

# Sustainable and open ocean research software

*Prof. Dr. Wilhelm (Willi) Hasselbring*

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

.....  
digitale  
woche 2017  
.....

**Kiel**  
**K!el**

Sailing.City.



**KIEL  
MARINE  
SCIENCE**  
CONNECTED RESEARCH



**ozean der zukunft**  
DIE KIELER MEERESWISSENSCHAFTEN

# Abgrenzung zu Forschungsdaten

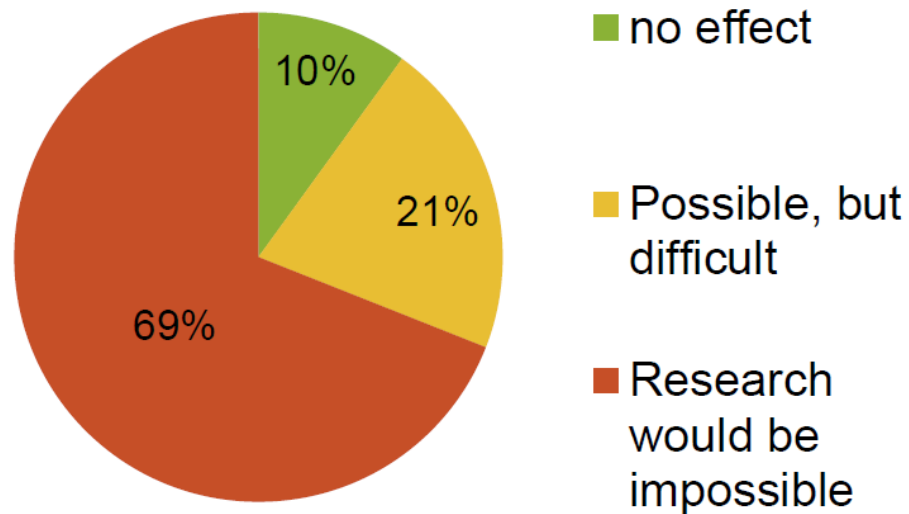
- Software is **executable**,
  - data is not.
- Data provides **evidence**,
  - software provides tools.
- Software is a **creative work**,
  - data are facts or observations.
- Software is updated more **frequently** than papers or data.

Source: Daniel S. Katz et al.

<https://github.com/danielskatz/software-vs-data>

# Relevance of Research Software

**What would happen if you could not longer use research software?**



Hettrick. S. J., et al. (2014). UK Research Software Survey 2014 [Data set]. doi:10.5281/zenodo.14809  
<https://www.software.ac.uk/blog/2016-07-26-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers>

# Gene name errors are widespread in the scientific literature

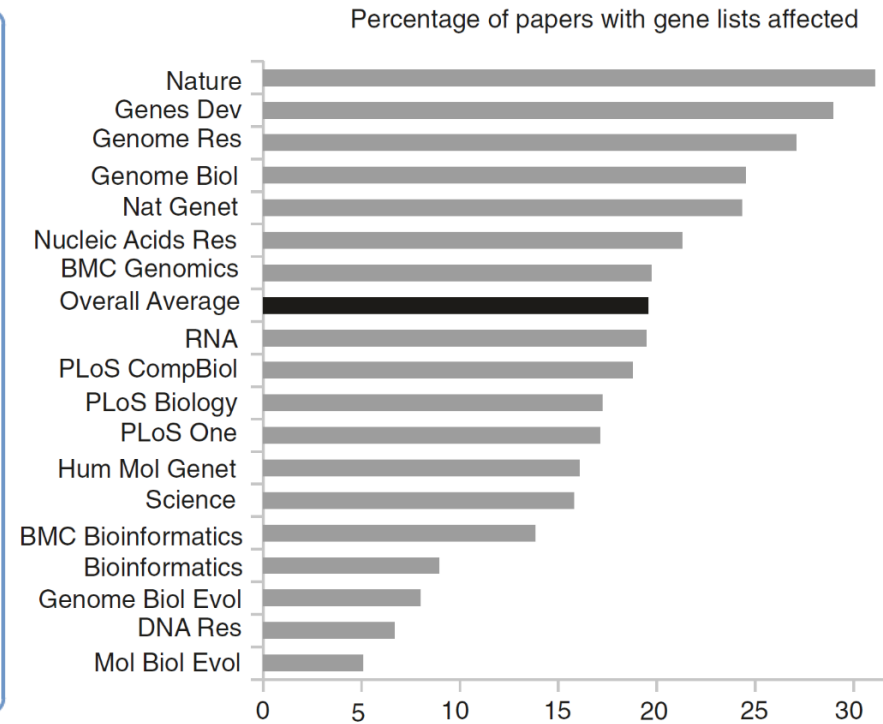
Mark Ziemann<sup>1</sup>, Yotam Eren<sup>1,2</sup> and Assam El-Osta<sup>1,3\*</sup>

## 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



# Research Software Sustainability

Without software, modern research would not be possible. This report recommends practices of software sustainability to minimise the risks of reliability and reproducibility of research.

📅 1 October 2015 - 3 March 2016, 00:00 - 00:00, Berlin, Germany



<http://www.knowledge-exchange.info/event/software-sustainability>



RESEARCH SOFTWARE  
ENGINEERS ASSOCIATION

<http://www.de-rse.org>

# Was tun andere? (hier UK)



Software  
Sustainability  
Institute

<https://www.software.ac.uk/>

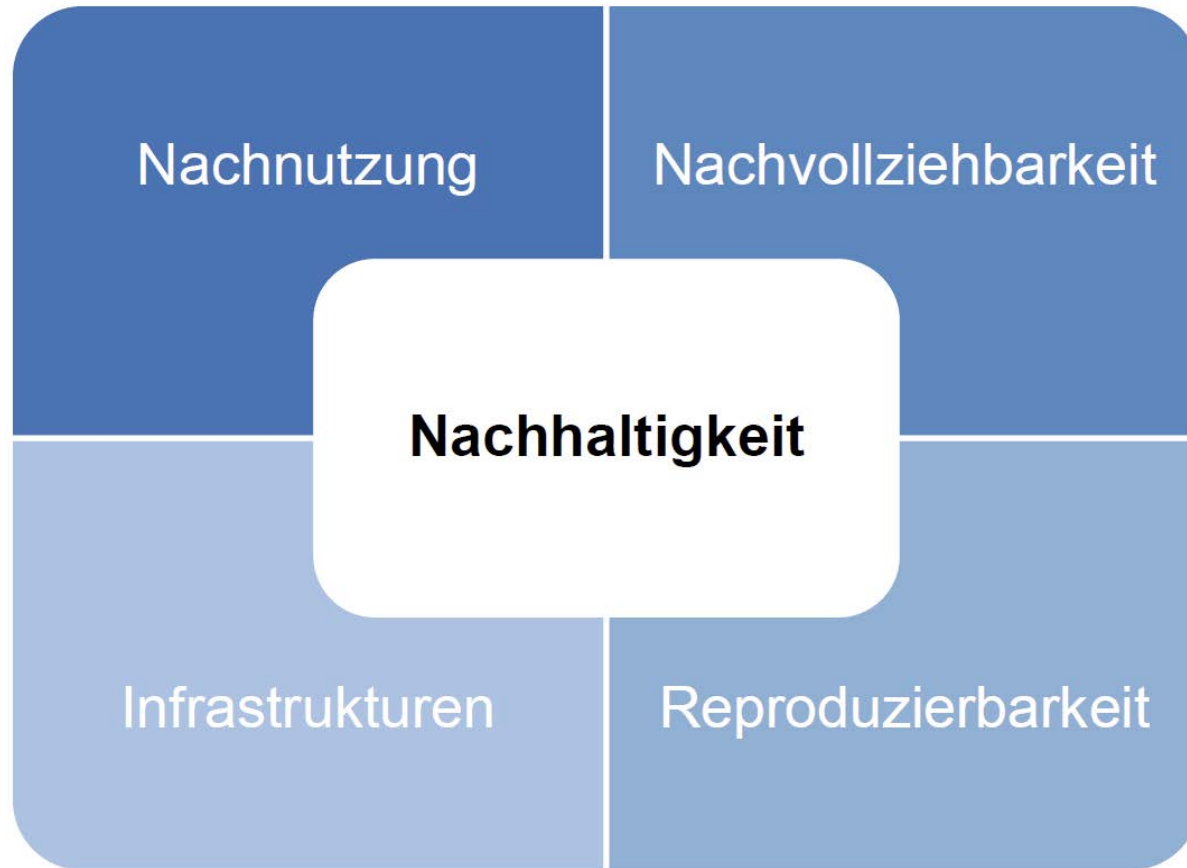


WORKING TOWARDS SUSTAINABLE  
SOFTWARE FOR SCIENCE:  
PRACTICE AND EXPERIENCES

<http://wsspe.researchcomputing.org.uk/>

# Nachhaltigkeit von Forschungssoftware

## Relevanz aus Sicht der DFG als Forschungsförderer



Quelle: Dr. Matthias Katerbow (DFG LIS): „Nachhaltigkeit von Forschungssoftware“,  
Helmholtz Open Science Workshop, Dresden, 2016

# Nachnutzung von Forschungssoftware

- Schaffung von Rahmenbedingungen
  - Bereitstellung, Archivierung, Versionierung, etc.
  - Nutzbarhaltung, Wartung, Support
  - Qualitätssicherung
  - Lizenzmodelle
- Herausforderungen
  - Geschäftsmodelle
  - Langfristige Finanzierung
  - Integration von Weiterentwicklungen und Anpassungen durch Dritte



Quelle: Dr. Matthias Katerbow (DFG LIS)



# Nachvollziehbarkeit der Methodik

- Wissenschaftliche Ergebnisse können nur reproduziert werden, wenn alle wichtigen Schritte vollständig nachvollziehbar sind
  - Was gilt hier für die Verwendung von Forschungssoftware?
    - Verfügbarkeit der **Software**?
    - Verfügbarkeit und **Dokumentation** des Quellcodes?
    - **Parametrisierung** der Software während der Nutzung?
    - **Anpassungen** der Software?
- Nachweise eigener und fremder Vorarbeiten (**Zitate**)
  - Was gilt hier für die Verwendung von Forschungssoftware?
    - Wie soll Software zitiert werden? (Standards z.B. zur executable citation)
    - Wie werden Anwendung und Parametrisierung von Software dokumentiert?



Quelle: Dr. Matthias Katerbow (DFG LIS)

# Reproduzierbarkeit mit und durch Forschungssoftware



Research Software: A missing link in the chain of Open Science!

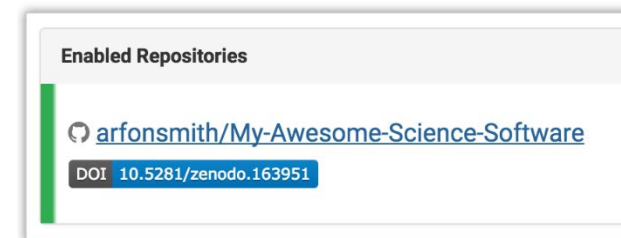


Quelle: Dr. Matthias Katerbow (DFG LIS)

# Infrastruktur zur Nachhaltigkeit von Forschungssoftware

Herausforderungen u.a.:

- Nachhaltige Software-Entwicklung
- Nachweissysteme, Metadaten
- Zitierbarkeit
  - Erste Ansätze z.B.: GitHub  
Release via Zenodo archivieren
- **Reproduzierbarkeit**
- Geschäftsmodelle



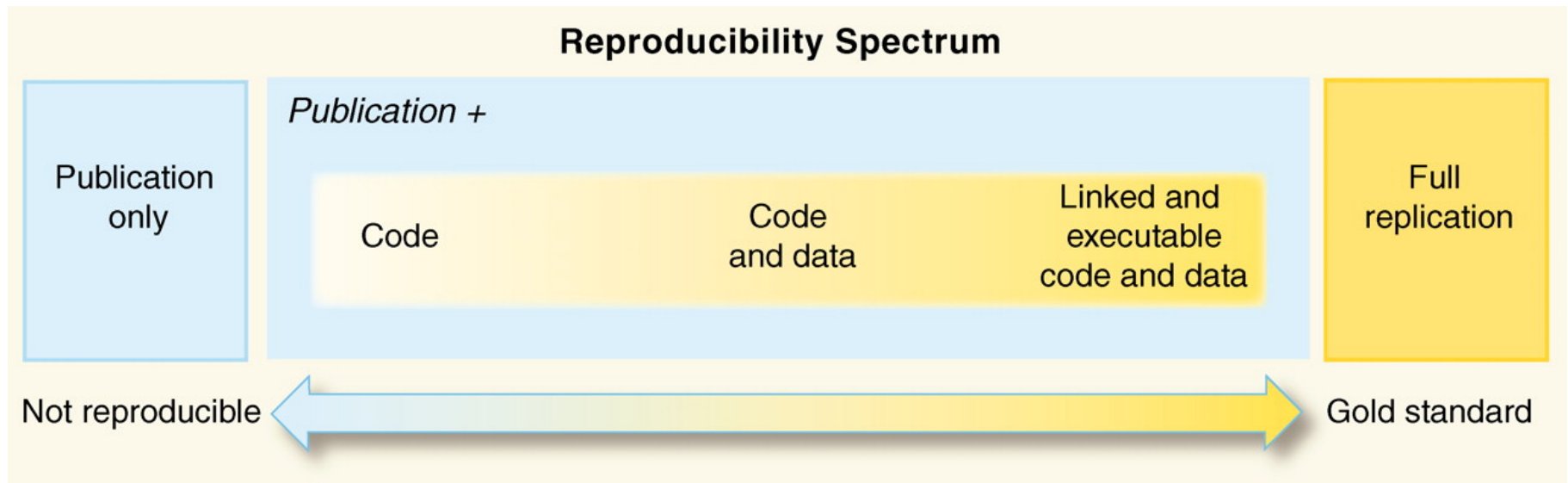
Quelle: Dr. Matthias Katerbow (DFG LIS)

PERSPECTIVE

## Reproducible Research in Computational Science

Roger D. Peng

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



# The case for open computer programs

Darrel C. Ince<sup>1</sup>, Leslie Hatton<sup>2</sup> & John Graham-Cumming<sup>3</sup>

- “We argue that, with some exceptions, anything less than the release of source programs is **intolerable** for results that depend on computation.
- The vagaries of hardware, software and natural language will always ensure that exact **reproducibility remains uncertain**,
  - but withholding code increases the chances that efforts to reproduce results will fail.”

# 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.”

***Repeatability***, not necessarily *reproducibility*

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

# 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

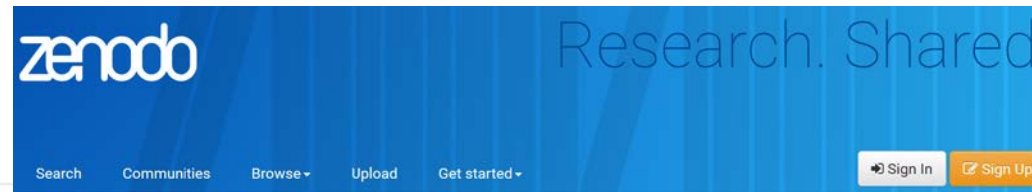
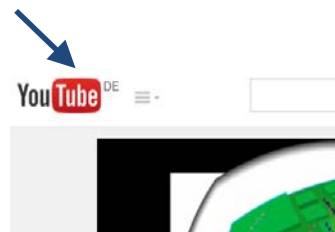


### Exploring Software Cities in Virtual Reality

Florian Fittkau, Alexander Krause, and Wilhelm Hasselbring

(Kiel University, Germany)

Preprint Available Video Info



06 August 2015

Dataset Open access

Publication date: 06 August 2015

DOI: 10.5281/zenodo.23168

Keyword(s): City Metaphor

Explore

Conference on (VISSOFT) January, 2015.

GitHub Search GitHub



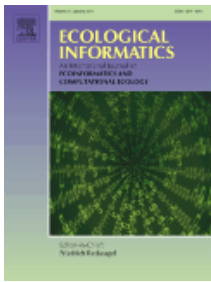
ExplorViz  
Live trace visualization for large software landscapes

<http://www.explorviz.net>

20.09.2017

<https://www.conference-publishing.com/list.php?Event=VISSOFT15>





## Publishing:

- **Paper:** <http://dx.doi.org/10.1016/j.ecoinf.2017.02.007/>
- **Code:** <https://github.com/a-johanson/oceantea> [Johanson et al. 2016]
- **Software service with data:** <http://maui.se.informatik.uni-kiel.de:9090/>

## Modeling Polyp Activity of *Paragorgia arborea* Using Supervised Learning

Arne 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

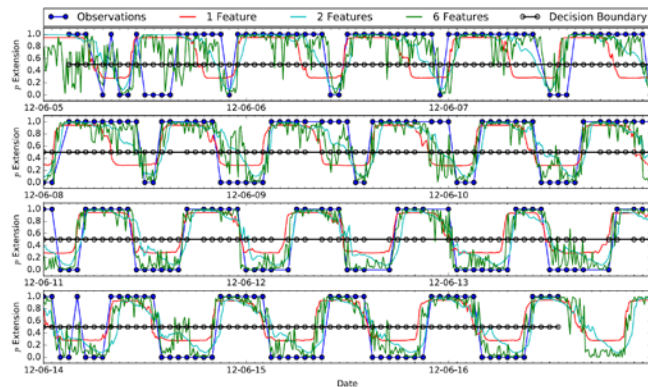
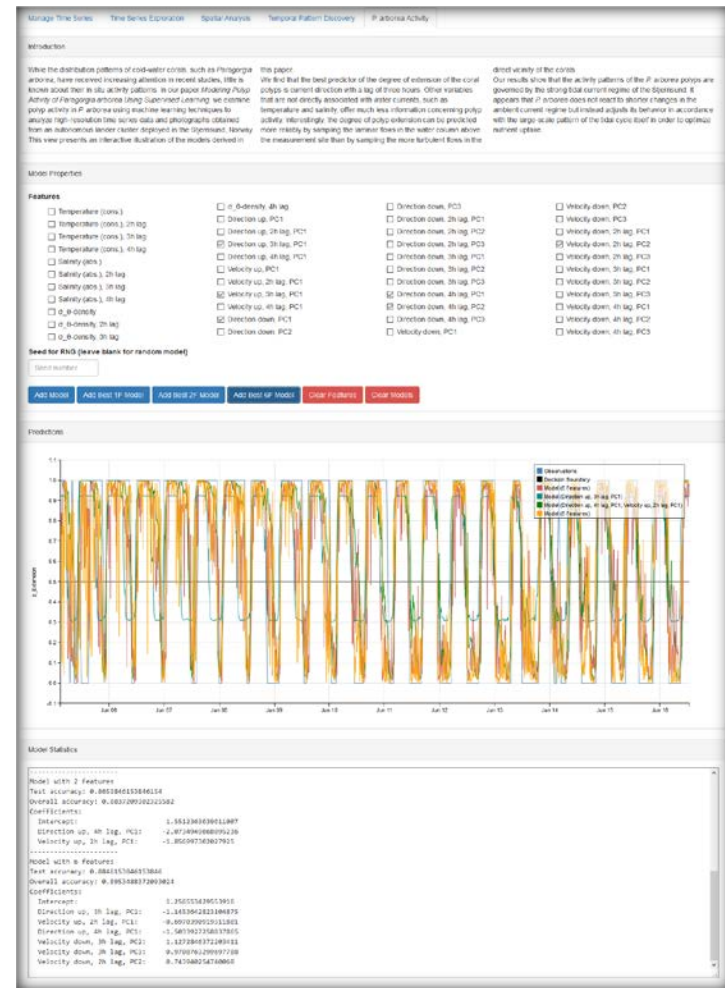


Fig. 7. Degree of extension of coral polyps according to observations as well as to models with one, two, and six features. For an interactive illustration of this figure follow the link provided at: <https://github.com/a-johanson/paragorgia-arborea-activity>.

[Johanson et al. 2017]







## ROLL: Fast In-Memory Generation of Gigantic Scale-free Networks

Full Text: [PDF](#)

2016 Article

see [source materials](#) below for [more options](#)

[Artifacts](#)

**Note:**

**Computationally Replicable.** The experimental results of this paper were replicated by a SIGMOD Review Committee and were found to support the central results reported in the paper. Details of the review process are found [here](#)

- Artifacts Available
- Results Replicated

Authors: [Ali Hadian](#)

[Sadeqh Nobari](#)  
[Behrooz Min](#)

[Qiang Qu](#)

Published in:

- Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data
- Pages 182–193
- San Francisco, CA, USA: ACM New York, 2016
- [table of contents](#)

**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.

# Impact of Artifact Evaluation

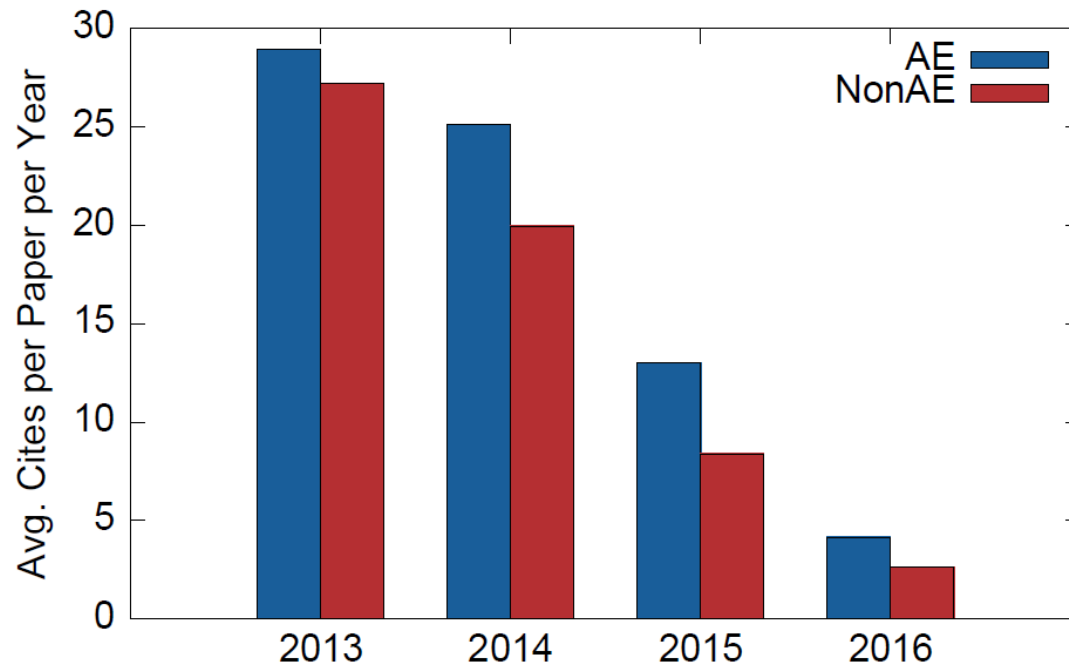
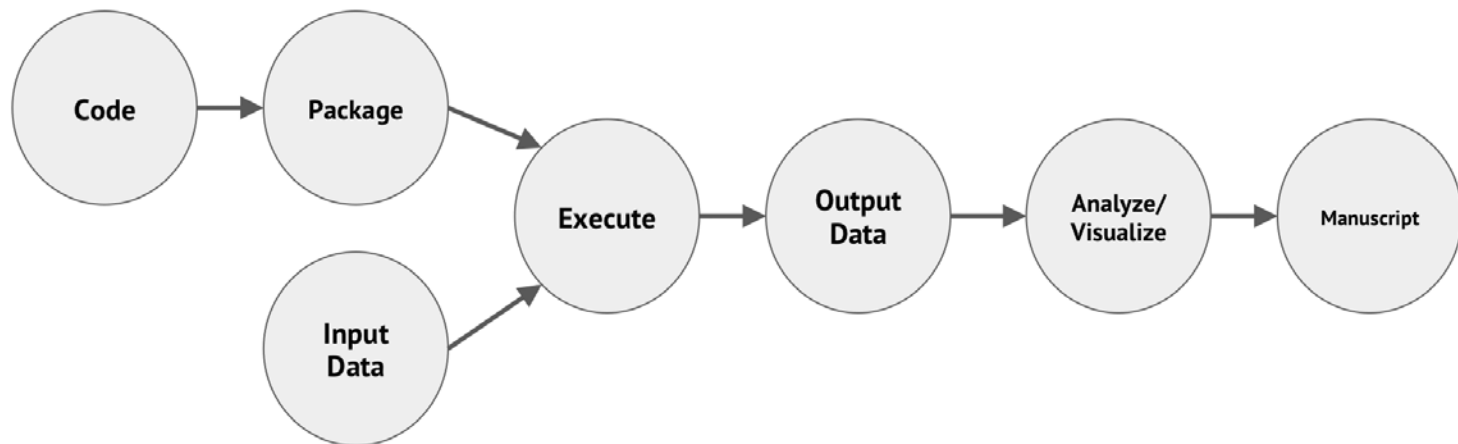


Fig. 1. Average citation counts of AE and non-AE papers for conferences that used AE in 2013 to 2016 (conferences: VISSOFT, PPOPP, POPL, PLDI, PACT, OOPSLA, ISSTA, FSE, ECRTS, ECOOP, CGO, CAV).

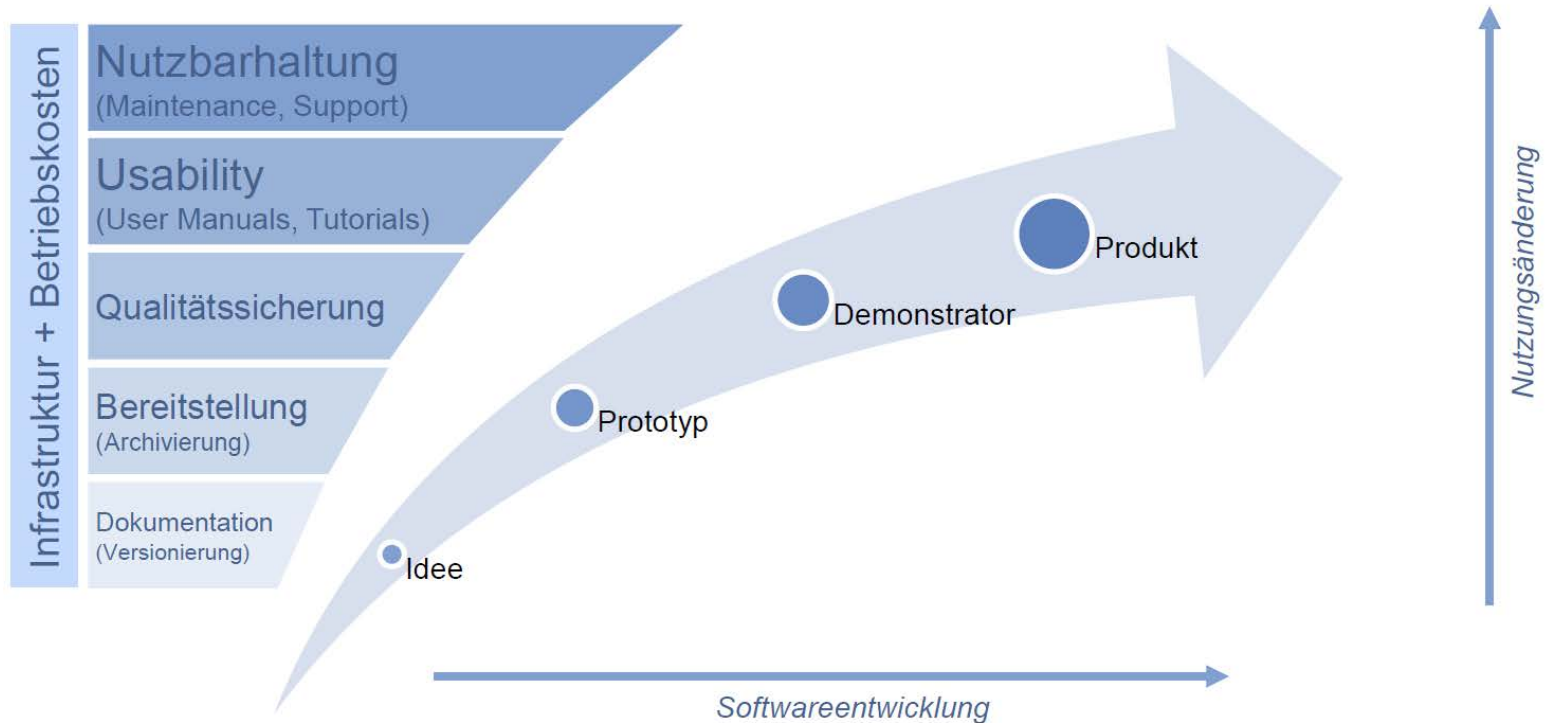
Source: Bruce R. Childers, Panos K. Chrysanthis: „Artifact Evaluation: Is it a Real Incentive?“  
In: Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE5.2), 2017

# Popper – Practical Falsifiable Research

<http://falsifiable.us/>



# Förderangebote der DFG



**Information für die Wissenschaft Nr. 71 | 2. November 2016**

## Nachhaltigkeit von Forschungssoftware

**DFG erbittet Anträge zur Nutzbarmachung und Nachhaltigkeit von Forschungssoftware**

Quelle: Dr. Matthias Katerbow (DFG LIS)

# Summary

Key question:

- How to achieve **repeatability, extensibility** and **sustainability** of ocean research software?

Solution approach:

- Digital Ocean could invent and introduce new digital research infrastructures as supporting structures (mainly software) that are themselves **sustainable**.

Sustainable Development Goal 9

- “Build resilient **infrastructure**, promote inclusive and sustainable industrialization and foster innovation. [...]”



# 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. 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. 2016] 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
- [Ince et al. 2012] D.C. Ince, L. Hatton, J. Graham-Cumming: “The case for open computer programs”. Nature, 482-485, 2012.
- [Johanson et al. 2016] 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. 2017] A. Johanson, S. Flögel, C. Dullo, P. Linke, W. Hasselbring: “Modeling Polyp Activity of *Paragorgia arborea* Using Supervised Learning”, In: Ecological Informatics, Elsevier, 2017.
- [Krishnamurthi & Vitek 2015] S. Krishnamurthi, J. Vitek: “The real software crisis: Repeatability as a core value”. Communications of the ACM, 58(3), 34-36, 2015.
- [Peng 2011] R.D. Peng: “Reproducible research in computational science. Science”, 334(6060), 1226-1227, 2011.
- [Ziemann et al. 2016] M. Ziemann, Y. Eren, A. El-Osta: “Gene name errors are widespread in the scientific literature”. Genome biology, 17(1), 2016.