Empirical user studies in semantic web contexts

Catia Pesquita 1, Valentina Ivanova 2, Steffen Lohmann 3, Patrick Lambrix 4

1 LASIGE, Universidade de Lisboa, Lisbon, Portugal

2 RISE, Sweden

3 Fraunhofer IAIS, Sankt Augustin, Germany

4 Linkoping University, Linkoping, Sweden

EKAW 2018

International Workshop on Visualization and Interaction for Ontologies and Linked Data

VOILA! 2015

VOILA! 2016

VOILA! 2017

VOILA! 2018

Special Theme 2018:
Empirical Research

What is the state of **empirical evaluation** in SW contexts?

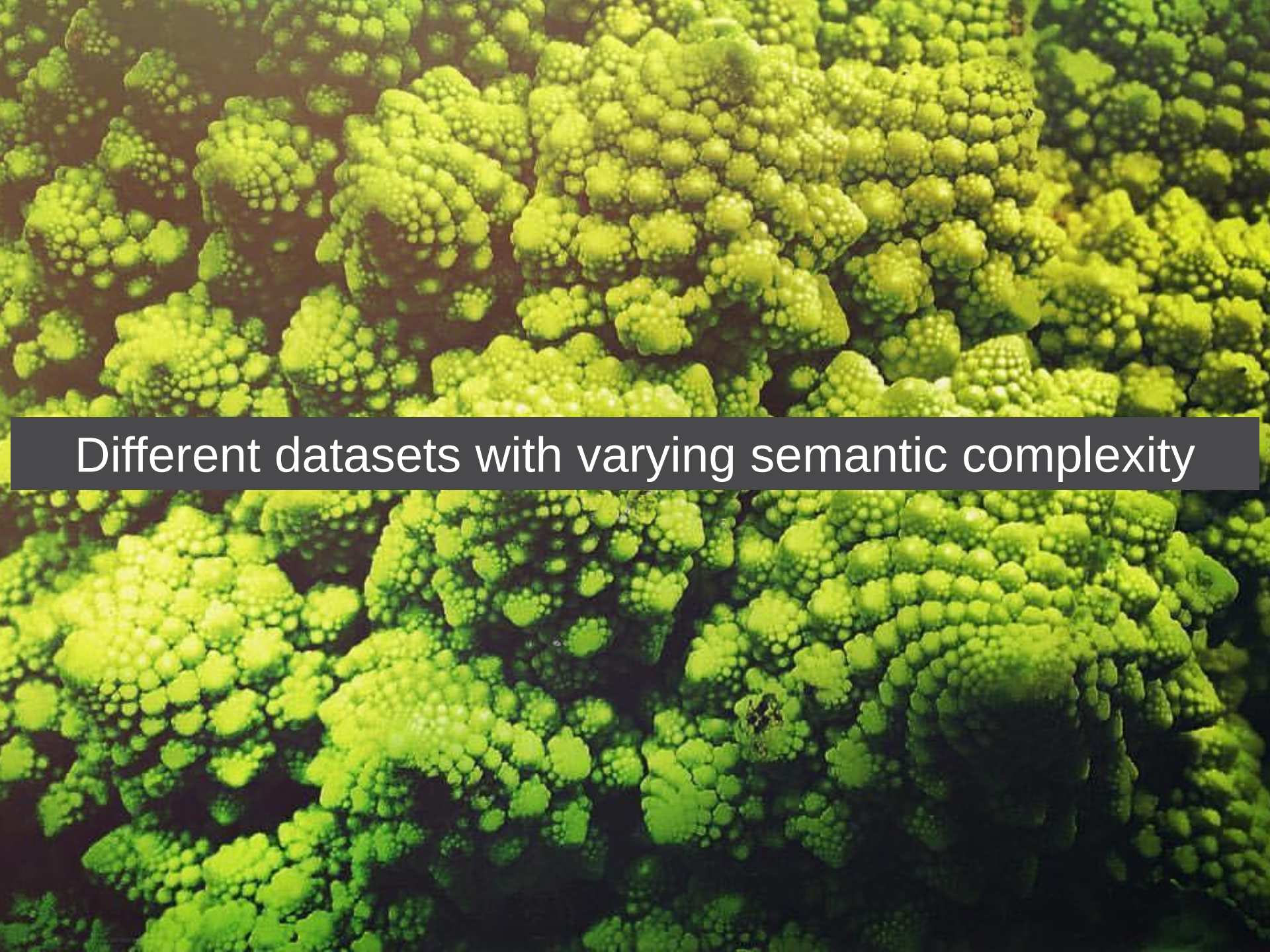
What is the state of **empirical evaluation** in SW contexts?

Can we help improve it?

Why are user studies **different** in Semantic Web?

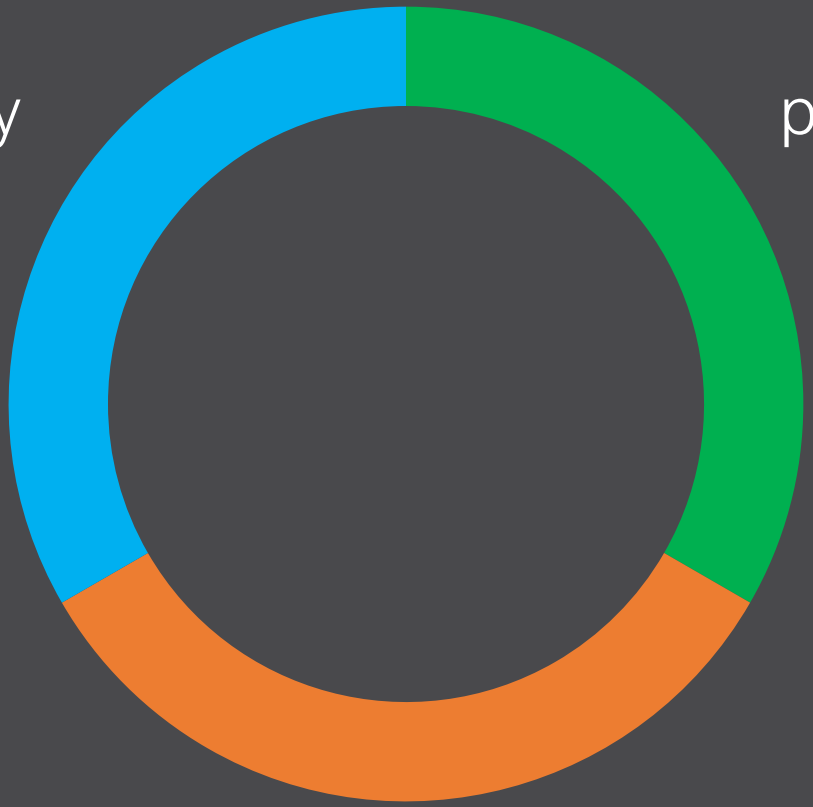
Knowledge Engineer vs. Domain Expert

Knowledge Engineering expert vs. novice


Different datasets with varying semantic complexity


Different users and datasets

external validity



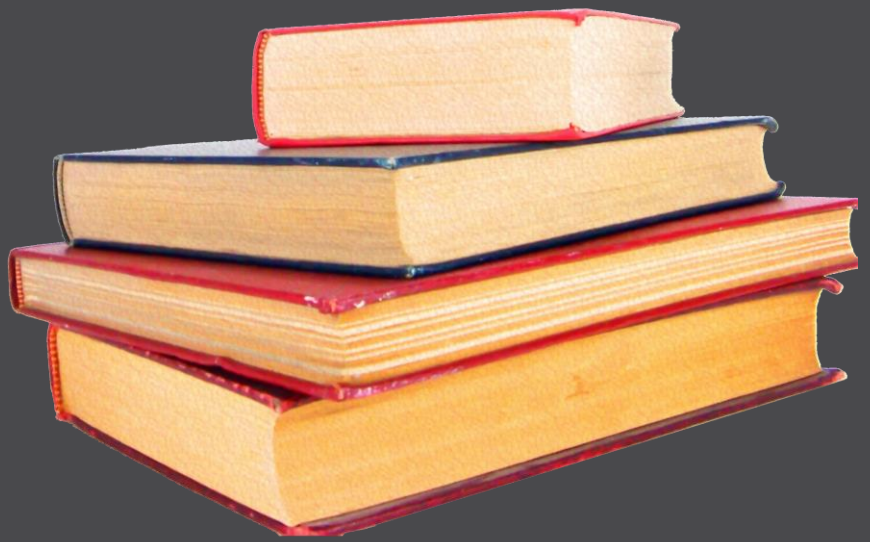
population validity

ecological validity

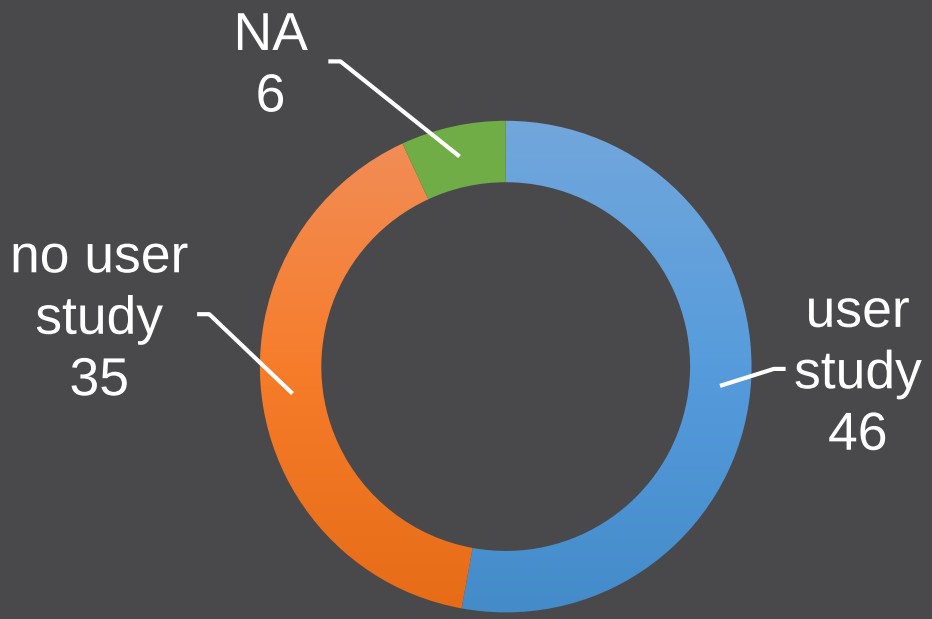
2015-2017

- ISWC (International Semantic Web Conference)
- ESWC (Extended Semantic Web Conference)
- VOILA (International Workshop on Visualization and Interaction for Ontologies and Linked Data)
- IESD (International Workshop on Intelligent Exploration of Semantic Data)

- “user study”
- “user evaluation”
- “empirical evaluation”
- “interaction”
- “visualization”



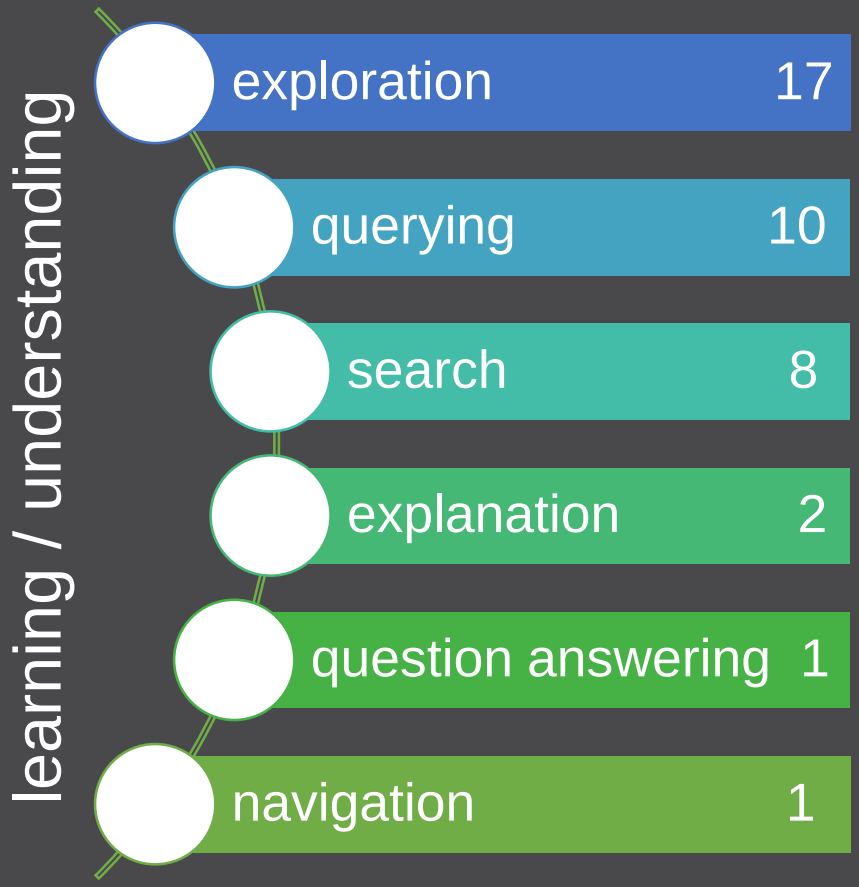
Literature review

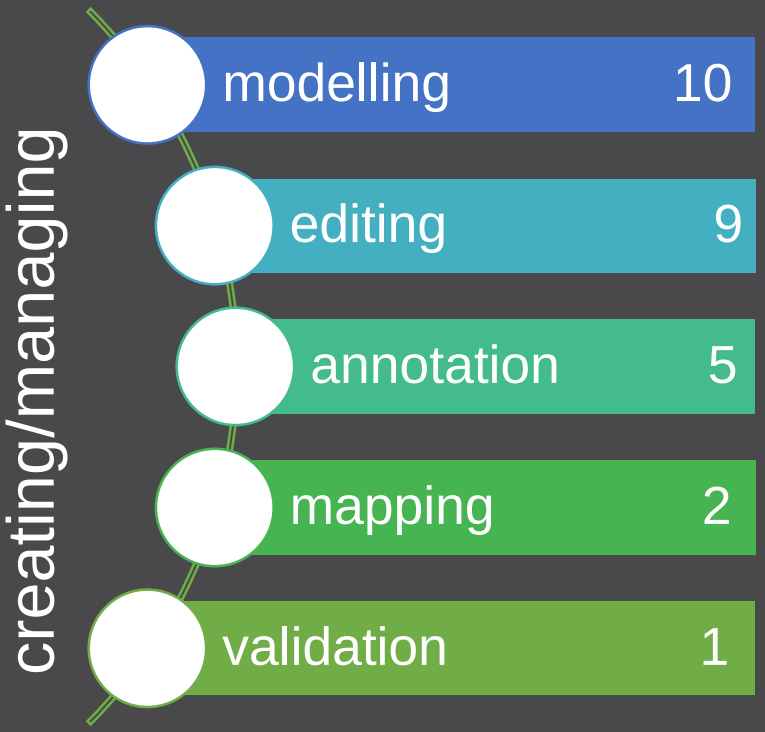


87 reviewed papers
46 user study papers were coded

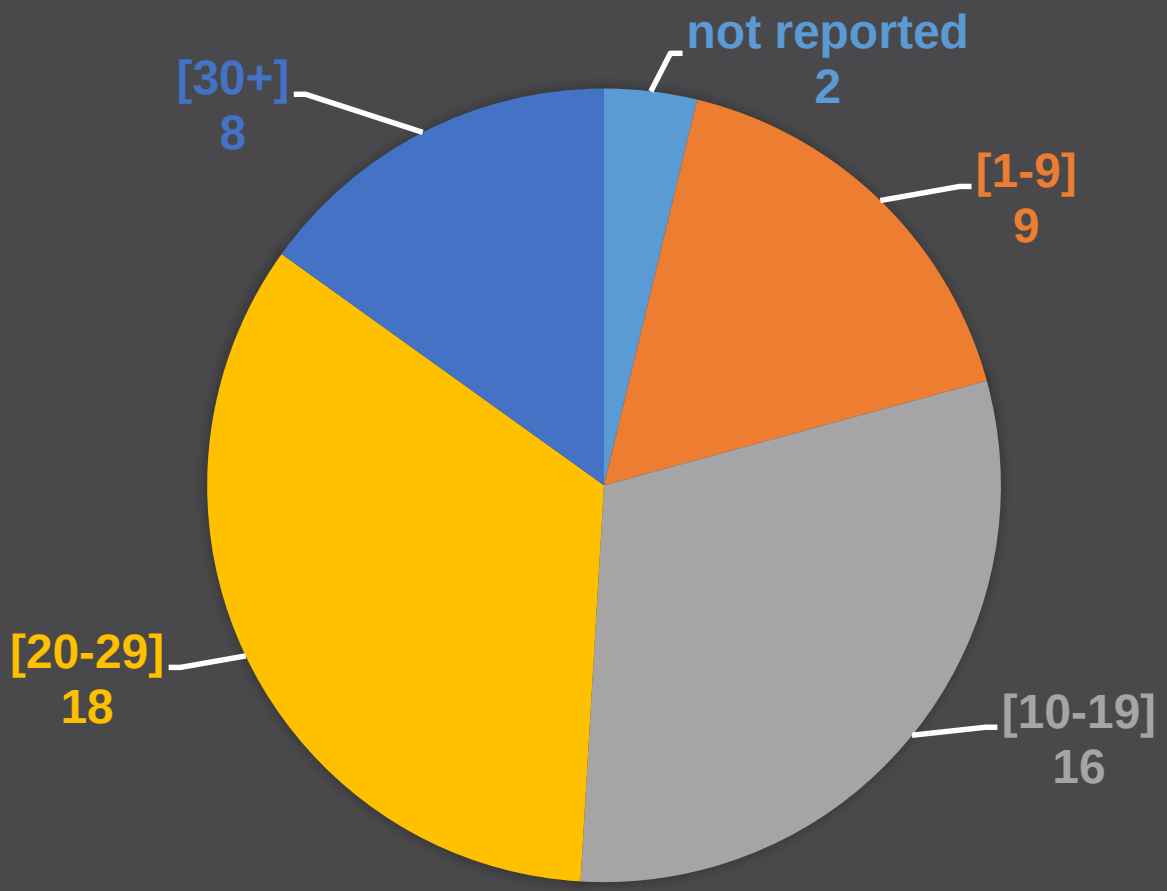
inductive analysis coding → purpose, users, evaluation methods → codes shared by all researchers → list of codes is revised → assigned codes checked by second researcher

Purpose

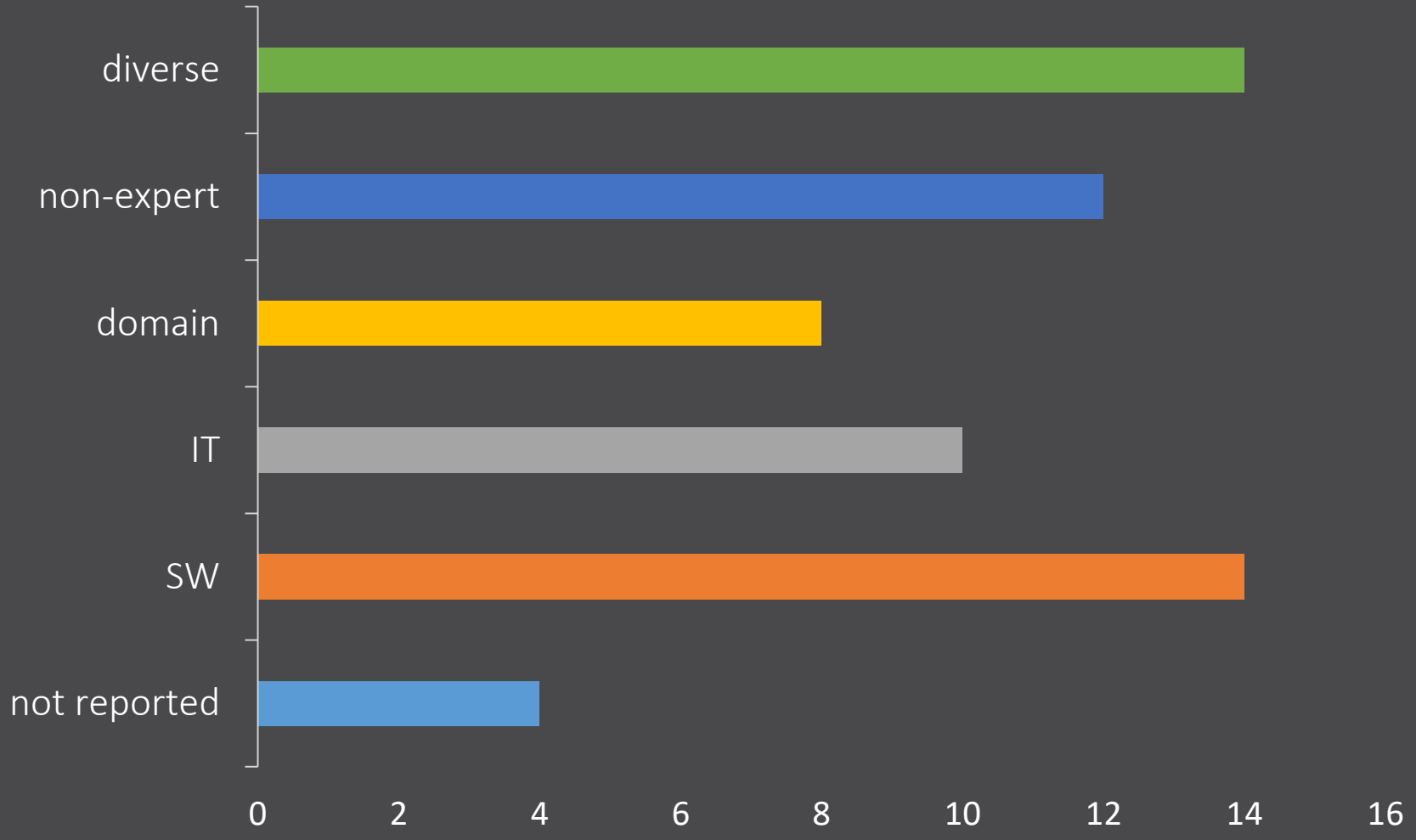

learning / understanding
exploration 17
querying 10
search 8
explanation 2
question answering 1
navigation 1


creating/managing
modelling 10
editing 9
annotation 5
mapping 2
validation 1

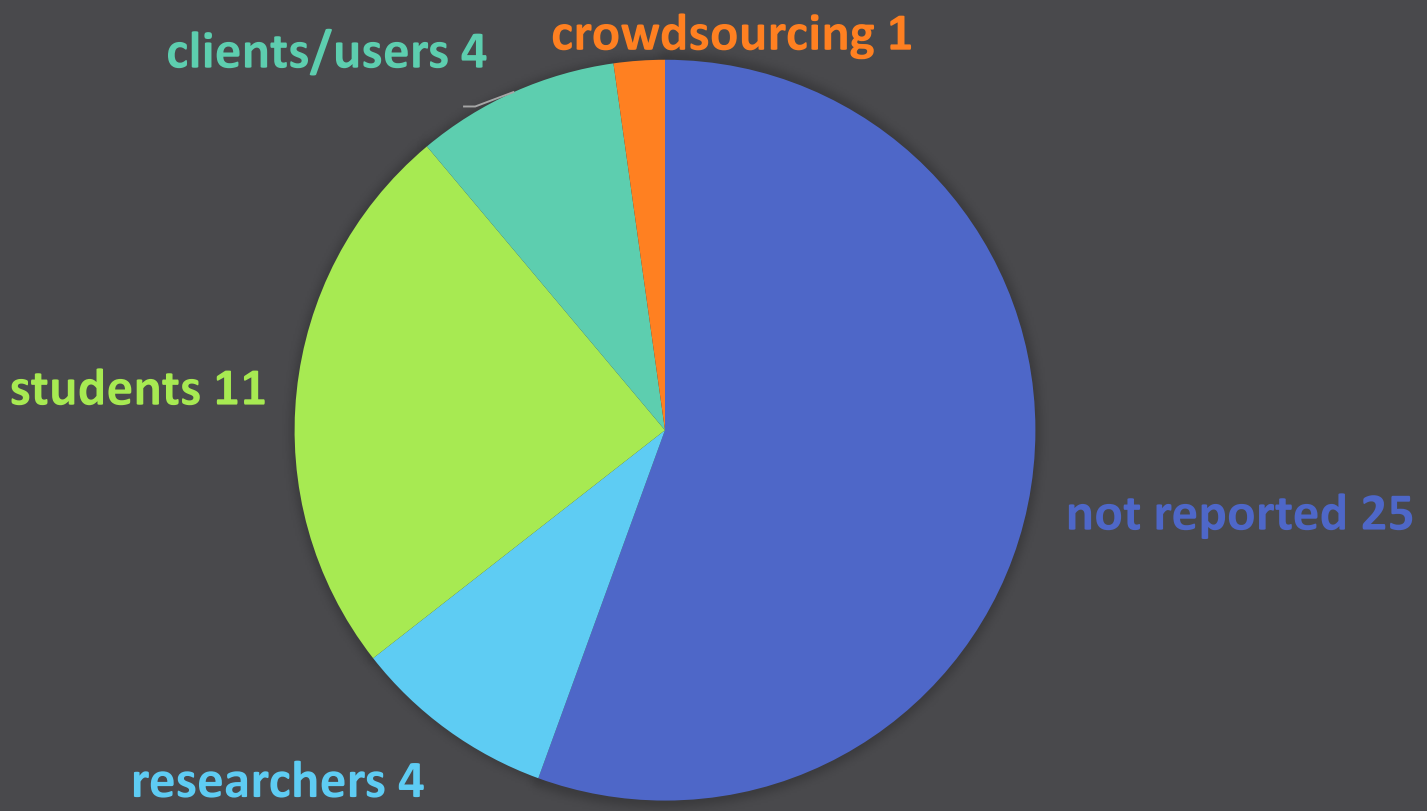
Number of users in reported study



Users expertise



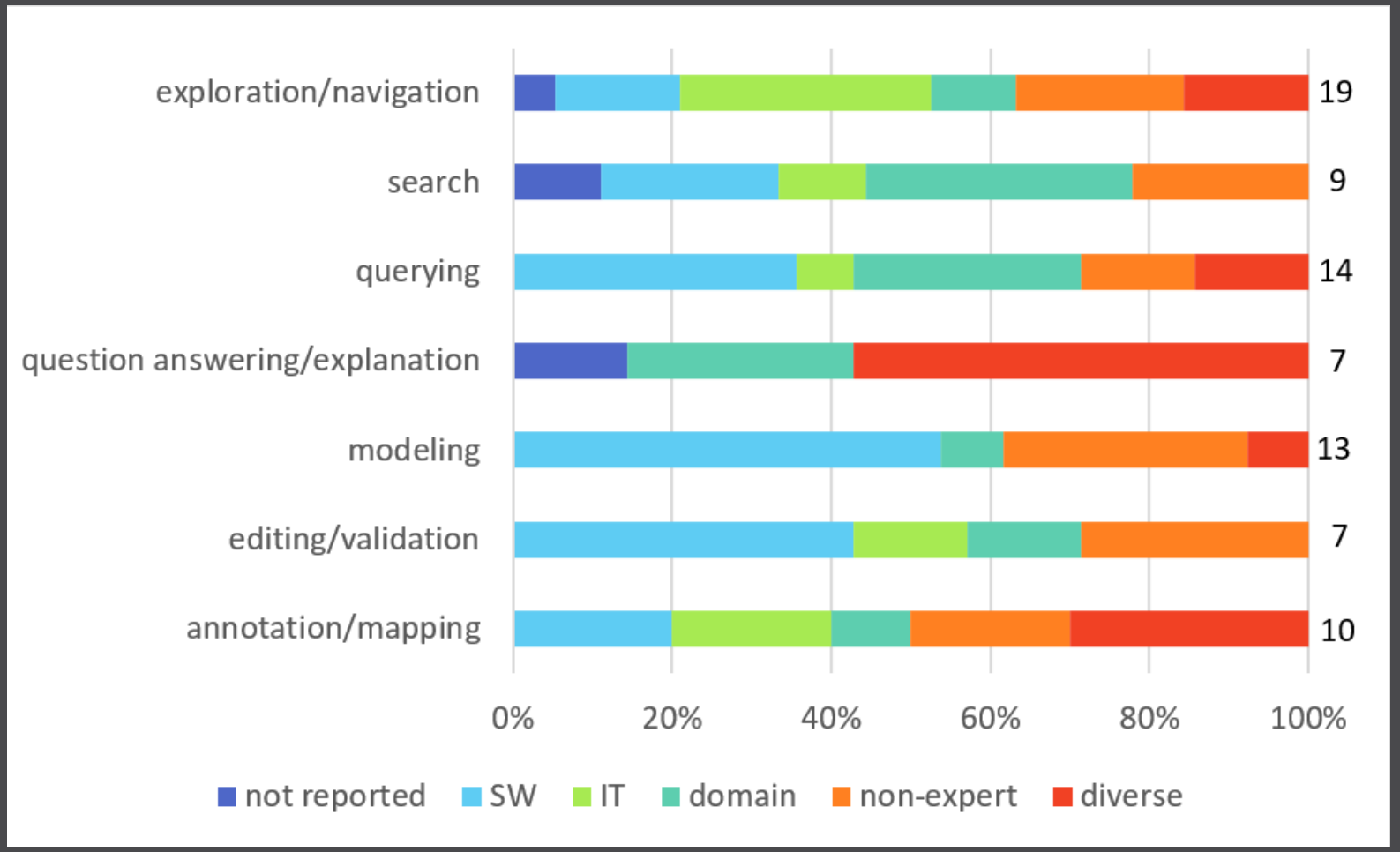
Users recruitment method



Evaluation Methods

- Qualitative 12
- Quantitative
  - Questionnaire
    - Standard 8
    - **Custom 21**
  - Task performance
    - **Success 10**
    - Time 2
    - **Success & Time 10**
    - Non-tracked 3
  - Design
    - Comparative within subjects 5
    - Comparative between subjects 3

User expertise by type of purpose and operation

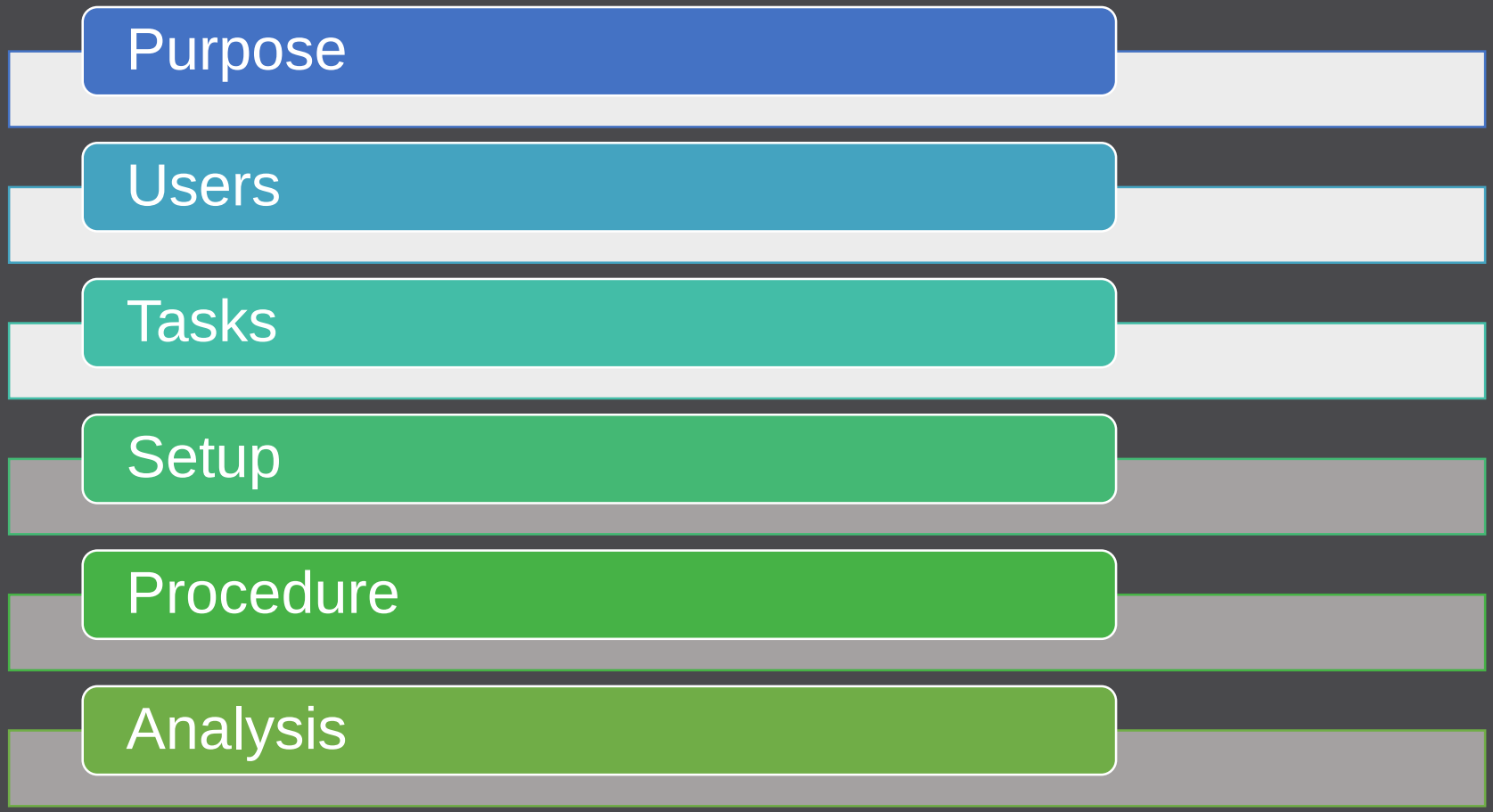


How to design SW user studies?

How to report on SW user studies?

- support the interpretation of the conducted user study
- enable the comparison to similar evaluations
- permit the replication of the user study

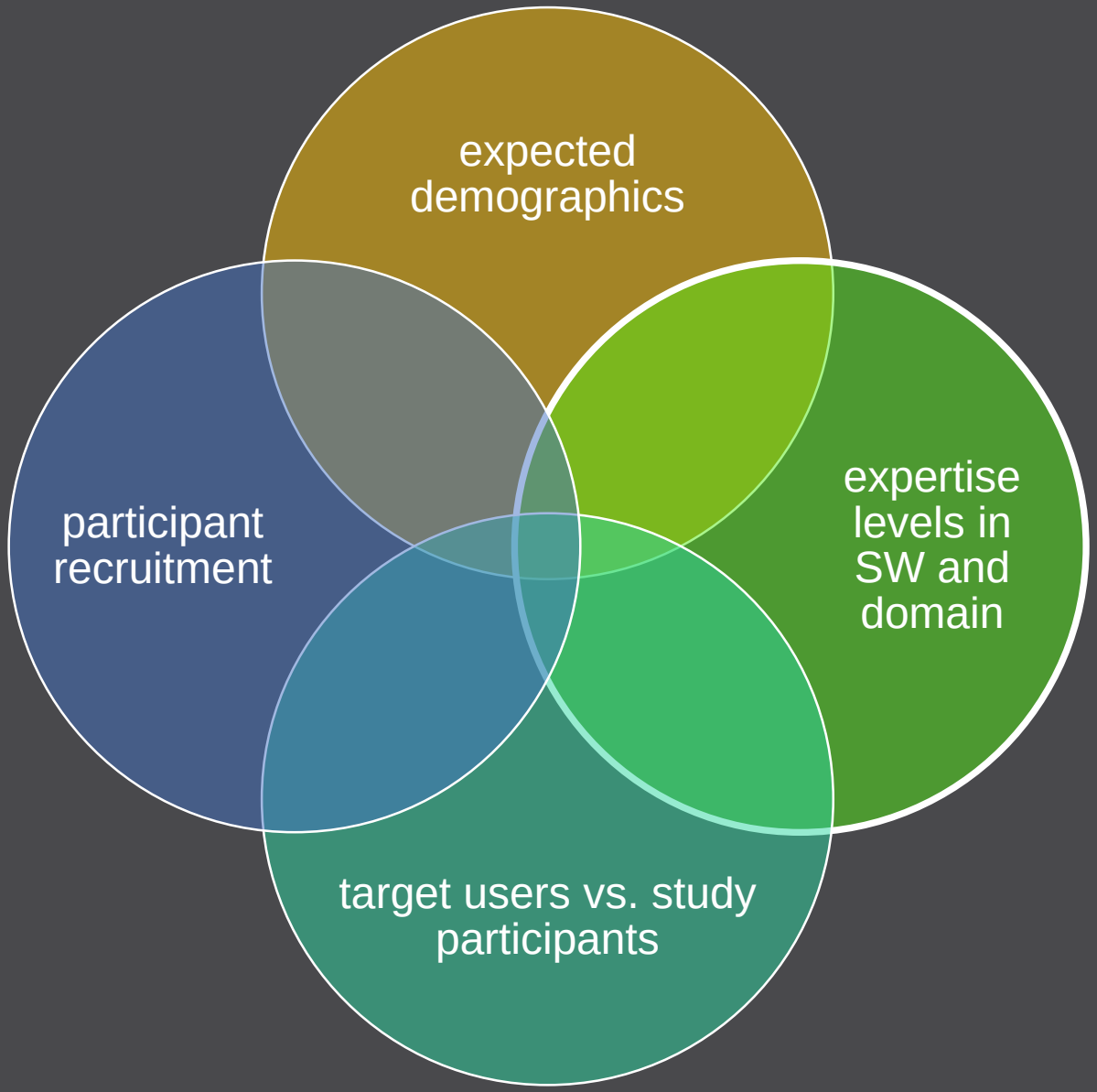
Minimum information about a user study in SW



Purpose

exploration	goal is general discovery and insight generation
search	focused examination of SW content with a clear information need
creation	includes operations such as modeling ontologies or RDF content, and creating mappings between SW resources
management	includes assessment, validation, annotation and editing of SW resources;

Users


expected demographics
participant recruitment
expertise levels in SW and domain
target users vs. study participants

Tasks

Include the **exact task descriptions**

For multi-purpose systems, **categorize tasks according to purpose**

Describe **data** and make it **available**

Final thoughts

- Nearly half of the papers did **not present a user study**
- Of the 46 papers with user study reports we reviewed, only **four** conducted studies for **SW experts and domain** experts, and **three** for **SW experts vs. novices**

- Are we struggling to find the right participants?
- Are we lacking space in our publications to devote to user testing results? Are these considered a *lesser* contribution?
- *Are SW researchers publishing their user studies elsewhere?*
- Or is there something else holding user testing and reporting in SW back?

Acknowledgements

Valentina Ivanova, RISE Research Institutes of Sweden

Patrick Lambrix, Linköping University

Steffen Lohmann, Fraunhofer IAIS

VOILA community

CP is funded by FCT (UID/CEC/00408/2013, PTDC/EEI-ESS/4633/2014).
P is funded by the Swedish e-Science Society (SeRC).
SL is partly funded by the Fraunhofer Cluster of Excellence Cognitive Internet Technologies (CIT).