

Linked Data Cubes: Research results so far

A. Karamanou, E. Kalampokis, E. Tambouris, K. Tarabanis



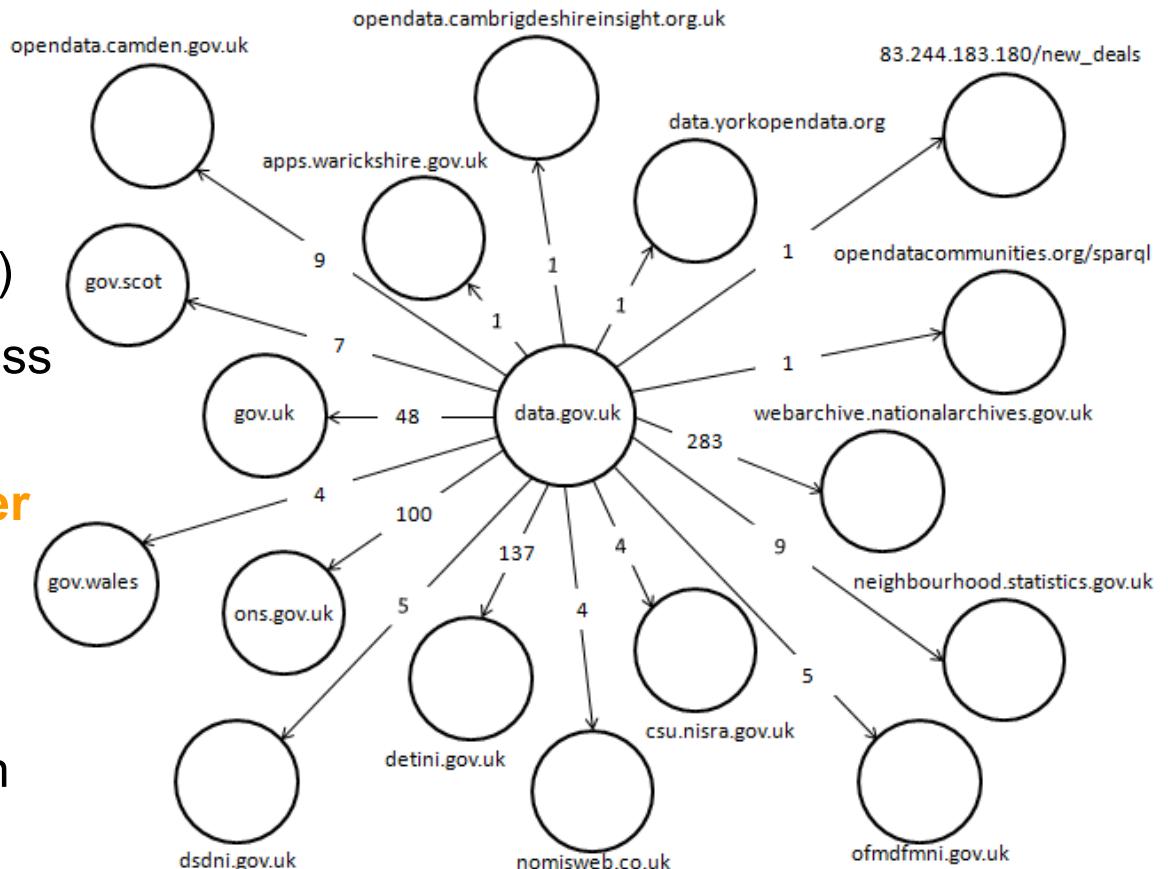
Open Statistical Data

- A large part of open data is numerical and, more specifically, concerns statistics.
- Examples include demographics (e.g. census data), social data (e.g. on unemployment and poverty), economic data (e.g. number of new businesses) etc.



“Unemployment” in data.gov.uk

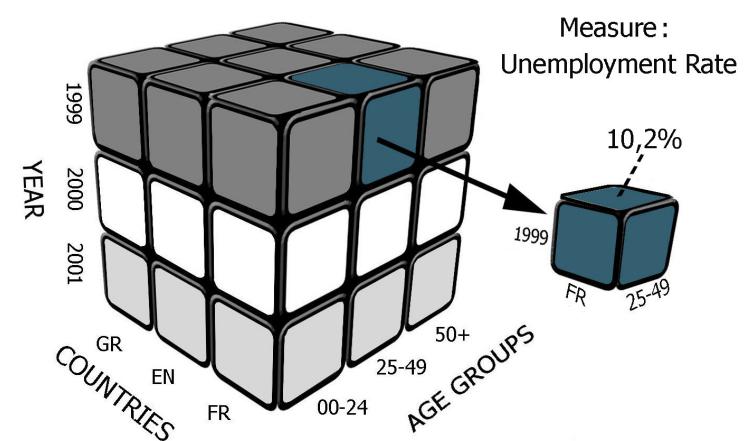
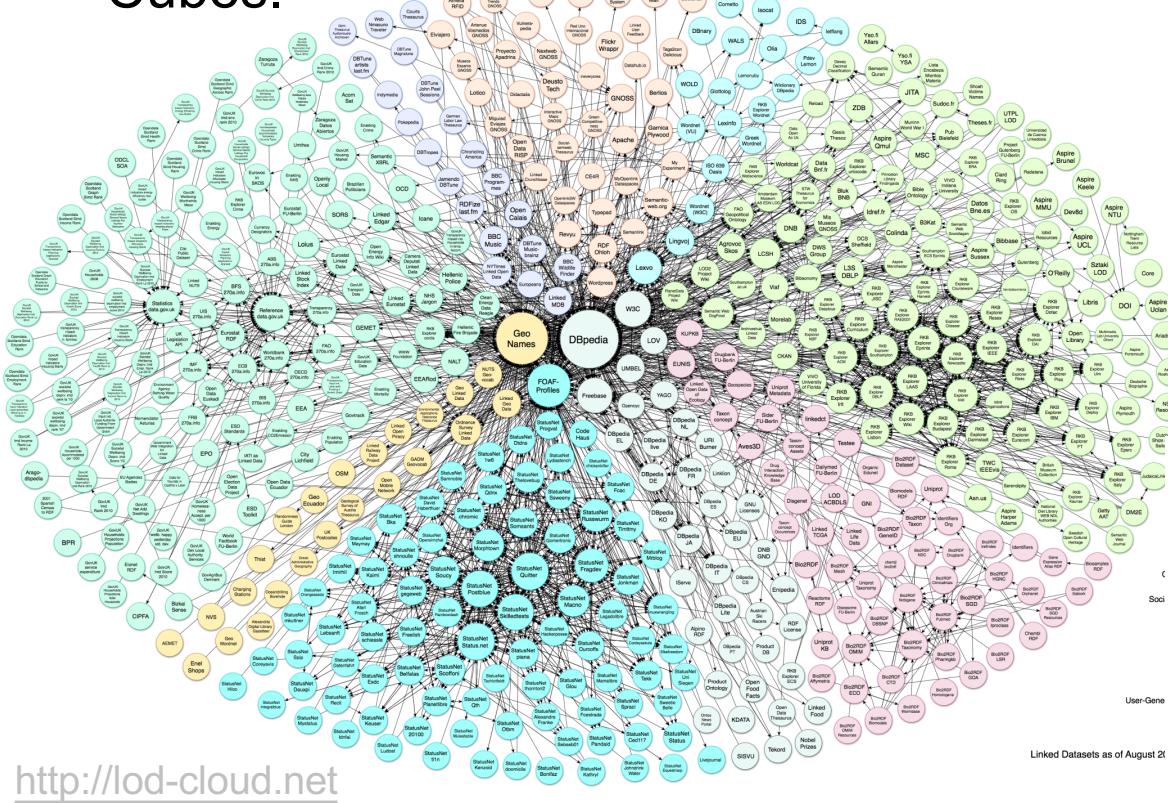
- Search on data.gov.uk for datasets using the keyword “unemployment”:
 - **122 results** (links and files)
 - These results provide access to **56 files** and **610 links**
 - These links lead to **18 other portals** (e.g. the Office for National Statistic and the National Archives)
 - Through them to more than **2000 other files**



Kalampokis, E., Tambouris, E., Tarabanis, K.: Linked open cube analytics systems: Potential and challenges.
IEEE Intelligent Systems, 31(5), 2016.

Linked Data Cubes

- Linked data promises to facilitate data integration
 - During the last years a growing body of literature studies Linked Data Cubes.



Aim of the paper

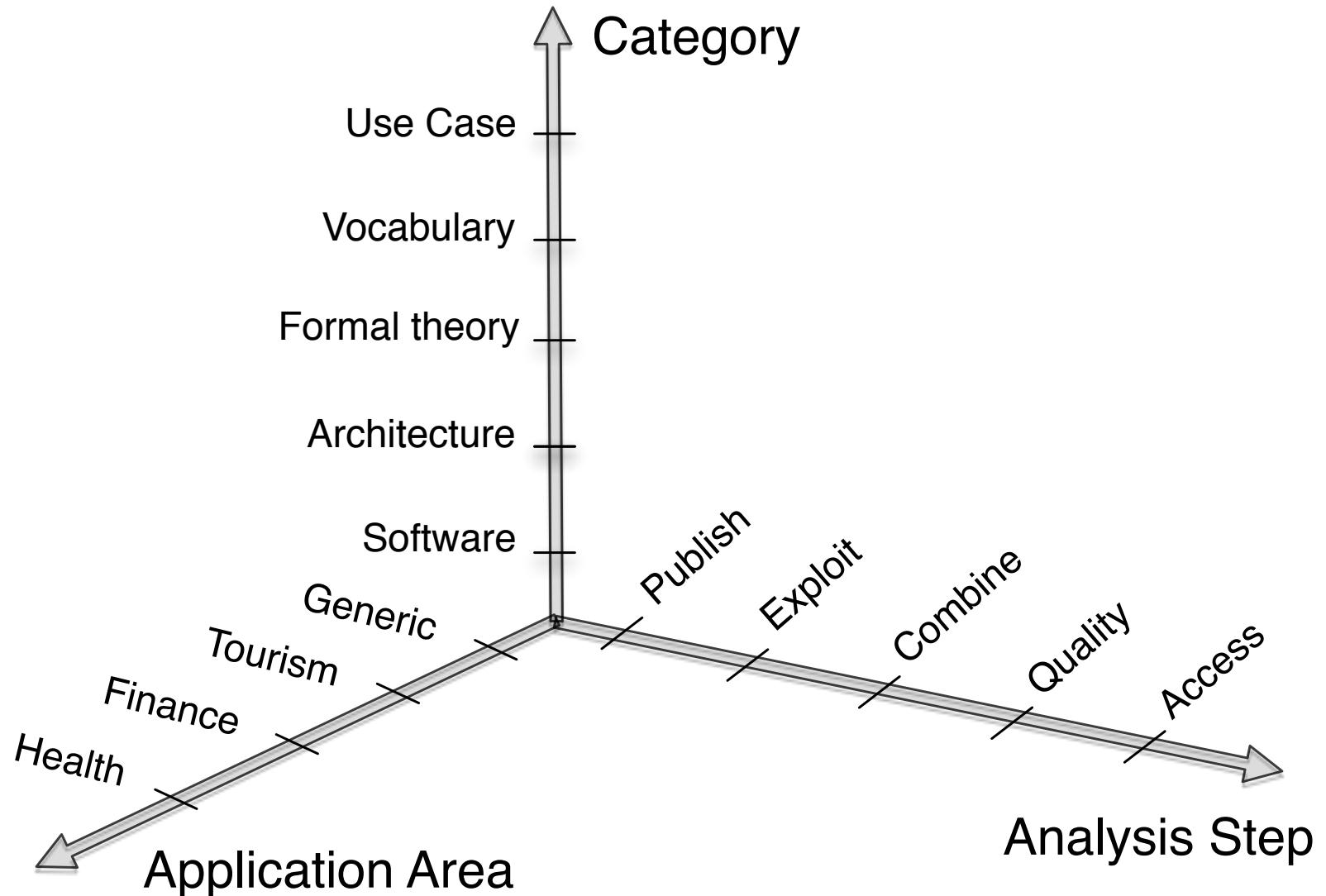
- The aim of this paper is to consolidate this body of knowledge and provide a preliminary understanding of the research results in the area so far.
- This will enable
 - understanding of research results so far
 - proposing future research directions
- This presentation includes preliminary results as the analysis of the literature is still in progress.

Approach

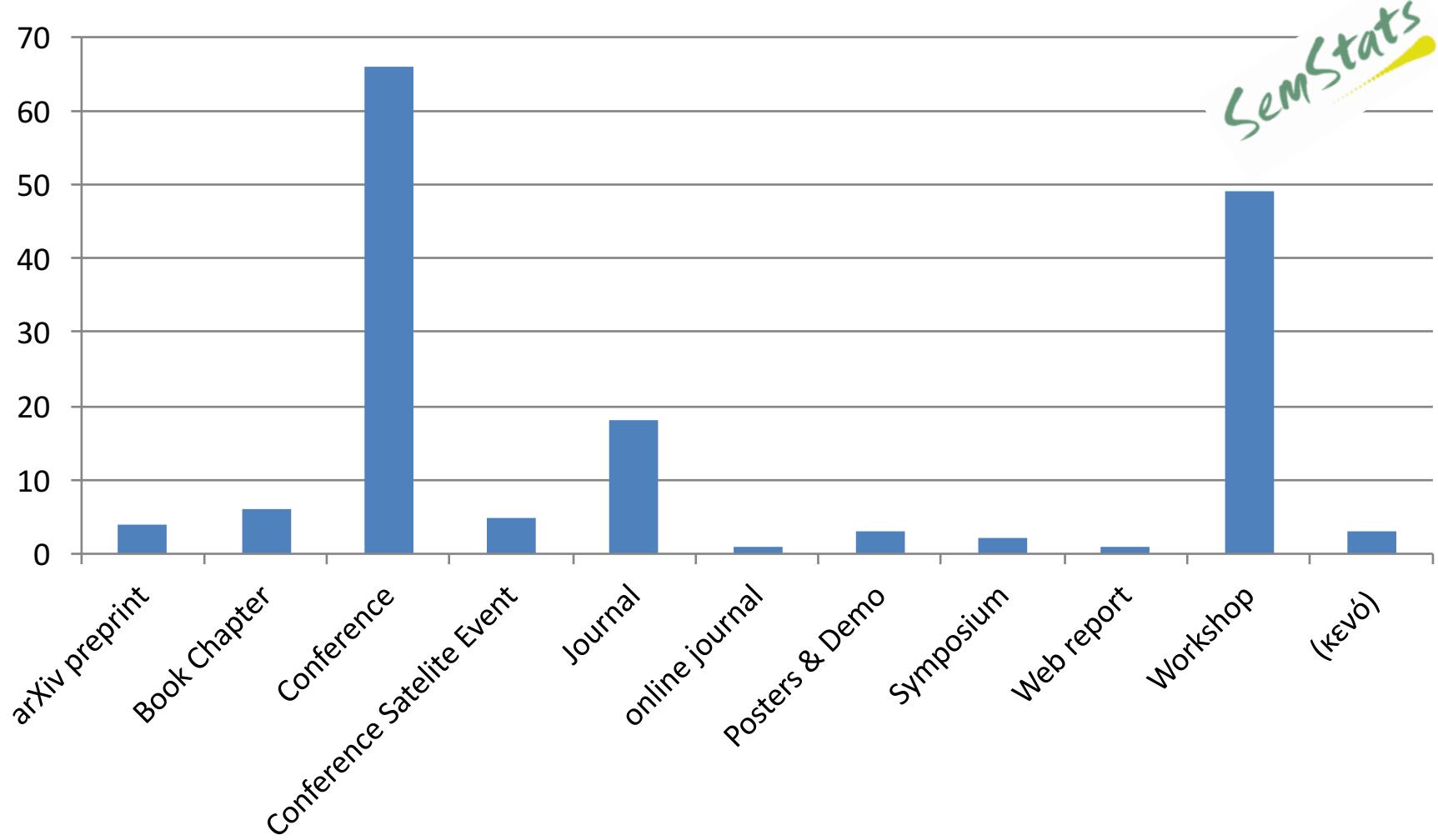
- Systematic search in order to accumulate a relatively complete body of relevant scientific literature.
 - Google Scholar using the keywords “**linked data**” AND “**data cube**”.
 - Backward (citations in the identified articles) and forward (articles citing the previously identified articles).
 - Final set that was included in our research.
- This approach resulted in a set of **138 papers**
- Synthesize the accumulated knowledge (concept-centric analysis).
 - The main characteristics of the area were extracted and **a conceptual framework for linked data cubes** was created in order to structure the area.
- The framework was employed to **categorise** and further analyse the literature and to extract insights into the area of linked data cubes

Webster, J. Watson, R. T.: Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly, pages xiii–xxiii, 2002.

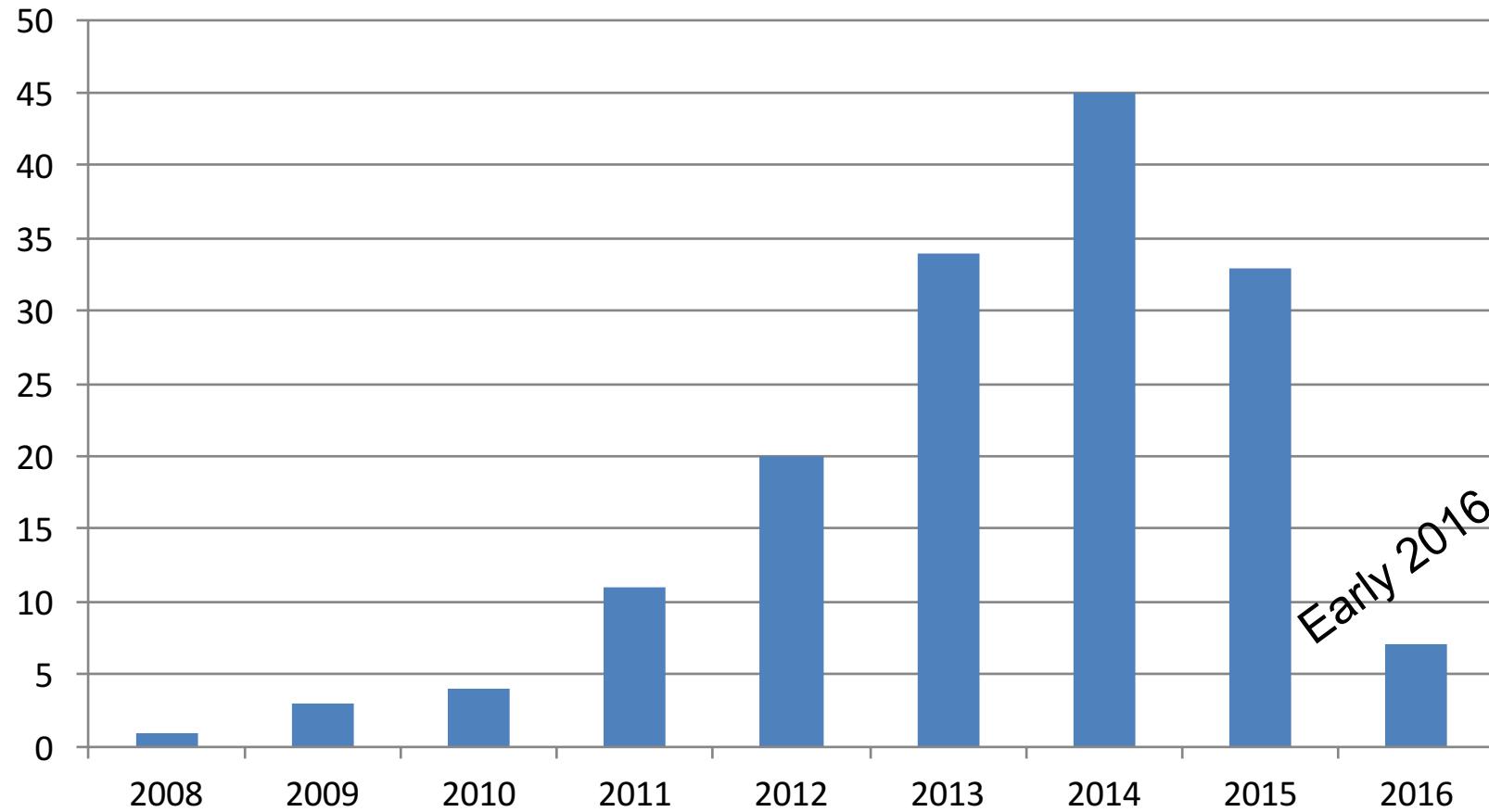
Conceptual Framework



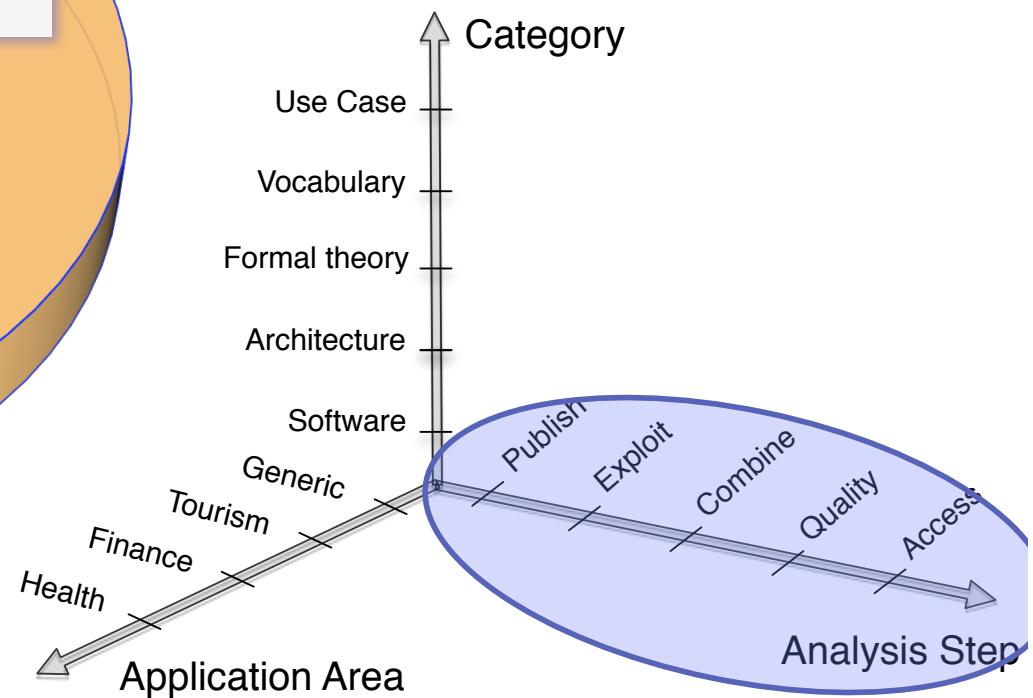
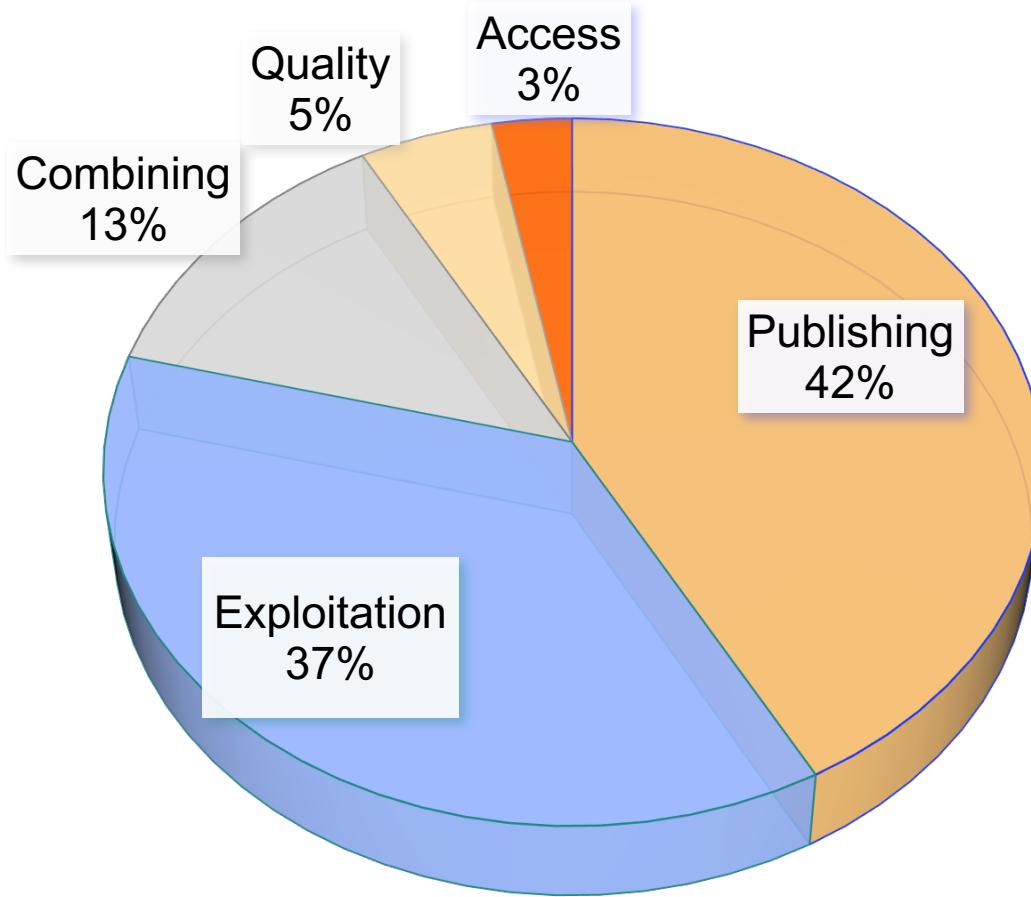
Number of contributions per source type



Number of contributions per year

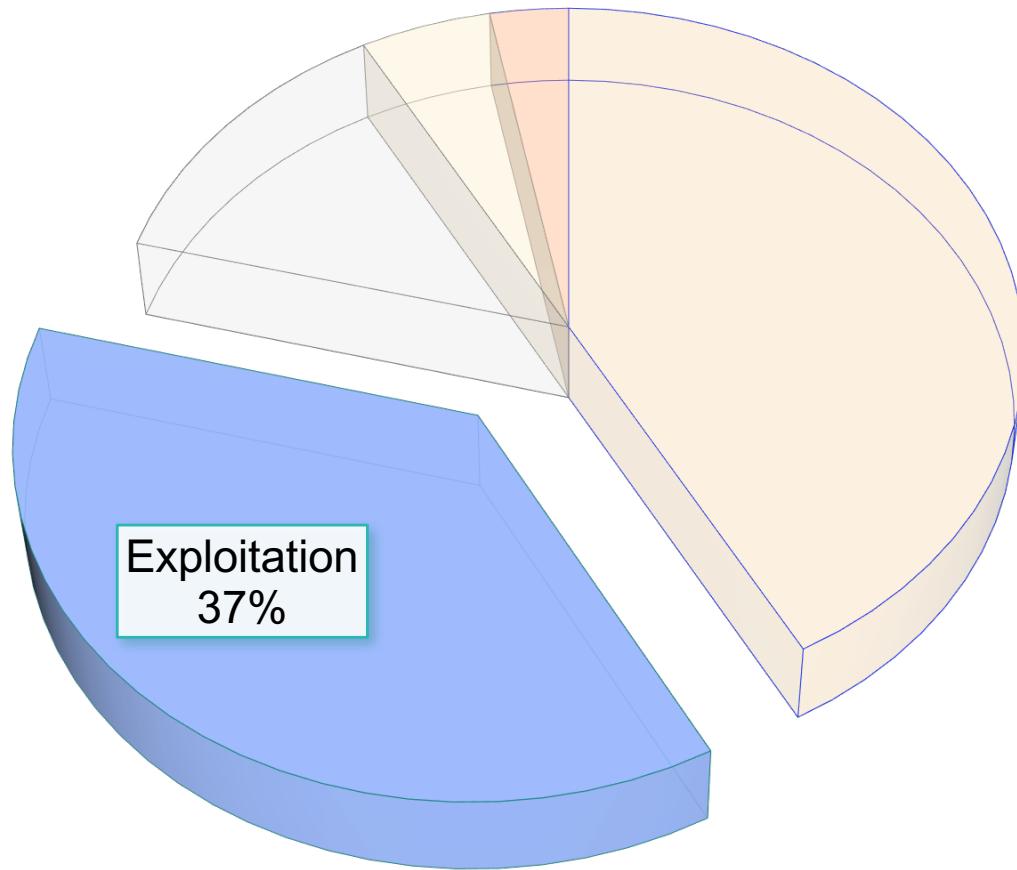


Dimension: Analysis Step



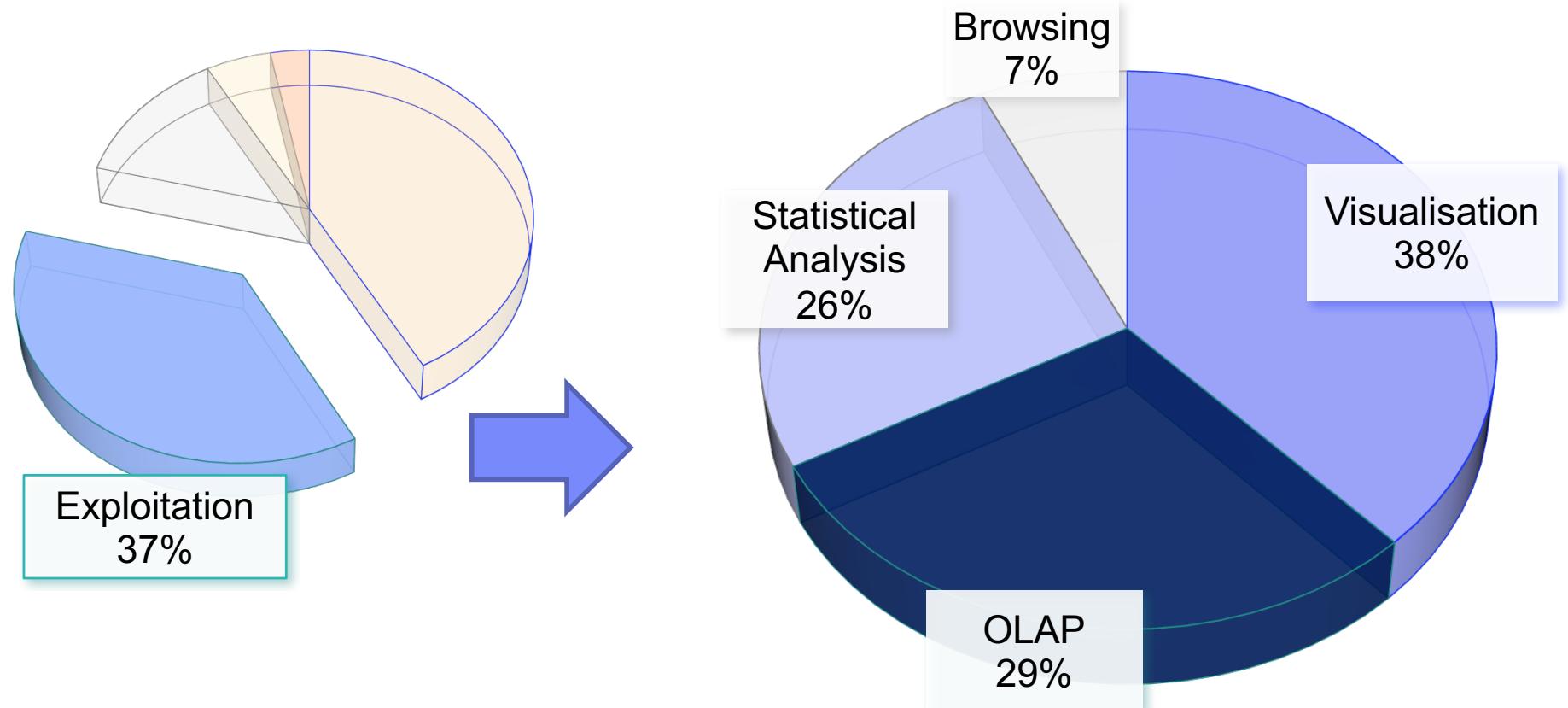
Analysis step – exploitation

- Drill-down on “Exploitation”

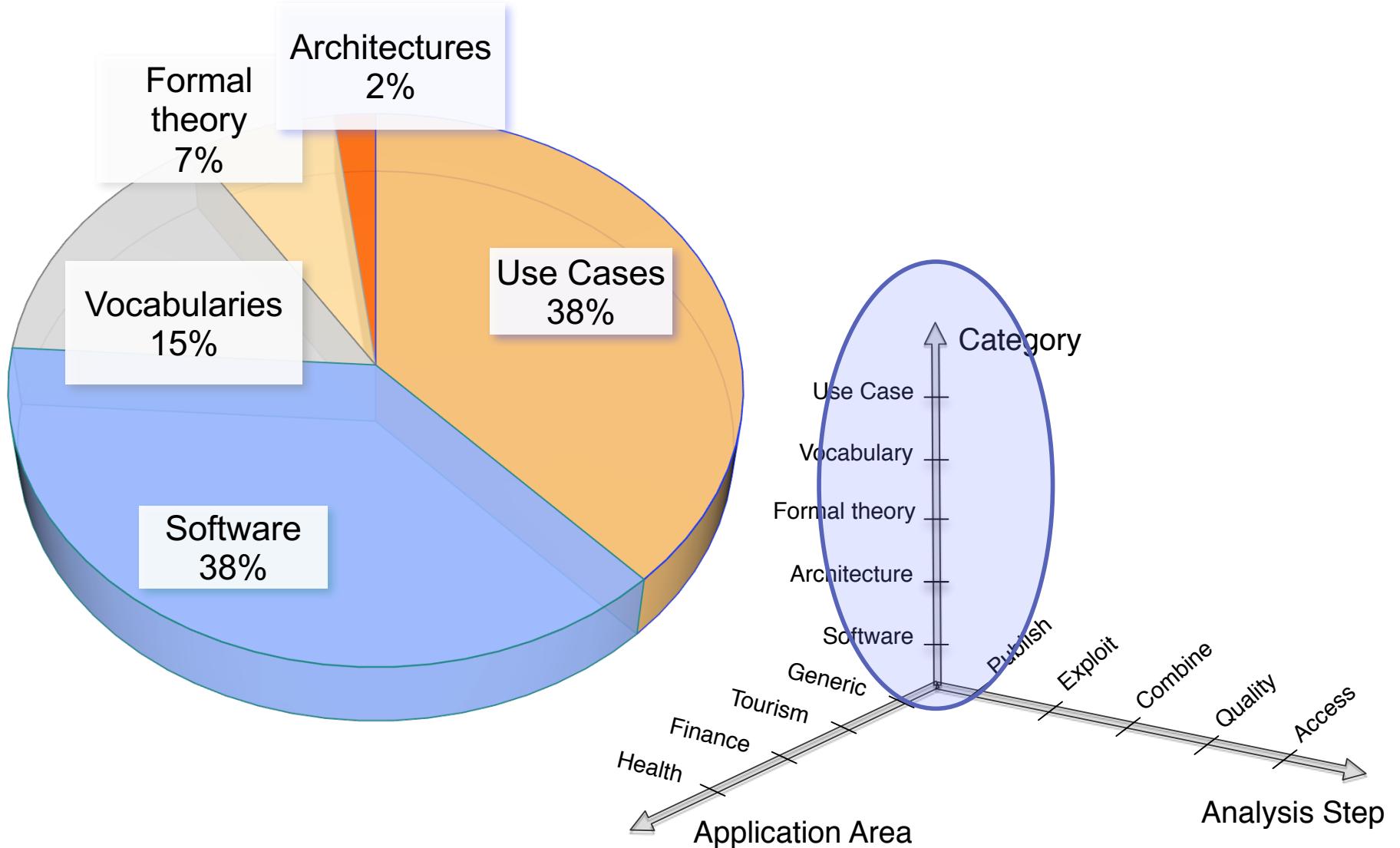


Analysis step – exploitation

- Drill-down on “Exploitation”

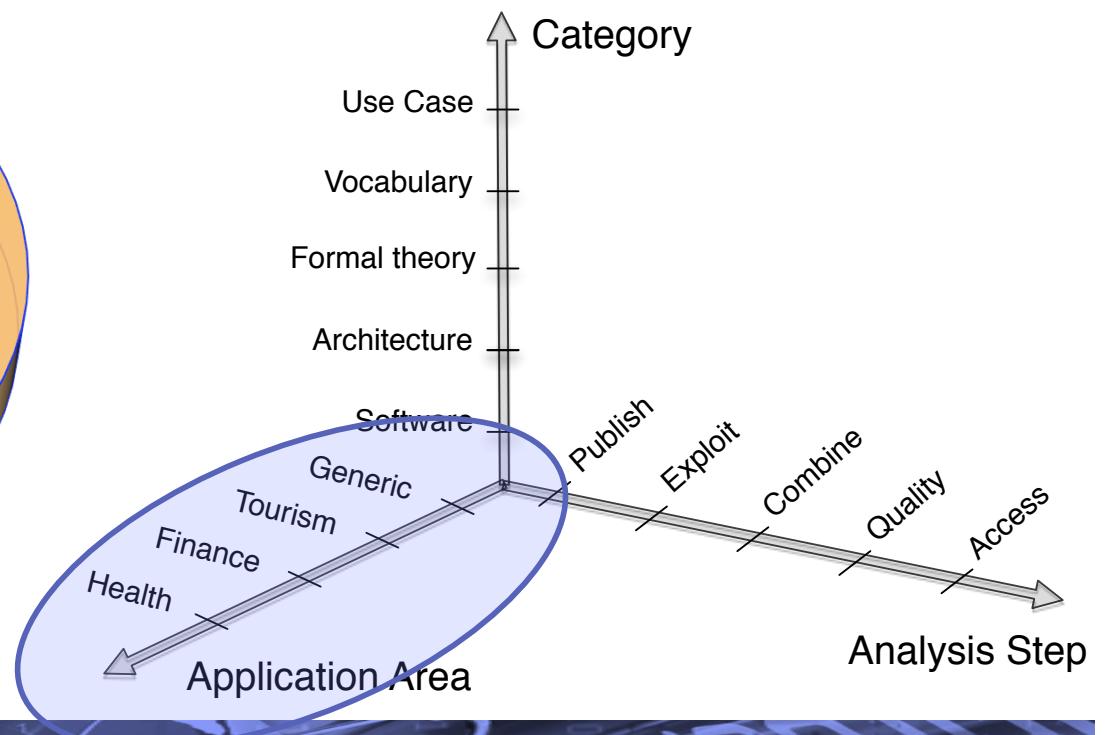
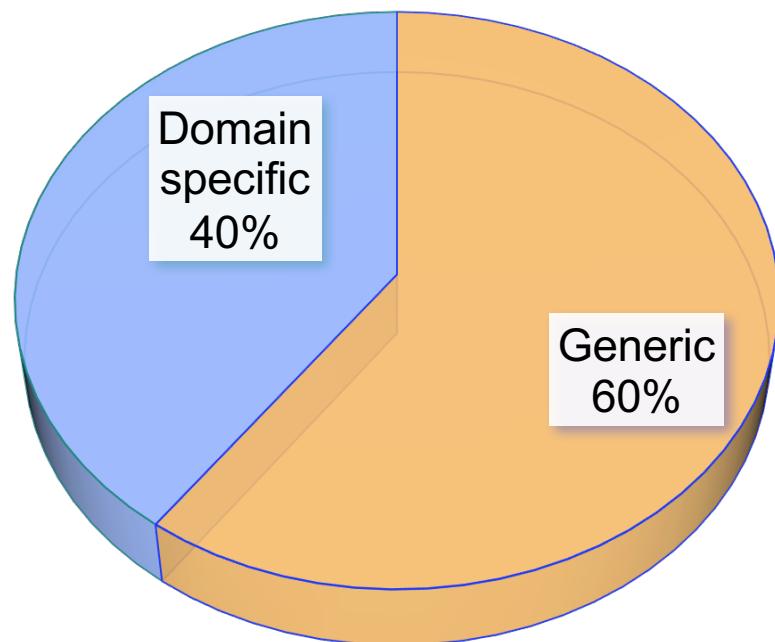


Dimension: Category

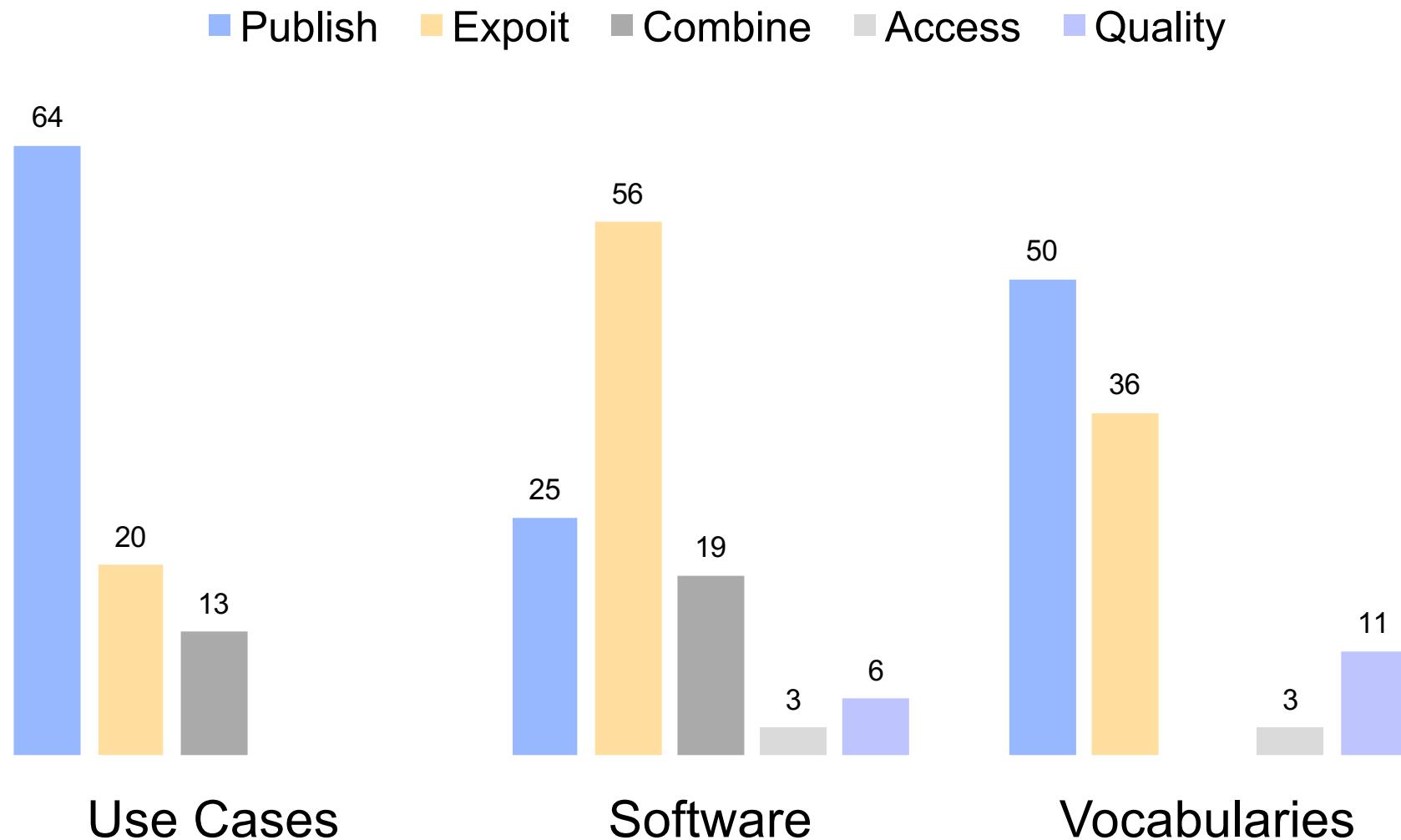


Dimension: Application Area

- Domains include health, policy-making, government, environment, economics, biology, and tourism.
- The most notable domain is **health** that characterises 25% of all domain-specific contributions



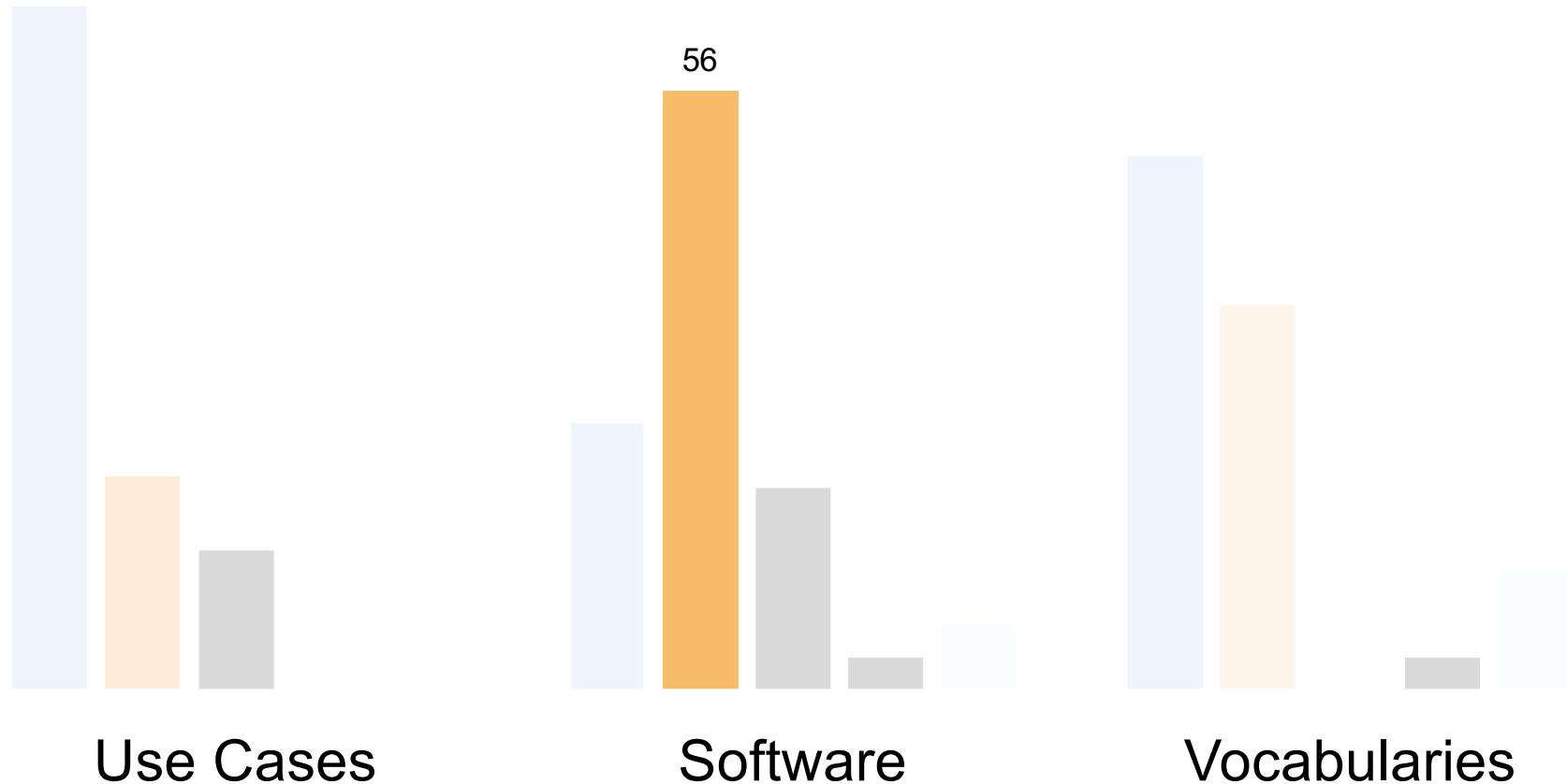
Category vs Analysis Step



Category vs Analysis Step

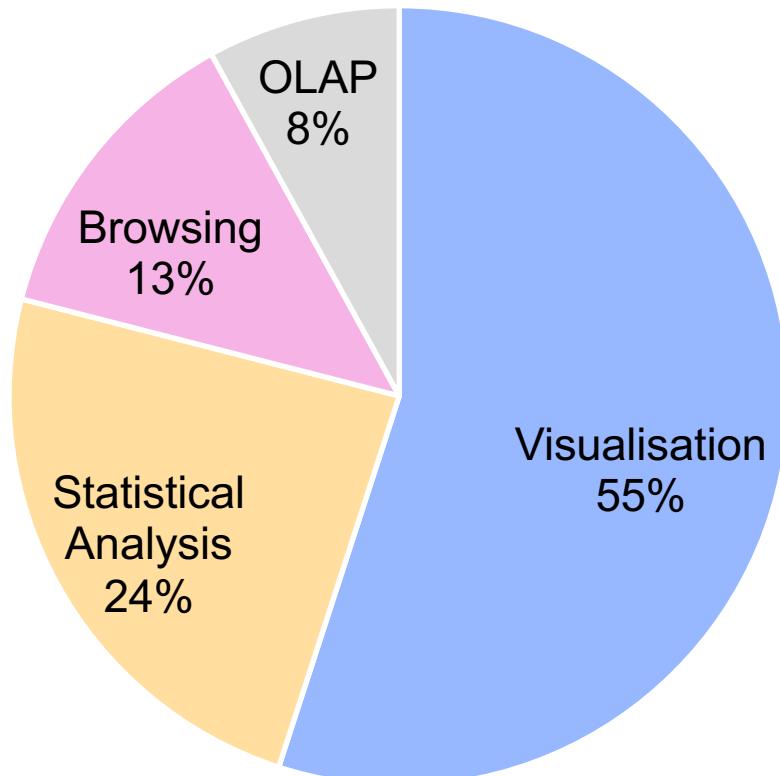
- If we drill down on exploitation software contributions

■ Publish ■ Exploit ■ Combine ■ Access ■ Quality

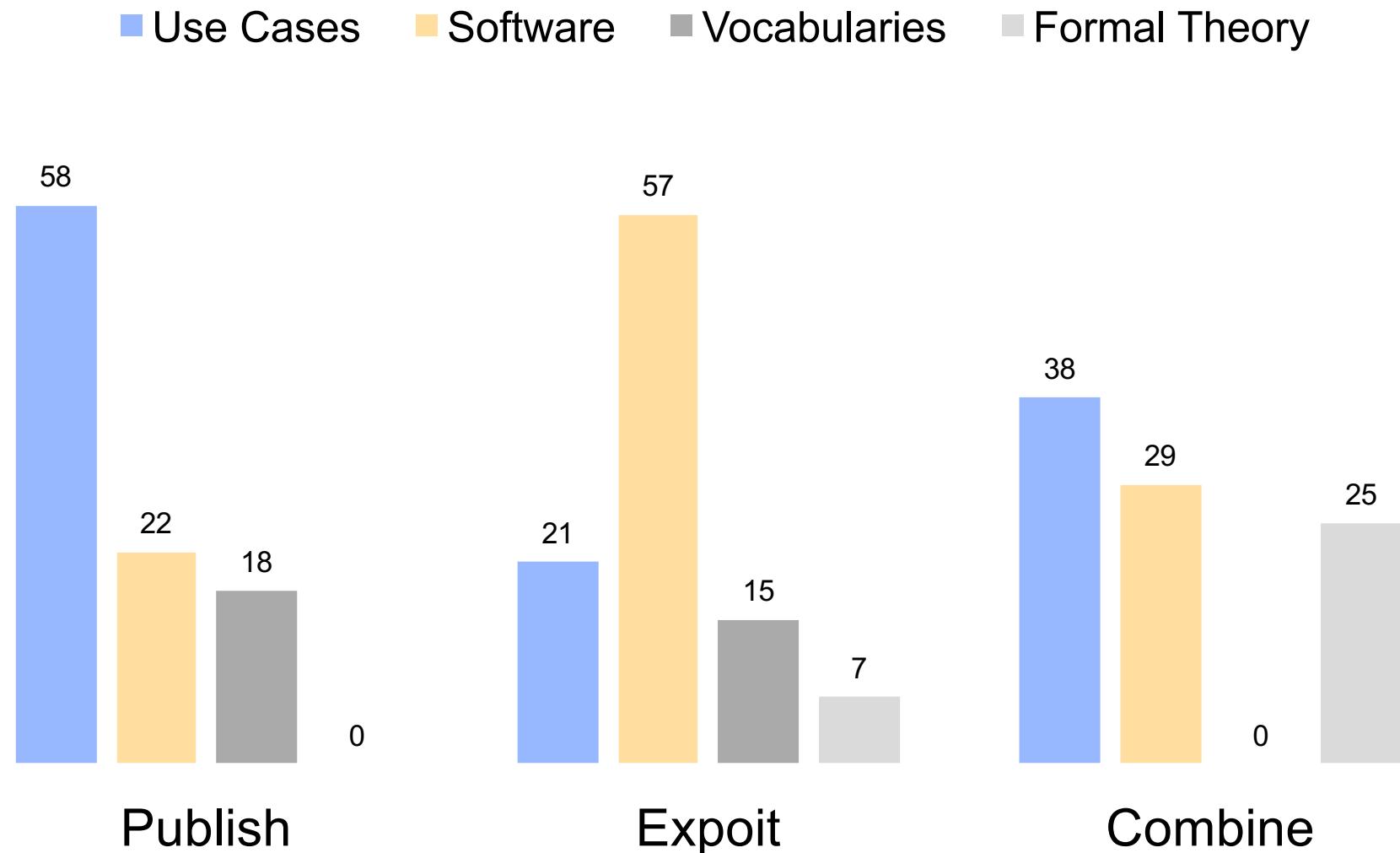


Category vs Analysis Step

- If we drill down on “exploitation software” contributions



Analysis Step vs Category



Application Area vs Category

- 77% of all software contributions are domain independent
 - This is 82% and 79% in the case of software for publishing and exploitation respectively
 - This goes down to 57% in software that support combining of linked data cubes
- 78% of all vocabulary contributions are domain independent
- 31% of the use cases are generic

Conclusion & Future Work

- The majority of the contributions that focus on publishing linked data cubes are cases while the majority of the exploitation contributions are software tools
- Although linked data promises to facilitate data integration, integration of data cubes remains largely unexamined in the literature (13% of total contributions).
- Future work
 - Update the conceptual framework
 - Enable an in more depth understanding
 - Come up with more suggestions for future research

Acknowledgments

- The work presented in the paper is partly funded by

opengovintelligence

Improving public services with multidimensional statistical data



<http://OpenGovIntelligence.eu>



@OpenGovInt