UNIVERSITY OF MANNHEIM

On the Status of Experimental Research on the Semantic Web

Heiner Stuckenschmidt, Michael Schuhmacher, Johannes Knopp, Christian Meilicke, Ansgar Scherp

Data and Web Science Research Group

Abstract

- Descriptive study about the role of experiments
- ISWC research papers from 2003 – '12 analyzed

- We study
  - The types of papers,
  - the importance of experimental work,
  - the quality of experimental work, and
  - if any of this has an influence on a paper's impact

Studies on the Role of Experiments in Computer Science

1. Computer Science (CS) in General
   - Far less experiments than in natural science or engineering [Tichy et al 1995]
   - Similar picture for ACM Publications [Wainer et al 2009]
2. Software Engineering
   - Extremely few experiments [Zelkowitz 1998]
   - Situation improved since then [Zelkowitz 2009]
3. Other special areas of CS
   - Neural Networks [Prechelt 1996]
   - Computer Supported Collaborative Work [Pinelle and Gutwin 2000]

Study Design

1. What is Semantic Web Research (SWR)?
   → ISWC research track papers as proxy

2. SWR is young, how is it developing?
   → All 500 papers from 2002 – 2012

3. How to ensure good annotations?
   → 2 annotators per paper
   → 1 junior and 1 senior research
   → Disagreements solved by 3rd senior researcher

Paper Classification Schema

Reusing classification schema from [Tichy et al 1995] and [Wainer et al 2009]:

1. Formal Theory
2. Design and Modeling:
   - main contributions are systems, techniques, algorithms, or models
   - claimed properties cannot formally be proven
3. Empirical Work / Hypothesis Testing
4. Other (e.g. surveys)

On this data we tested 5 hypothesis

H1) Like in CS in general, Design and Modeling work is the dominant form of research on the Semantic Web

H2) The importance of experimental work on the Semantic Web is comparable with computer science in general.

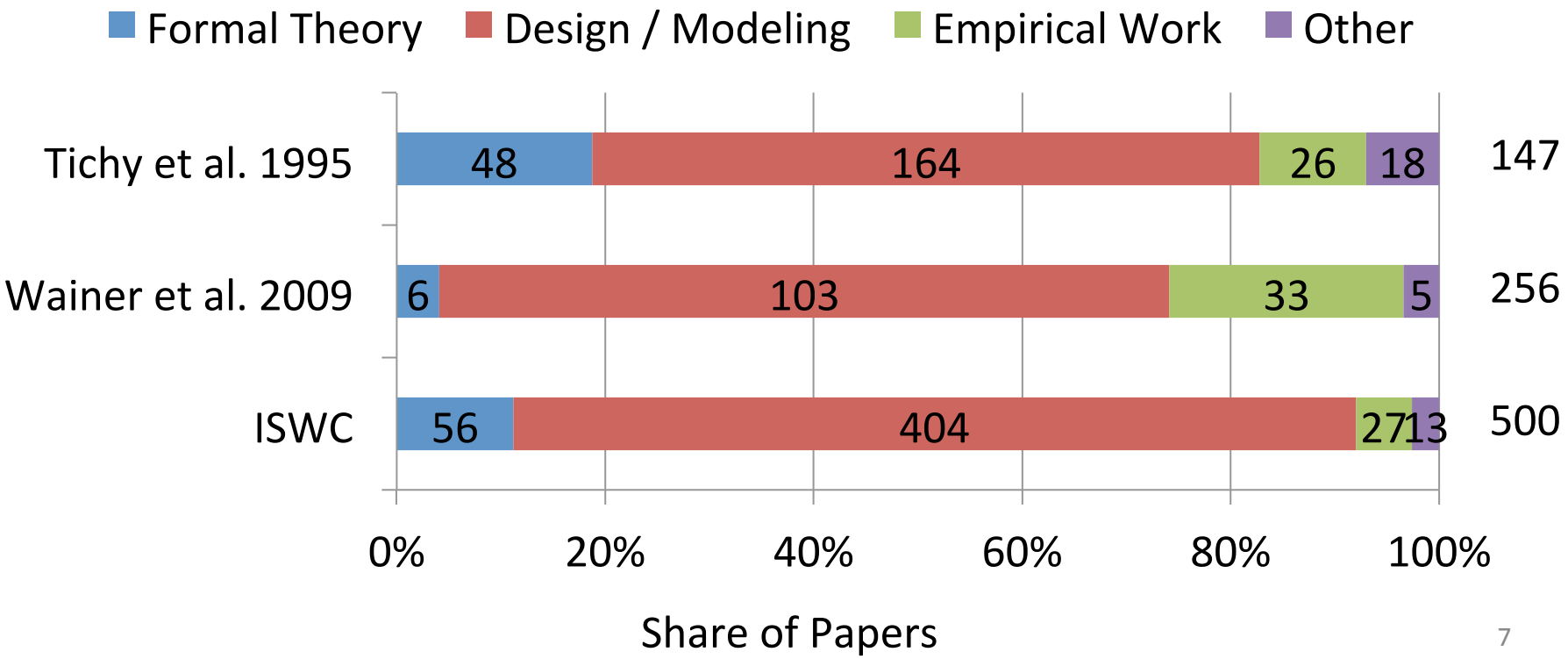
H3) The importance of experimental work on the Semantic Web is increasing over time.

H4) The quality of experimental work on the Semantic Web is increasing over time.

H5) Strong experimental work increases the impact of a paper.

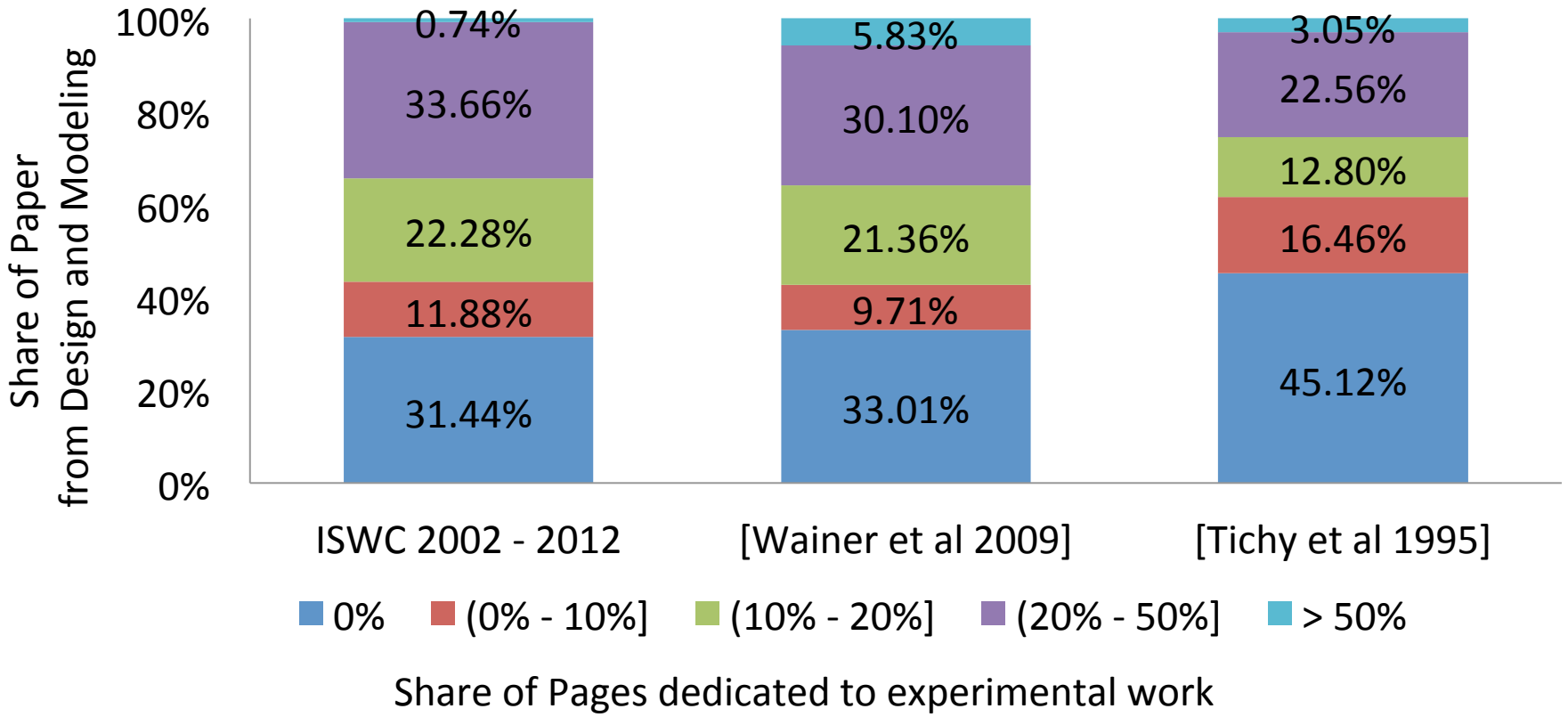
H1: Design and Modeling

Like in general CS, Design and Modeling work is the dominant form of research on the Semantic Web:



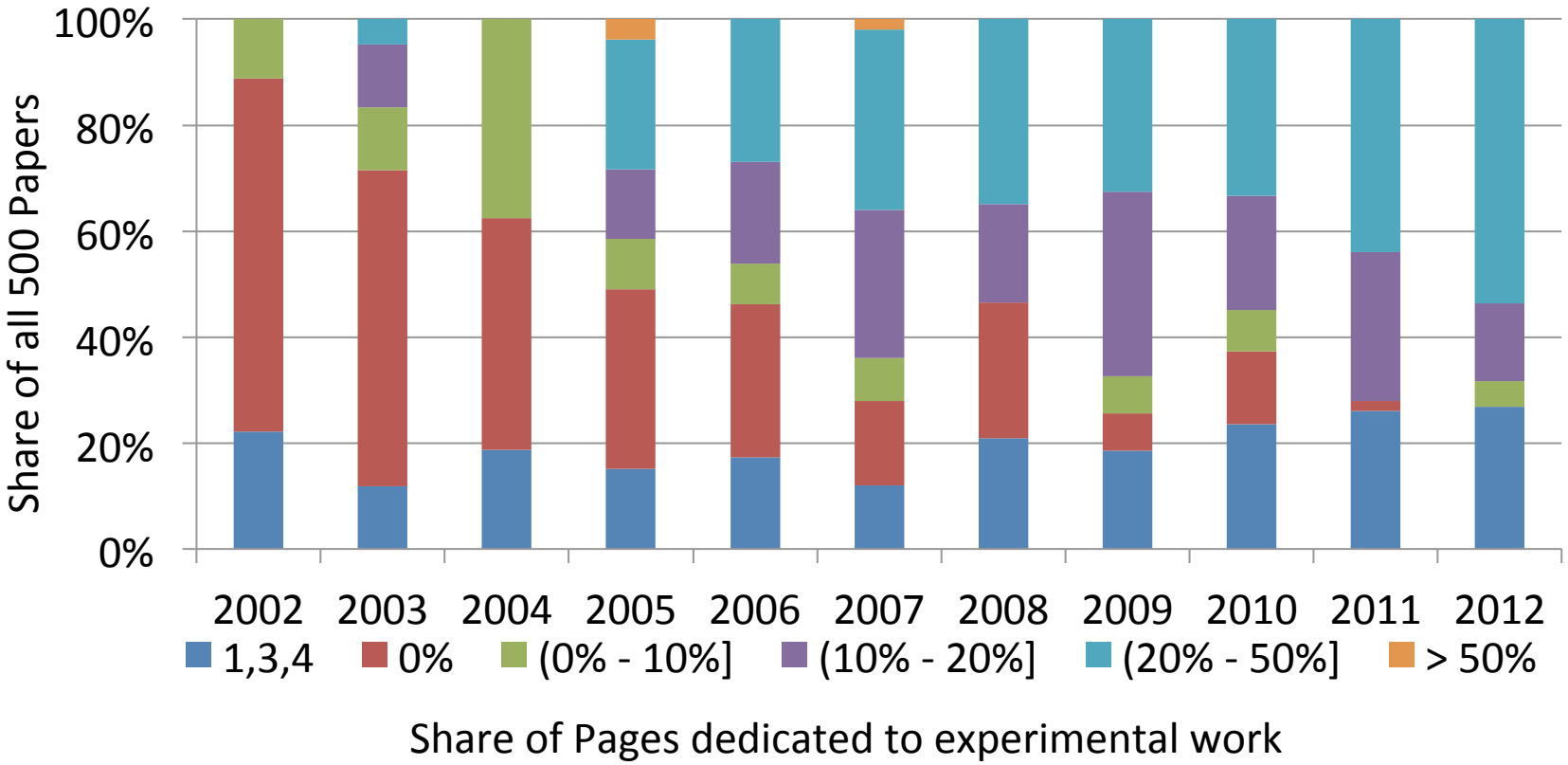
H2: Importance of experimental work

The importance of experimental work on the SW is comparable with CS in general



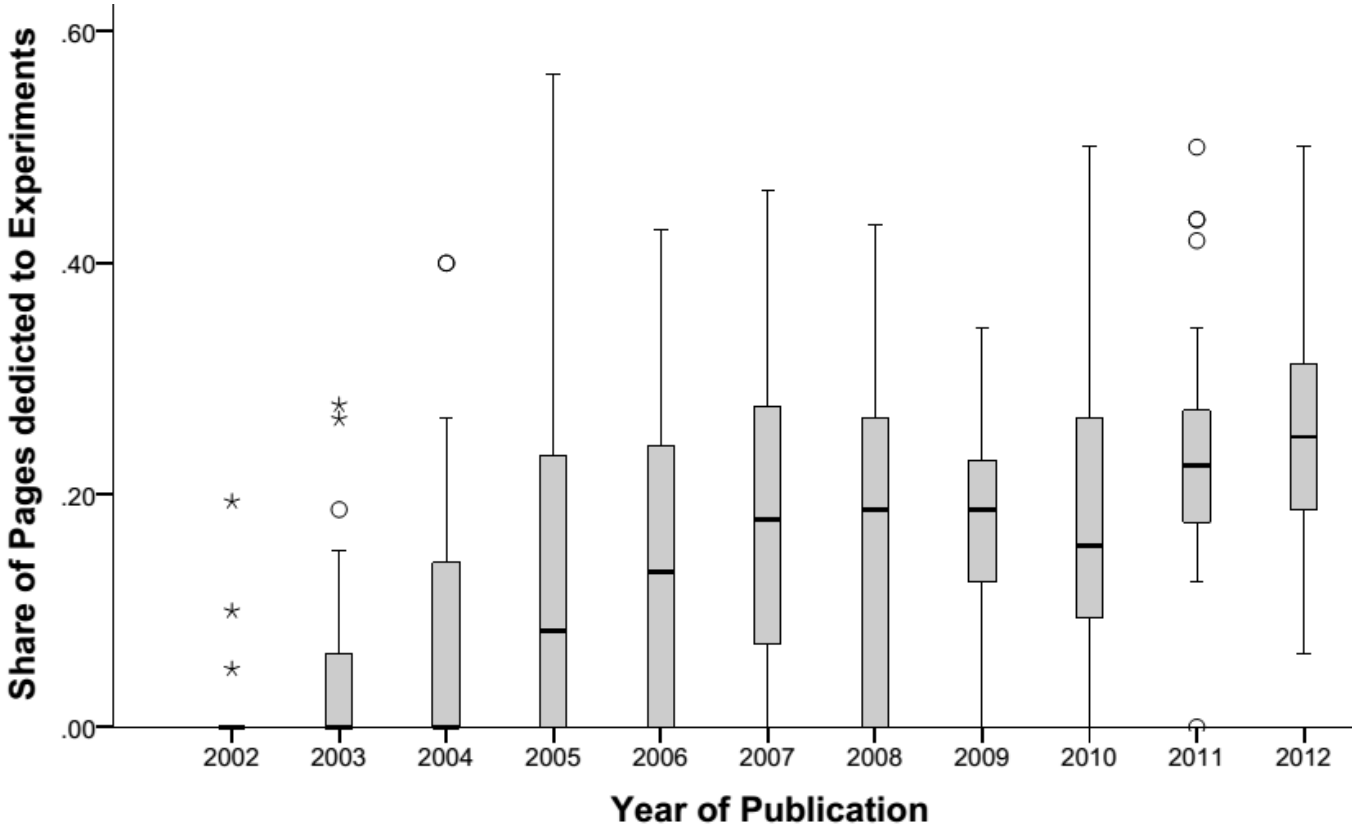
H3: Experimental work over time

The importance of experimental work on the semantic web is increasing over time



H3: Experimental work over time

The importance of experimental work on the semantic web is increasing over time



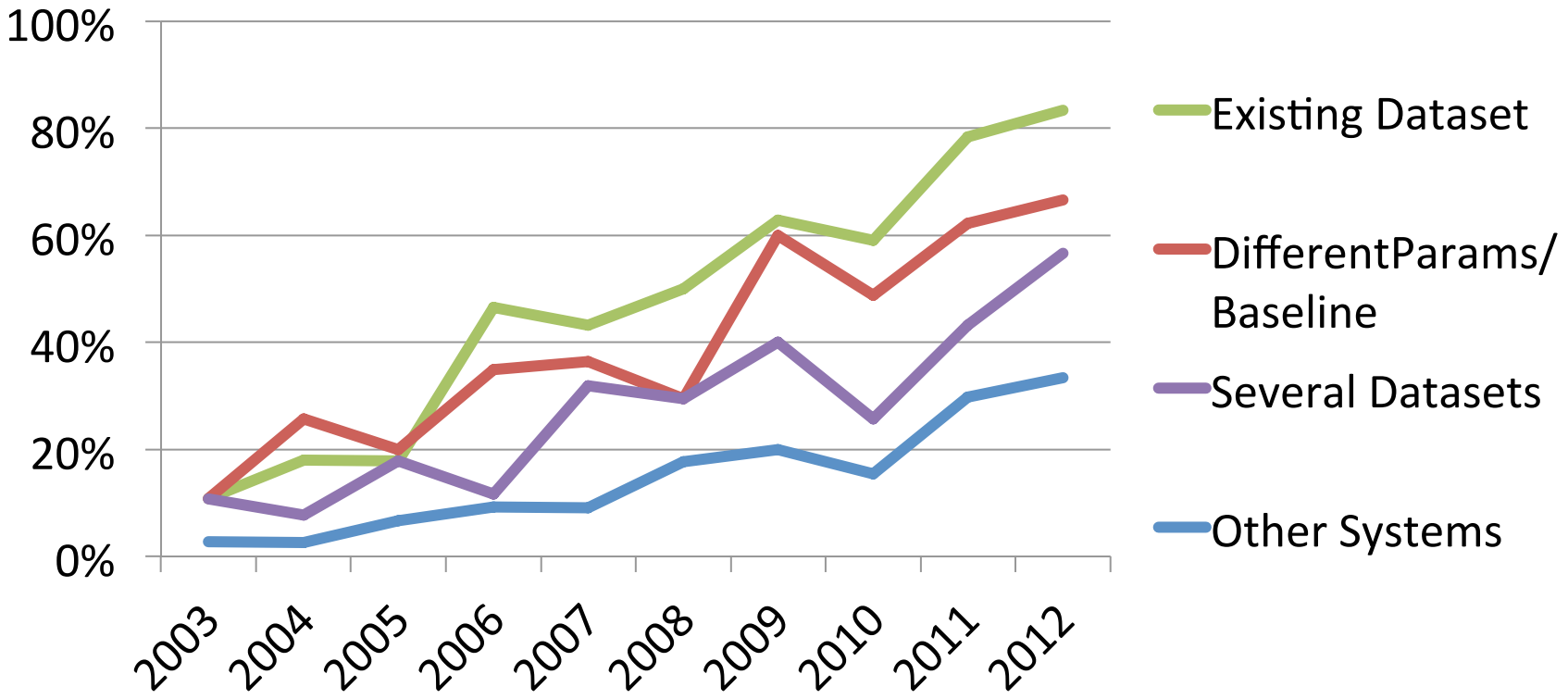
H4: Quality of experimental work

The quality of experimental work is increasing over time.

- **Existing Dataset**: Using an already existing dataset, not one that has been created for the purpose of conducting the experiments
- **DifferentParameters/ Baseline**: Comparing the proposed approach against a baseline or comparing different settings against each other
- **Several Datasets**: Using several datasets, not only one
- **Other Systems**: Comparing existing algorithms/systems

H4: Quality of experimental work

The quality of experimental work is increasing over time.



H5: Influence on Impact of paper

Strong experimental work increases the impact of a paper.

- **Measure**: Average Citation Count per year (based on Google Scholar)
- **Variables same as for H4 plus**
- **Age**: Age of the paper
- **RelPagesClass**: Share of pages describing experiments grouped into 5 classes
  0%, (0%-10%], (10%-20%], (20%-50%], (50%-100%]

H5: Influence on Impact of paper

Strong experimental work increases the impact of a paper.

?

	Wald Chi-Square	Deg. of Freedom	Significance
(Intercept)	281.760	1	.000
RELPAGESCLASS	5.510	4	.239
AGE	164.634	10	.000
BASEDIFF	3.139	1	.076
SYS	3.835	1	.050
OTHER	0.205	1	.651
SEVERAL	3.750	1	.053

Log-linear regression to explain citation count

Conclusion

H1) Like in CS in general, **Design and Modeling** work is the **dominant** form of research on the Semantic Web

H2) The **importance** of experimental work on the Semantic Web is **comparable** with computer science in general.

H3) The **importance** of experimental work on the Semantic Web is **increasing** over time.

H4) The **quality** of experimental work on the Semantic Web is **increasing** over time.

H5) Strong experimental work **increases** the **impact** of a paper.

UNIVERSITY OF MANNHEIM

Questions?

On the Status of Experimental Research on the Semantic Web

Heiner Stuckenschmidt, Michael Schuhmacher,
Johannes Knopp, Christian Meilicke, Ansgar Scherp

Data- and Web Science Research Group, University of Mannheim
http://dws.informatik.uni-mannheim.de