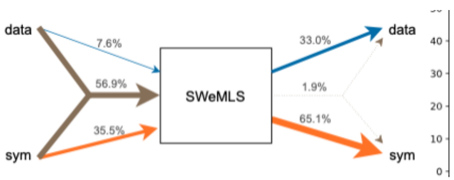


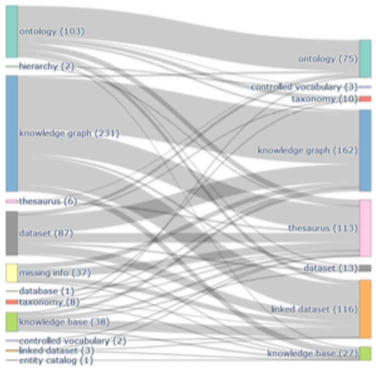
Follow for more detailed analysis in upcoming publications:

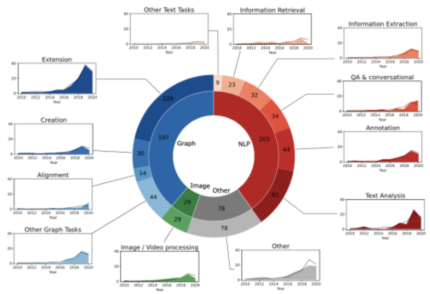
http://semsys.ifs.tuwien.ac.at

http://www.obaris.org

https://twitter.com/LaWaltersdorfer

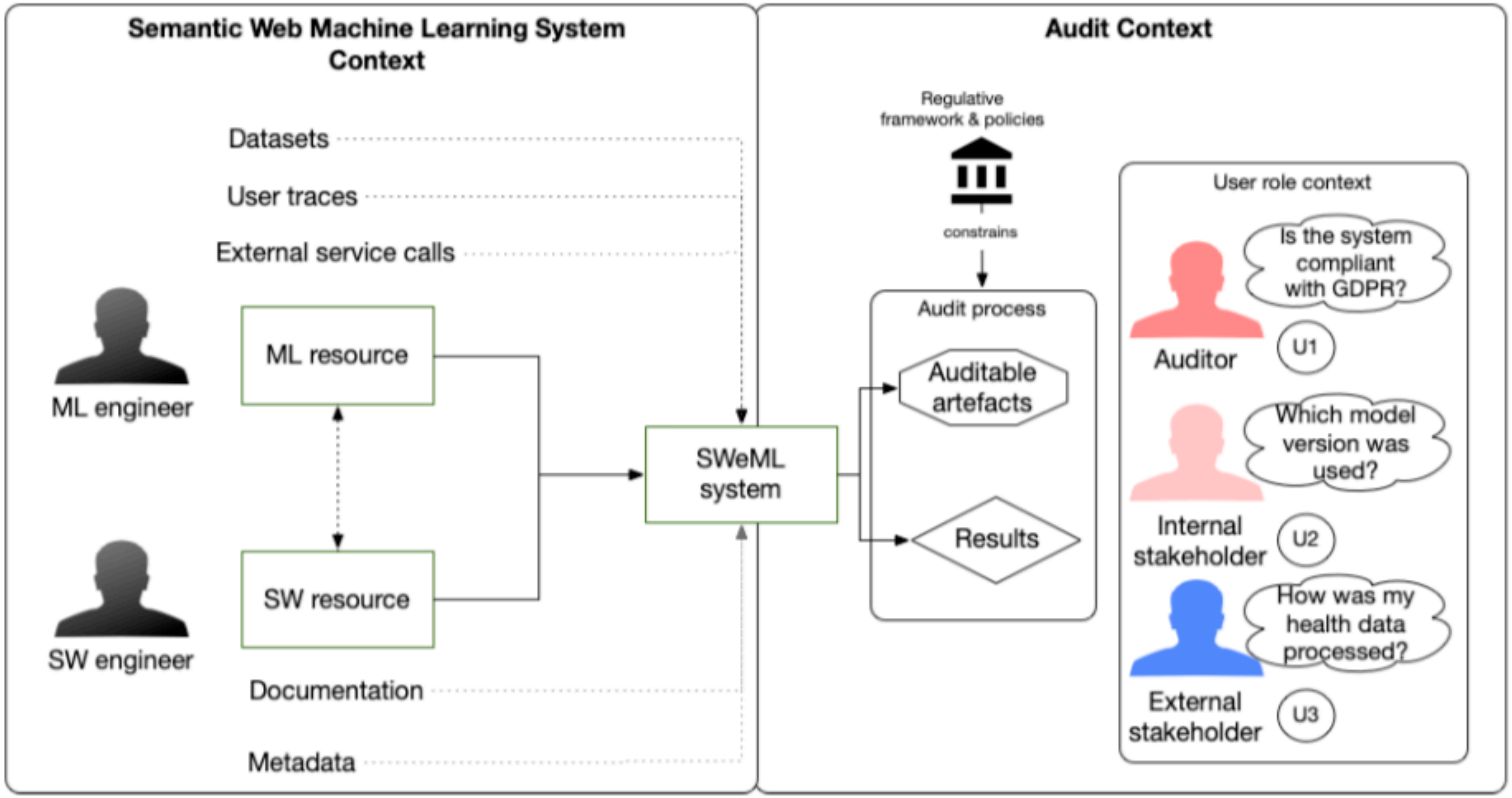






Auditing SWeML

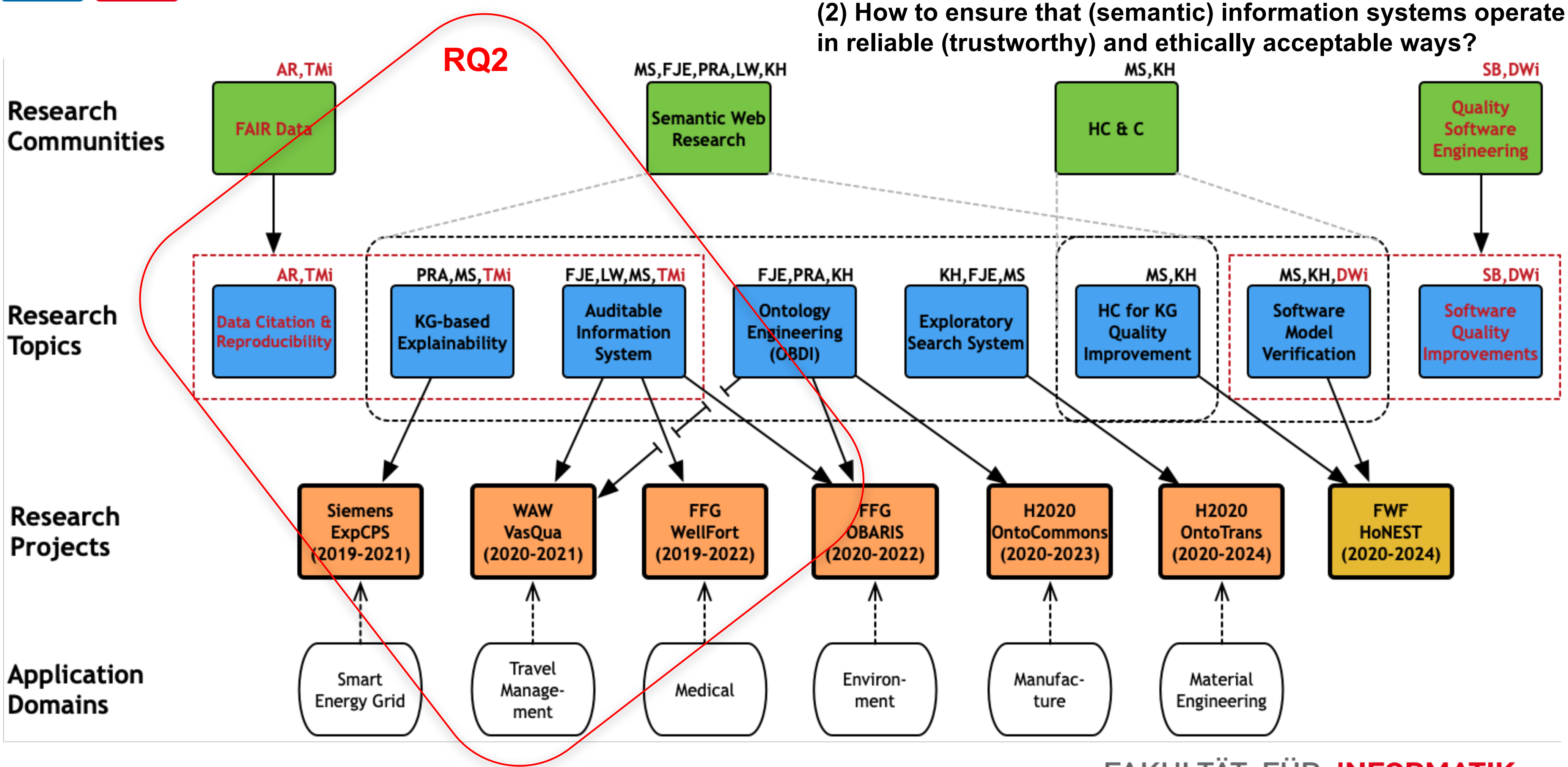
G4: Create an auditability framework for SAIS



Ongoing PhD work of Laura Waltersdorfer

Research Areas and Projects

(2) How to ensure that (semantic) information systems operate in reliable (trustworthy) and ethically acceptable ways?



FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Explainable Cyber-Physical Systems (ExpCPS)

- Siemens AT
- 2019-2021

Explainable Cyber Physical Energy Systems based on Knowledge Graph

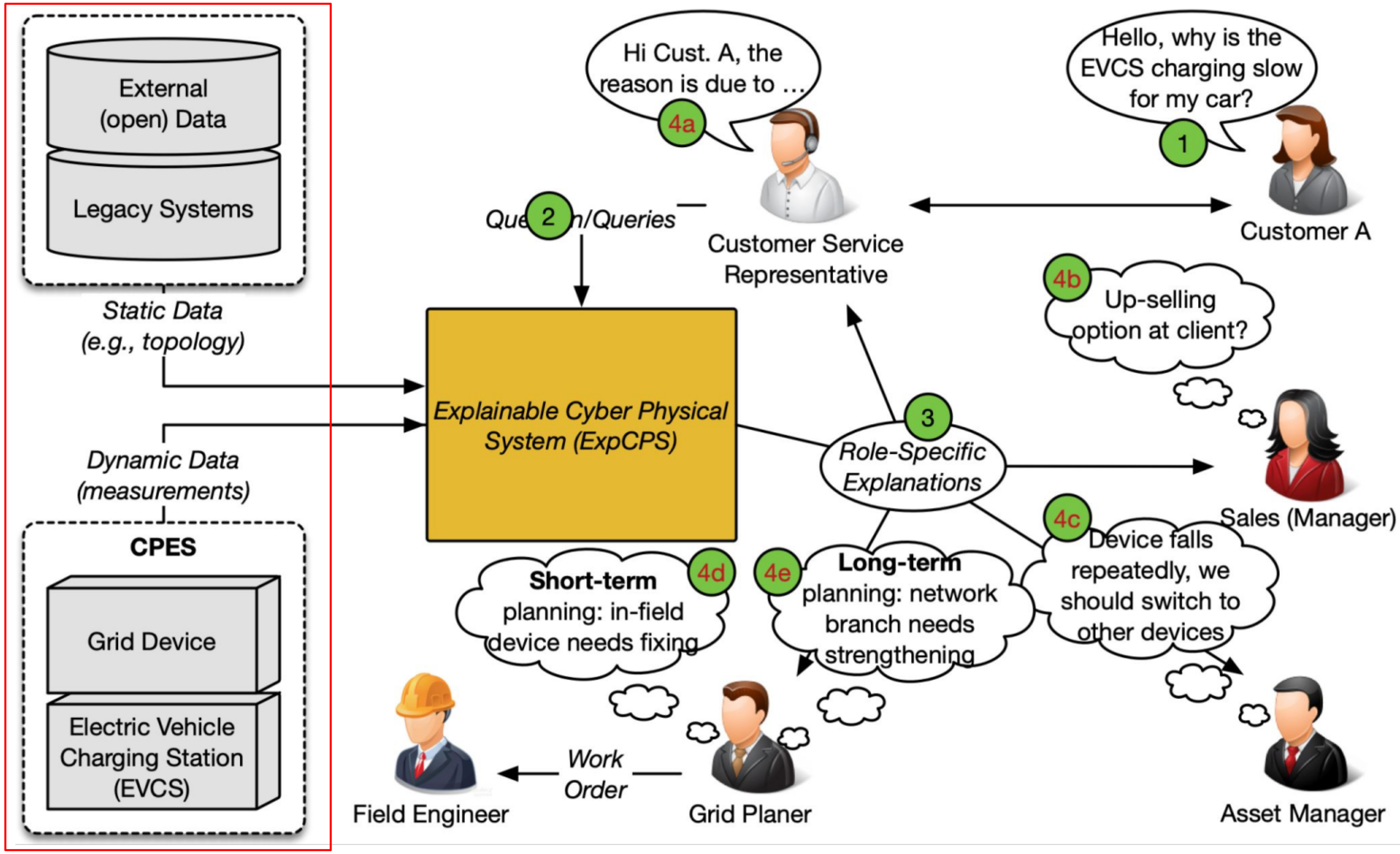
Peb Aryan, Fajar Ekaputra, Marta Sabou, Daniel Hauer Ralf Mosshammer, Alfred Einhalt, Tomasz Miksa, Andreas Rauber

Information & Software Engineering Research Group,
Institute of Information Systems Engineering,
Faculty of Informatics - TU Wien
Siemens AG

MSCPES 2021 18.05.2021

funded by the Austrian Research Promotion Agency (FFG)
in the PoSyCo project (FFG No. 3036508)

Motivation



Explanations could support the work of several Smart Grid stakeholders:

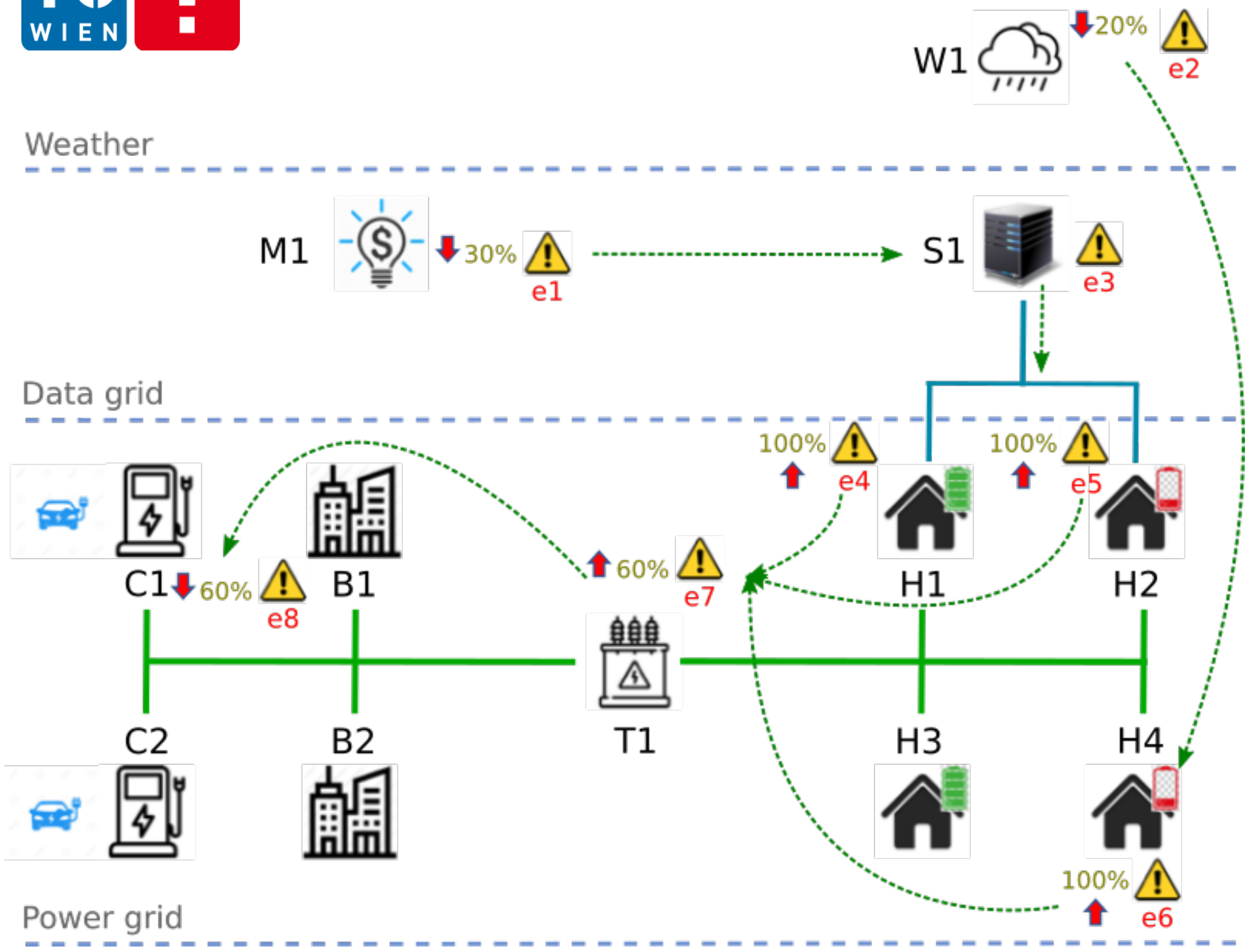
- Customers
- Grid operators (sales, operations, planning)
- Third parties (EVCS operators)

Explanations are the result of combining data from various sources:

- CPES assets sensors
- Open data (weather)
- Legacy systems

Aryan et.al. Simulation Support for Explainable Cyber-Physical Energy Systems. MSCPES 2020

Flexibility Scenario



Features

M1 : energy market
S1 : Flex application server
W1 : weather
H1-H4 : residential houses
B1-B2 : commercial buildings
C1-C2 : EV charging stations

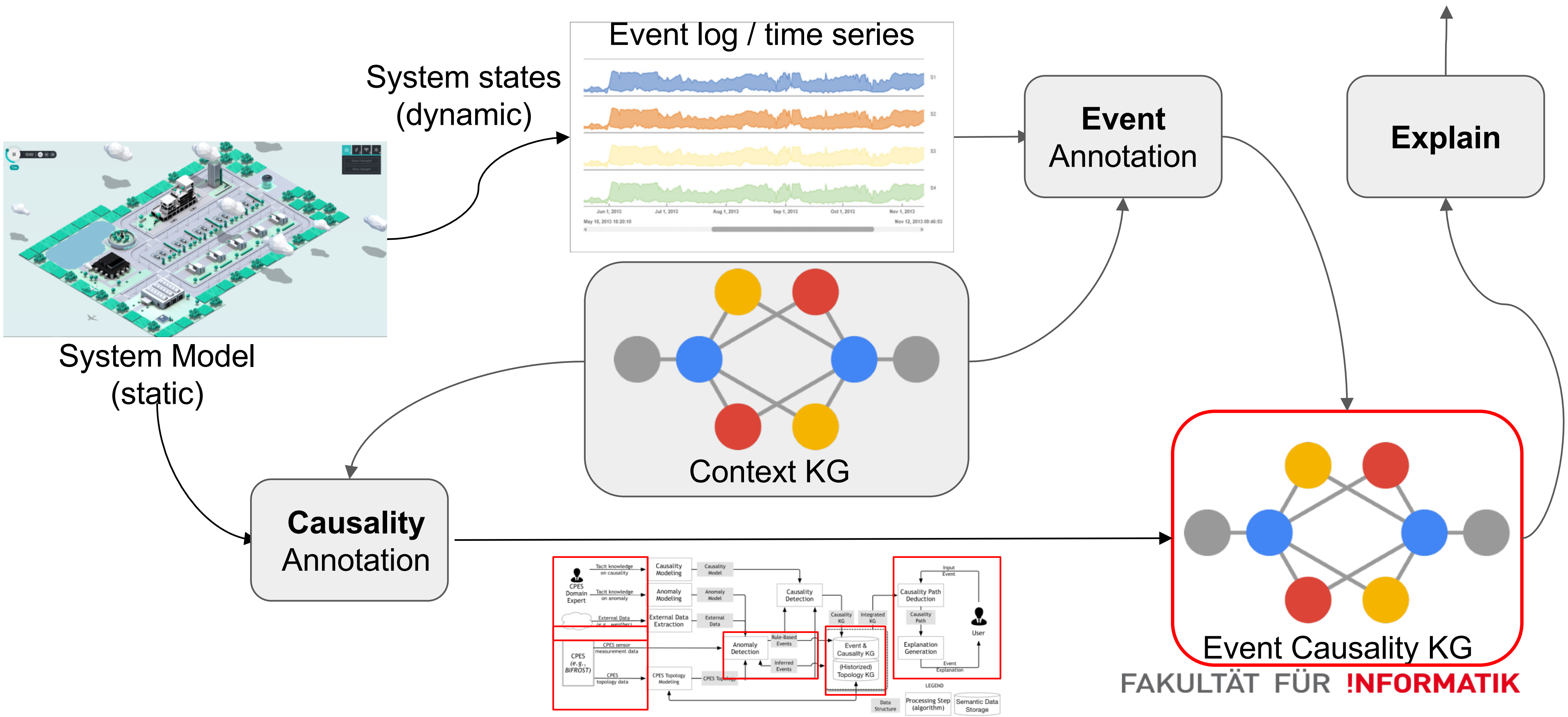
Events

e7 : trafo overload
e4,e5,e6 : high consumption
e3 : flex notification
e1 : energy market (real-time)

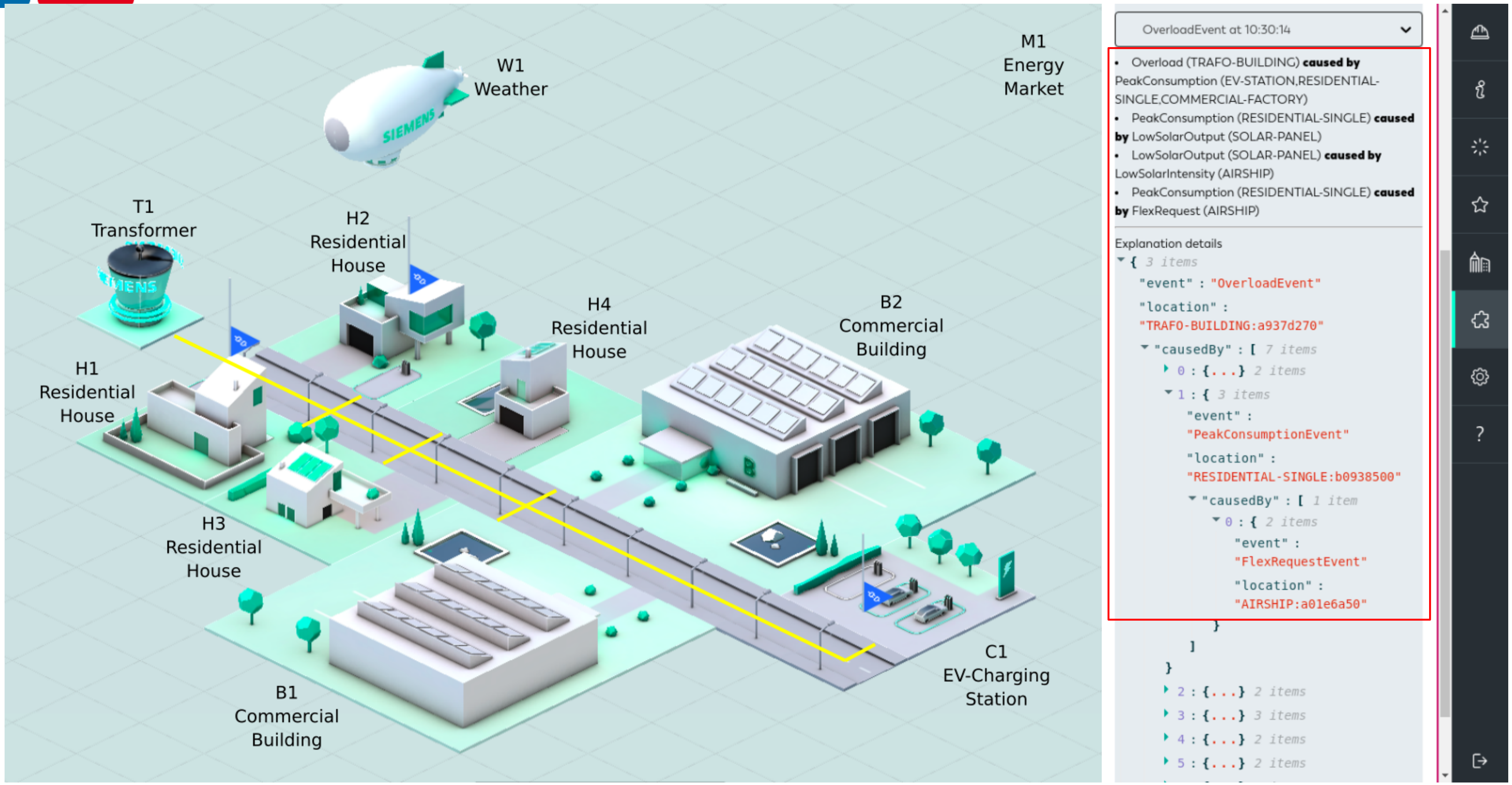
Example explanation

e7 because e(4,5,6)
e(4,5) because of e3; because of e1
e6 because of e2

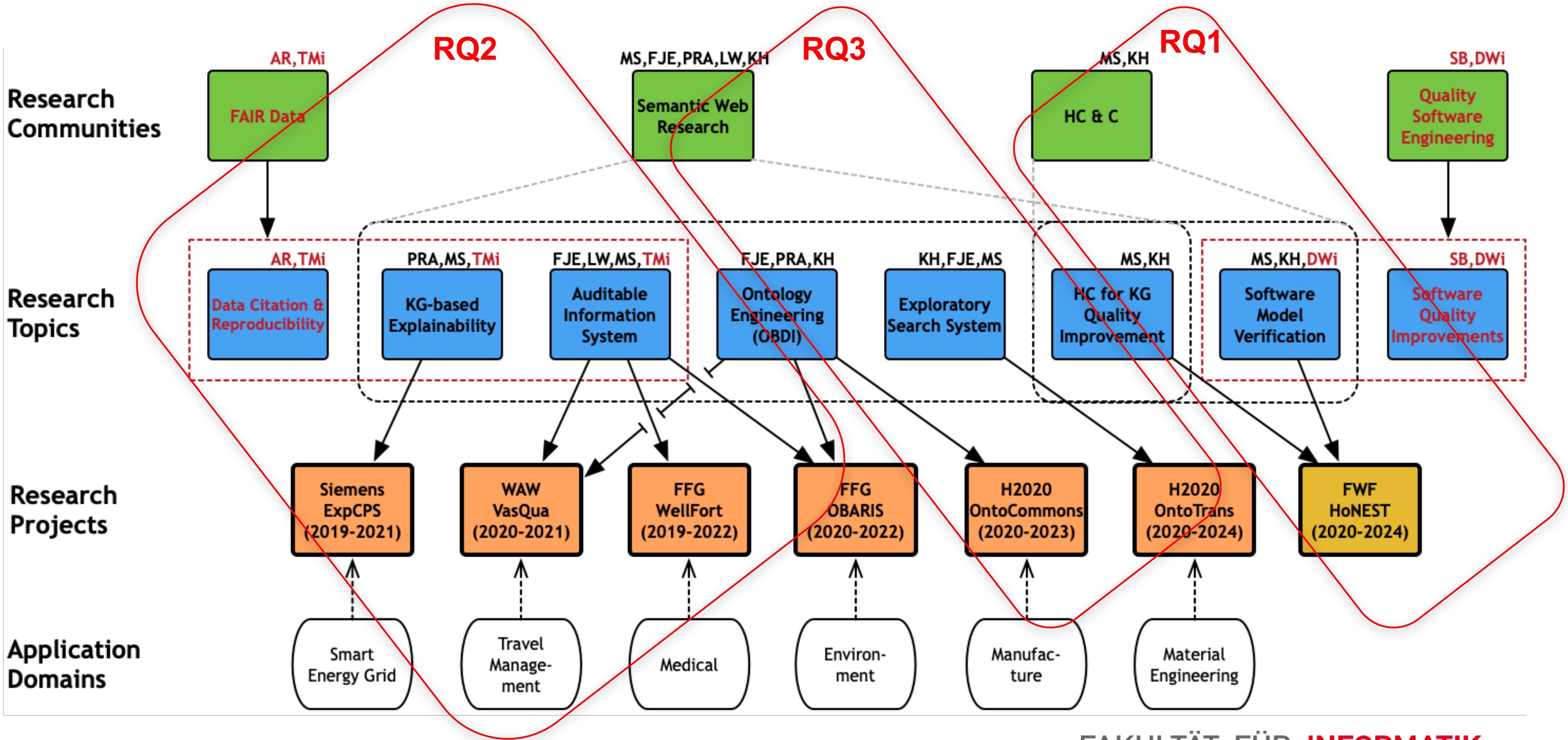
Explanation Generation Process



Implementation Result



Summary: Research Areas and Projects



FAKULTÄT
FÜR !NFORMATIK

Faculty of Informatics




Semantic Systems Research Lab

Thank you!

Happy 2022!