

# Nachhaltige Softwareentwicklung für die Digitalisierung der Wissenschaft

Prof. Dr. Wilhelm (Willi) Hasselbring  
Software Engineering Group

<http://se.informatik.uni-kiel.de/>

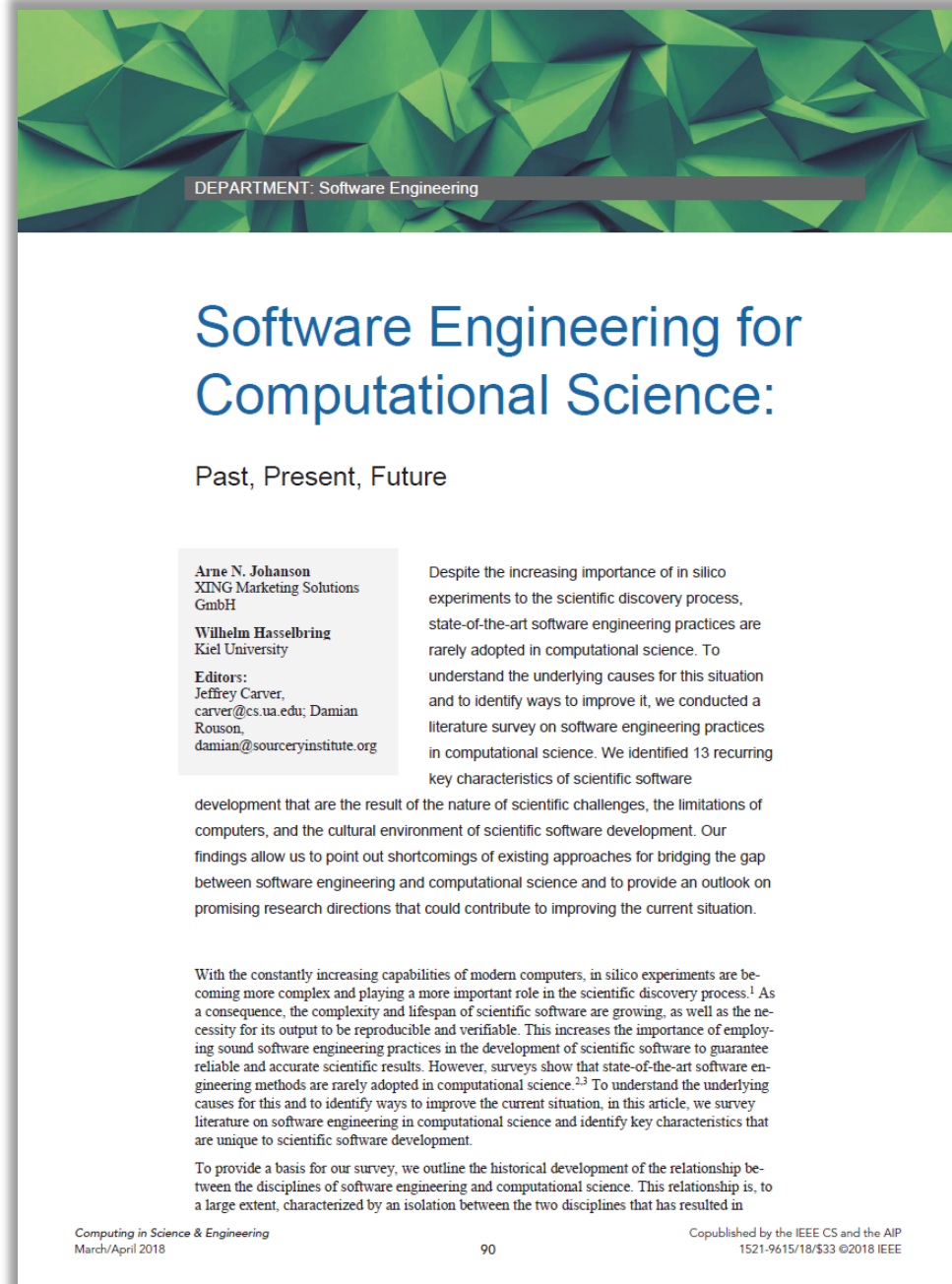
06. Juni 2018



Thema hier:  
**Digitalisierung der  
Wissenschaft  
und  
Reproduzierbarkeit**

<https://doi.org/10.1109/MCSE.2018.108162940>

[Johanson & Hasselbring 2018]



# Gene name errors are widespread in the scientific literature

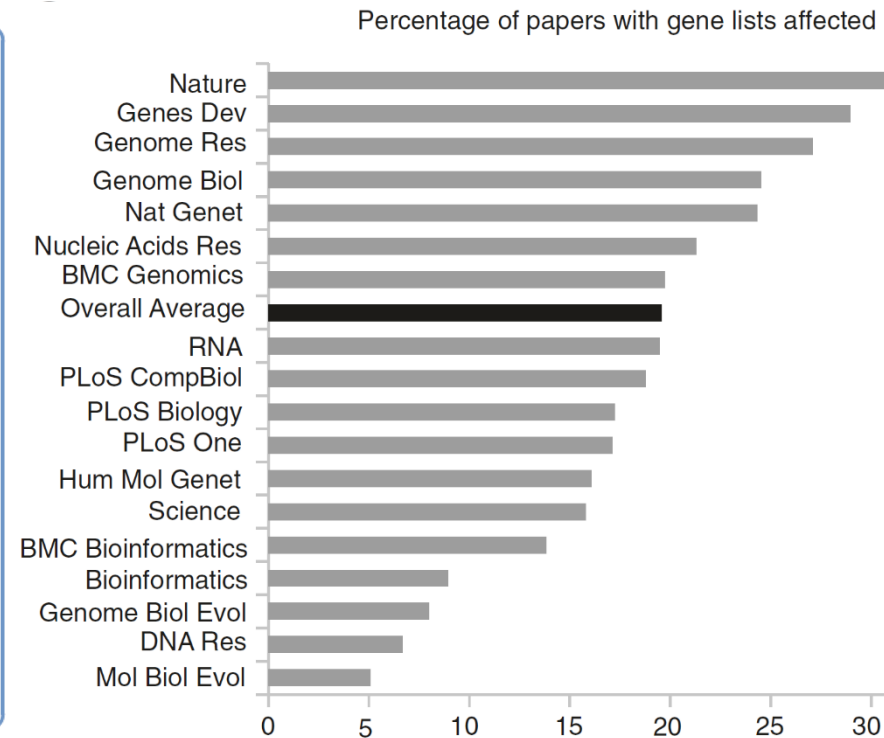
Mark Ziemann<sup>1</sup>, Yotam Eren<sup>1,2</sup> and Assam El-Osta<sup>1,3\*</sup>

## Abstract

The spreadsheet software Microsoft Excel, when used with default settings, is known to convert gene names to dates and floating-point numbers. A programmatic scan of leading genomics journals reveals that approximately one-fifth of papers with supplementary Excel gene lists contain erroneous gene name conversions.

**Keywords:** Microsoft Excel, Gene symbol, Supplementary data

**Abbreviations:** GEO, Gene Expression Omnibus; JIF, journal impact factor



# A Challenge for Arne's PhD research



## Marine Biology Research

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/smar20>

## Estimating the horizontal and temporal overlap of pelagic fish distribution in the Norwegian Sea using individual-based modelling

Kjell Rong Utne<sup>a</sup> & Geir Huse<sup>a</sup>

<sup>a</sup> Institute of Marine Research, Bergen, Norway

Version of record first published: 25 Apr 2012.

<http://dx.doi.org/10.1080/17451000.2011.639781>

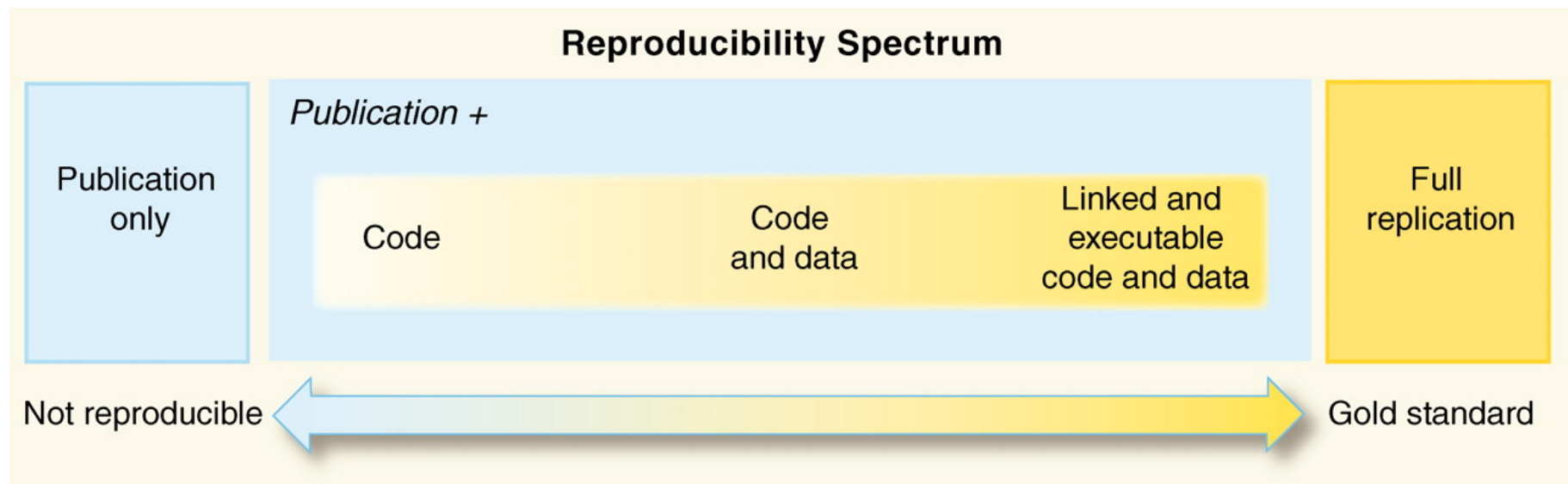
- Utne & Huse provide an abstract (in part mathematical) description of their individual-based model, but:
  - We cannot reconstruct the implementation from the provided information
  - Sources for calibration data are named (some are unpublished) but again we cannot reconstruct the specific input data and parameters used.
- Without releasing the source code **and** the input/configuration data of the model, **reproducibility** of the results is hard or even impossible.

PERSPECTIVE

## Reproducible Research in Computational Science

Roger D. Peng

“Replication is the ultimate standard by which scientific claims are judged.”



## Viewpoint

# The Real Software Crisis: Repeatability as a Core Value

*Sharing experiences running artifact evaluation committees for five major conferences.*

“Science advances faster when we can build on existing results, and when new ideas can easily be measured against the state of the art.”

*At least Repeatability, not necessarily reproducibility*

Several ACM SIGMOD, SIGPLAN, and SIGSOFT conferences have initiated artifact evaluation processes.



### Artifact Review and Badging:

A variety of research communities have embraced the goal of reproducibility in experimental science.

[\[more information\]](#)



#### **Artifacts Evaluated – Functional**

The artifacts associated with the research are found to be documented, consistent, complete, exercisable, and include appropriate evidence of verification and validation.



#### **Artifacts Evaluated – Reusable**

The artifacts associated with the paper are of a quality that significantly exceeds minimal functionality.



#### **Artifacts Available**

Author-created artifacts relevant to this paper have been placed on a publically accessible archival repository.



#### **Results Replicated**

The main results of the paper have been obtained in a subsequent study by a person or team other than the authors, using, in part, artifacts provided by the author.



#### **Results Reproduced**

The main results of the paper have been independently obtained in a subsequent study by a person or team other than the authors, without the use of author-supplied artifacts.



## Artifact Evaluation Track

Chairs: Wilhelm Hasselbring (Kiel University) & Petr Tuma (Charles University)

Some numbers for ICPE 2018:

- 59 submitted full research papers
- 14 accepted full research papers
- 6 submitted artifacts
- 2 accepted artifacts, evaluated as functional
- 0 accepted artifacts, evaluated as reusable



# Is it worth making the effort?

IF I HAVE SEEN FURTHER,  
IT IS BY STANDING  
**ON THE SHOULDERS  
OF GIANTS.**

- ISAAC NEWTON

1676



“Science advances faster when we can build on existing results, and when new ideas can easily be measured against the state of the art.”

[Krishnamurthi & Vitek 2015]

# Impact of Artifact Evaluation

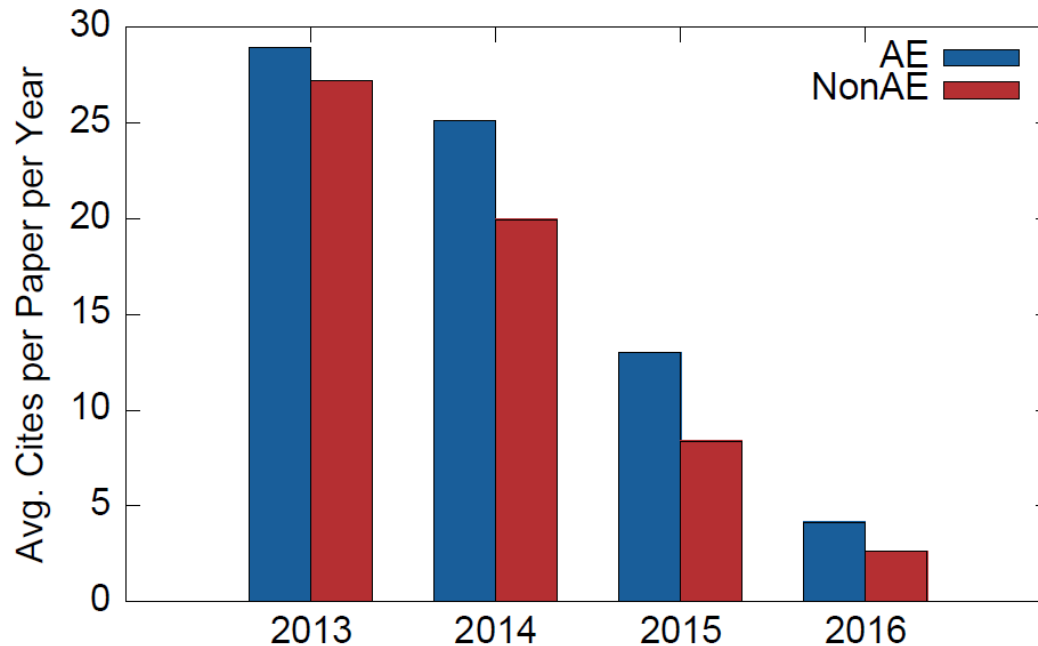


Fig. 1. Average citation counts of AE and non-AE papers for conferences that used AE in 2013 to 2016 (conferences: VISSOFT, PPOPP, POPL, PLDI, PACT, OOPSLA, ISSTA, FSE, ECRTS, ECOOP, CGO, CAV).

[Childers & Chrysanthis 2017]



Software  
Sustainability  
Institute

# What are we doing?

## Hierarchical Software Landscape Visualization for System Comprehension: A Controlled Experiment

Florian Fittkau, Alexander Krause, and Wilhelm Hasselbring  
Software Engineering Group, Kiel University, Kiel, Germany  
Email: {ffi, akr, wha}@informatik.uni-kiel.de



### Exploring Software Cities in Virtual Reality

Florian Fittkau, Alexander Krause, and Wilhelm Hasselbring

(Kiel University, Germany)

Preprint Available Video Info →



06.06.2018

YouTube DE



GitHub

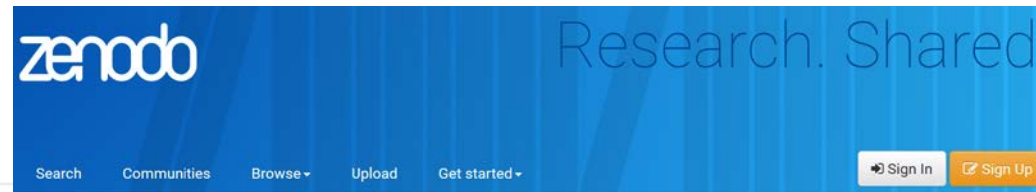
Search GitHub



# ExplorViz

Live trace visualization for large software landscapes

<http://www.explorviz.net>



06 August 2015

Dataset Open access

Experimental Data for: Exploring Software Cities through Virtual Reality

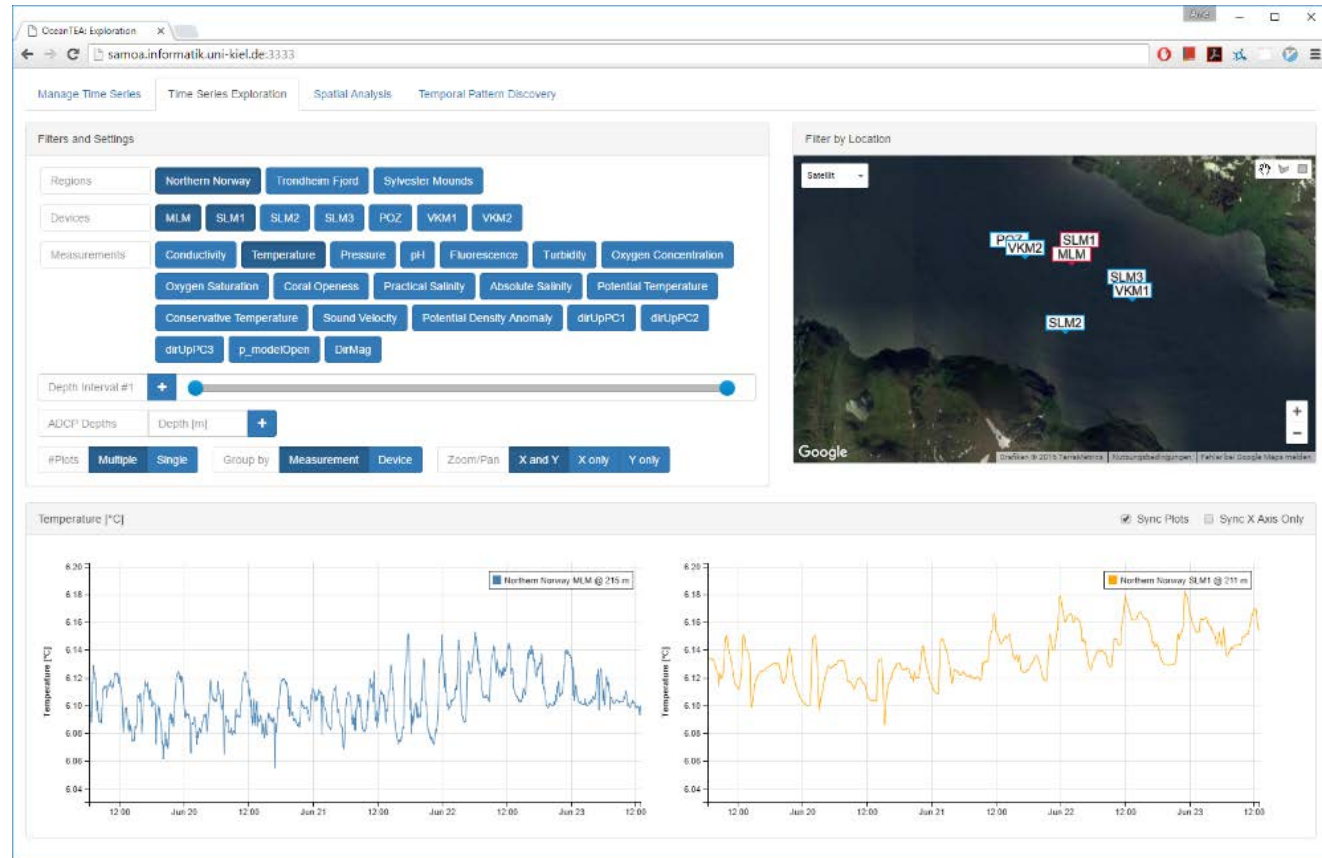
Publication date:  
06 August 2015  
DOI  
DOI 10.5281/zenodo.23168  
Keyword(s):

Explore

City Metaphor  
Conference on  
VISsOFT  
any, 2015.

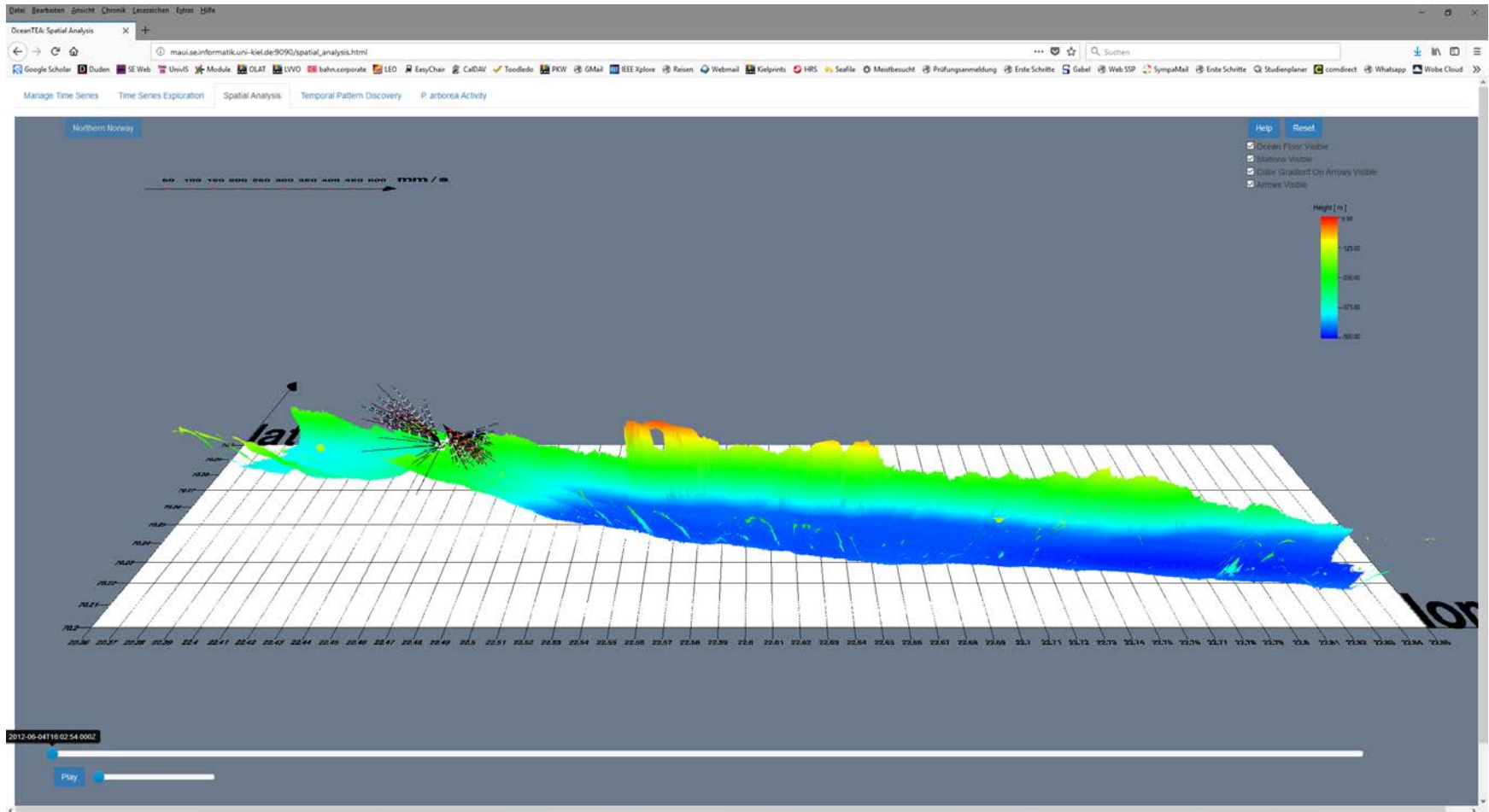
# Cloud-Based Platform for Repeatable Ocean Observation Data Processing

OceanTEA



[Johanson et al. 2016a]

# 4D Spatial Analysis with OceanTEA







# Machine Learning on Ocean Observation Data with OceanTEA

- Paper: <http://dx.doi.org/10.1016/j.ecoinf.2017.02.007>
- Source code: <https://github.com/a-johanson/oceantea>
- Software service with data: <http://maui.se.informatik.uni-kiel.de:9090/> (URL will change, refer to the GitHub repository for updates)

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: [www.elsevier.com/locate/ecoinf](http://www.elsevier.com/locate/ecoinf)

Modeling polyp activity of *Paragorgia arborea* using supervised learning

Arne N. Johanson<sup>a,\*</sup>, Sascha Flögel<sup>b</sup>, Wolf-Christian Dullo<sup>b</sup>, Peter Linke<sup>b</sup>, Wilhelm Hasselbring<sup>a</sup>

<sup>a</sup> Software Engineering Group, Kiel University, Germany  
<sup>b</sup> GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany

CrossMark

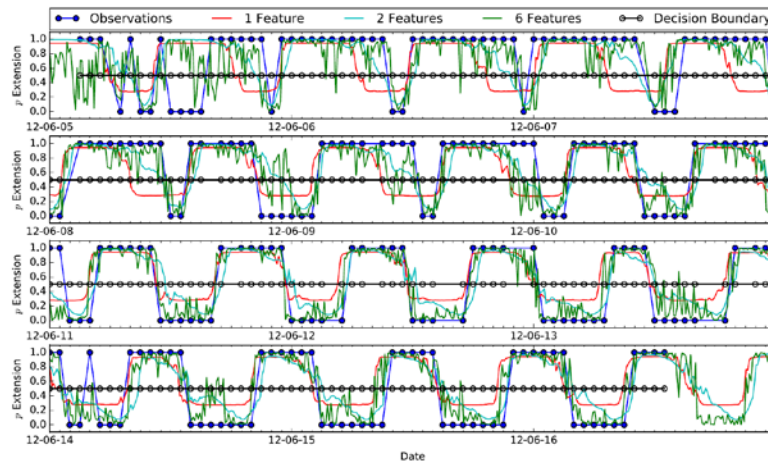
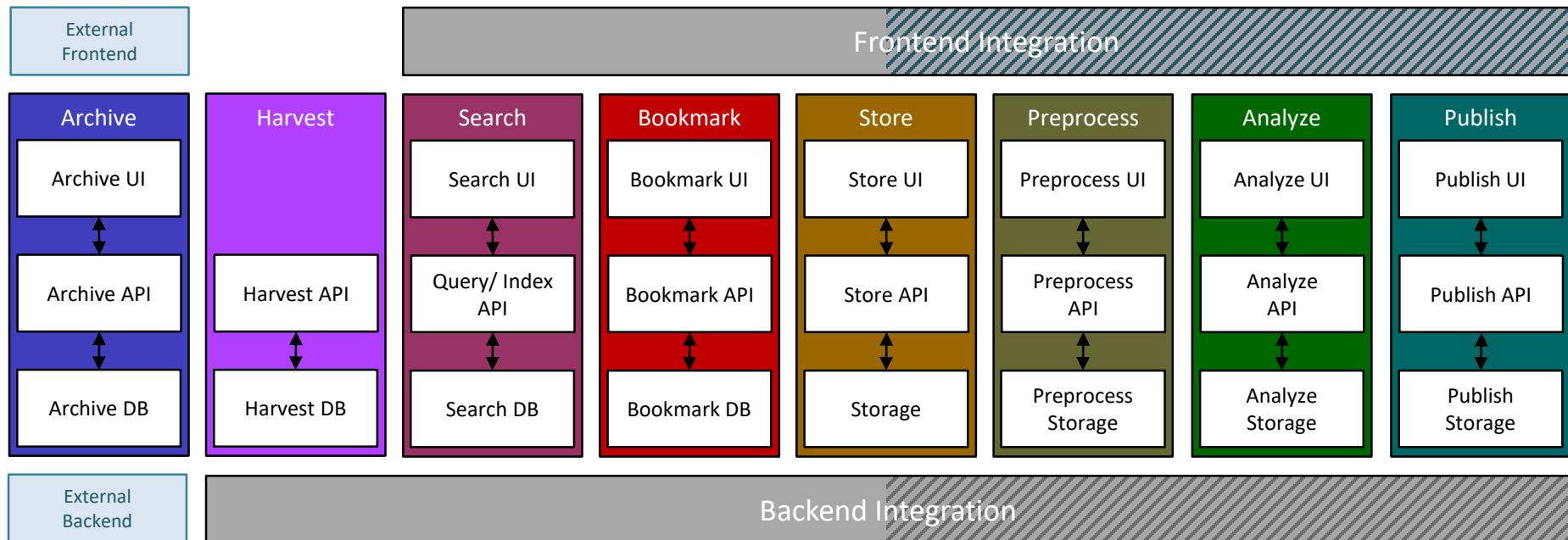


Fig. 7. Degree of extension of coral polyps according to observations as well as to models with one, two, and six features. For an interactive illustration of this figure follow the link provided at: <https://github.com/a-johanson/paragorgia-arborea-activity>.

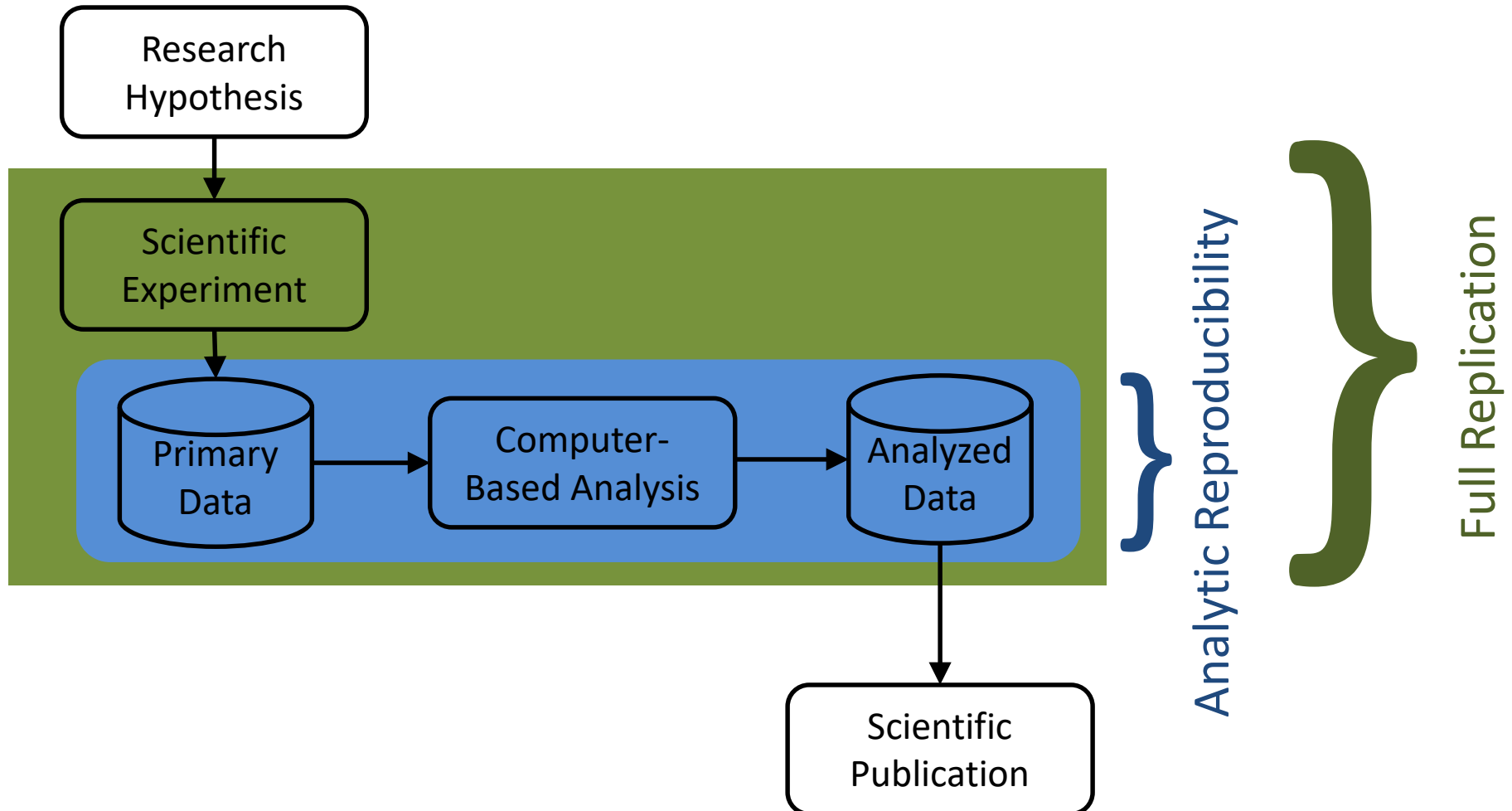


# Generic Research Data Infrastructure





# Analytic Reproducibility vs. Full Replication



# Ausblick: Digital Ocean 2018

<https://digitalewochekiel.de/>

.....

**digitale  
woche 2018**

.....

**Kiel.**  
**Kiel**  
Sailing.City.



ozean der zukunft



**KIEL  
MARINE  
SCIENCE**  
CONNECTED RESEARCH

# References

- [Childers & Chrysanthis 2017] B.R. Childers, P.K. Chrysanthis: “Artifact Evaluation: Is it a Real Incentive?”. In: Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE5.2), 2017
- [Fittkau et al. 2013] F. Fittkau, J. Waller, C. Wulf, W. Hasselbring: “Live Trace Visualization for Comprehending Large Software Landscapes: The ExplorViz Approach”, In: 1st IEEE International Working Conference on Software Visualization (VISOFT 2013).
- [Fittkau et al. 2014] F. Fittkau, P. Stelzer, W. Hasselbring: “Live Visualization of Large Software Landscapes for Ensuring Architecture Conformance”. In: ECSAW 2nd International Workshop on Software Engineering for Systems-of-Systems 2014 (SESoS 2014), 2014, Vienna, Austria.
- [Fittkau et al. 2015a] F. Fittkau, S. Roth, W. Hasselbring: “ExplorViz: Visual Runtime Behavior Analysis of Enterprise Application Landscapes”, In: 23rd European Conference on Information Systems (ECIS 2015).
- [Fittkau et al. 2015b] F. Fittkau, A. Krause, W. Hasselbring: “Hierarchical Software Landscape Visualization for System Comprehension: A Controlled Experiment”. In: 3rd IEEE Working Conference on Software Visualization, 2015.
- [Fittkau et al. 2015c] F. Fittkau, A. Krause, W. Hasselbring: “Exploring Software Cities in Virtual Reality”, In: 3rd IEEE Working Conference on Software Visualization, September 2015, Bremen, Germany.
- [Fittkau et al. 2015d] F. Fittkau, S. Finke, W. Hasselbring, J. Waller: “Comparing Trace Visualizations for Program Comprehension through Controlled Experiments”, In: 23rd IEEE International Conference on Program Comprehension (ICPC 2015), May 2015, Florence.
- [Fittkau et al. 2015e] F. Fittkau, E. Koppenhagen, W. Hasselbring: “Research Perspective on Supporting Software Engineering via Physical 3D Models”. In: 3rd IEEE Working Conference on Software Visualization, September 2015.
- [Fittkau & Hasselbring 2015] F. Fittkau, W. Hasselbring: “Elastic Application-Level Monitoring for Large Software Landscapes in the Cloud”. In: Fourth European Conference on Service-Oriented and Cloud Computing, September, 2015, Taormina, Italy.
- [Fittkau et al. 2017] F. Fittkau, A. Krause, W. Hasselbring: “Software landscape and application visualization for system comprehension with ExplorViz”, In: Information and Software Technology. DOI 10.1016/j.infsof.2016.07.004

# References

- [Hasselbring & Steinacker 2017] W. Hasselbring, G. Steinacker: “Microservice Architectures for Scalability, Agility and Reliability in E-Commerce”, In: Proceedings of the IEEE International Conference on Software Architecture (ICSA 2017), April 2017, Gothenburg, Sweden.
- [Johanson et al. 2016a] A. Johanson, S. Flögel, C. Dullo, W. Hasselbring: “OceanTEA: Exploring Ocean-Derived Climate Data Using Microservices”. In: Sixth International Workshop on Climate Informatics (CI 2016), September 2016, Boulder, Colorado.
- [Johanson et al. 2016b] A. Johanson, W. Hasselbring, A. Oschlies, B. Worm: “Evaluating Hierarchical Domain-Specific Languages for Computational Science: Applying the Sprat Approach to a Marine Ecosystem Model”. In: Software Engineering for Science. CRC Press, pp. 175-200. DOI 10.1201/9781315368924-9.
- [Johanson et al. 2017a] A. Johanson, A. Oschlies, W. Hasselbring, A. Worm: “SPRAT: A spatially-explicit marine ecosystem model based on population balance equations”, In: Ecological Modelling, DOI 10.1016/j.ecolmodel.2017.01.020.
- [Johanson et al. 2017b] A. Johanson, S. Flögel, C. Dullo, P. Linke, W. Hasselbring: “Modeling Polyp Activity of Paragorgia arborea Using Supervised Learning”, In: Ecological Informatics, Elsevier, 2017.
- [Johanson & Hasselbring 2016] A. Johanson, W. Hasselbring: “Effectiveness and efficiency of a domain-specific language for high-performance marine ecosystem simulation: a controlled experiment”, In: Empirical Software Engineering. DOI 10.1007/s10664-016-9483-z.
- [Johanson & Hasselbring 2018] A. Johanson, W. Hasselbring: “Software Engineering for Computational Science: Past, Present, Future”, In: Computing in Science & Engineering. 2018.
- [Krishnamurthi & Vitek 2015] S. Krishnamurthi, J. Vitek: “The real software crisis: repeatability as a core value”. In: Commun. ACM 58, 3 (February 2015), 34-36. DOI 10.1145/2658987
- [Knoche & Hasselbring 2018] H. Knoche, W. Hasselbring: “Using Microservices for Legacy Software Modernization”, In IEEE Software, 2018
- [Tavares de Sousa et al. 2018] N. Tavares de Sousa, W. Hasselbring, T. Weber, D. Kranzlmüller: “Designing a Generic Research Data Infrastructure Architecture with Continuous Software Engineering”, In: 3rd Workshop on Continuous Software Engineering (CSE 2018), March 2018, Ulm, Germany.