

# Artefakte aus der Softwareentwicklung als Daten

## aus der Perspektive der Software-Technik-Forschung

Webtalk NFDIxCS

Anne Koziolk

ARCHITECTURE-DRIVEN REQUIREMENTS ENGINEERING,  
INSTITUT FÜR PROGRAMMSTRUKTUREN UND DATENORGANISATION, KIT-FAKULTÄT FÜR INFORMATIK



# Überblick

- Welche Forschungsdaten?
- Wo stehen wir heute?
  - ACM Badges
  - Einschätzung zum Stand der FAIR Prinzipien
- Was sind Herausforderungen?

Dabei meine Perspektive: Software-Technik-Forschung

Für die Diskussion später: Passt es zu anderen Gebieten,  
in denen Software Forschungsdaten sind?

# Welche Forschungsdaten?

## ■ Von Forschern erzeugt

### ■ Software

- Skripte für Datenanalyse
- Programme, die neue Algorithmen umsetzen
- Software-Tools

### ■ Modelle



Fokus im Folgenden: Artefakte

## ■ Beobachtungen

- Unter Beteiligung von Menschen oder Umgebung beobachtet
- Automatisiert erzeugt

## ■ Weitere Eingabedaten

# Wo stehen wir heute in der SWT-Forschung?

- Zunehmende Bedeutung
- „Artifact Evaluation Track“ auf den meisten SWT-Konferenzen
- ACM Artifact Reviewing and Badging
  - “By "artifact" we mean a digital object that was either created by the authors to be used as part of the study or generated by the experiment itself. For example, artifacts can be software systems, scripts used to run experiments, input datasets, raw data collected in the experiment, or scripts used to analyze results.”
- Aber
  - Veröffentlichung der Artefakte keine Voraussetzung für Publikationen
  - Wiederverwendung von Artefakten weiterhin schwierig

# ACM Artifact Reviewing and Badging

## Functional



Artifacts documented, consistent, complete, exercisable, and include appropriate evidence of verification and validation

## Reusable



Functional + very carefully documented and well-structured to the extent that reuse and repurposing is facilitated. In particular, norms and standards of the research community for artifacts of this type are strictly adhered to.

## Available



Placed on a publically accessible archival repository. A DOI or link to this repository along with a unique identifier for the object is provided.

## Replicated



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.

## Reproduced



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.

<https://www.acm.org/publications/policies/artifact-review-badging>

# SWT Adaption der ACM Badges

## Functional

*No Badge*

Artifacts documented, consistent, complete, exercisable, and include appropriate evidence of verification and validation

## Reusable



Functional + very carefully documented and well-structured to the extent that reuse and repurposing is facilitated. In particular, norms and standards of the research community for artifacts of this type are strictly adhered to.

## Available



*Functional* + placed on a publicly accessible archival repository. A DOI or link to this repository along with a unique identifier for the object is provided.

## Replicated



*Available* + 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.

## Reproduced



*Available* + 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*.

vom Artifact Evaluation Track der ICSE Konferenz, <https://conf.researchr.org/track/icse-2020/icse-2020-Artifact-Evaluation#Call-for-Submissions>

# FAIRe Forschungsdaten in SWT?

**F**  
indable



Praktiken 

Umsetzung 

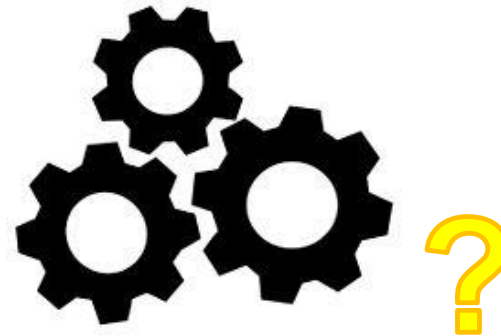
**A**  
ccessible



Praktiken 

Umsetzung 

**I**  
nteroperable



Praktiken 

Umsetzung 

**R**  
eusable



Praktiken 

Umsetzung 

Image „[FAIR data principles](#)“ by [SangyaPundir](#) is licensed under [CC-BY-SA-4.0](#) on Wikipedia

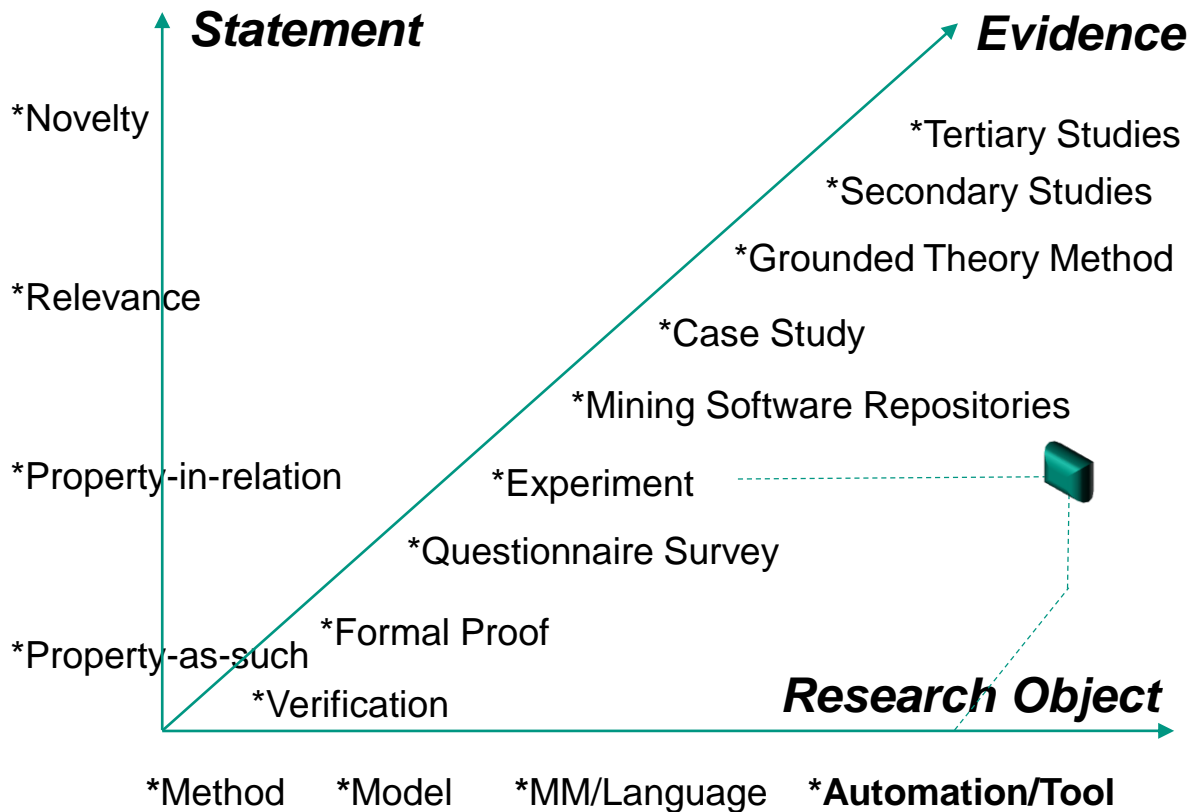


# Herausforderungen

- Anreizsysteme
  - Einreichungen von Forschungsdaten obligatorisch?
- Anforderungen an Artefakte im Begutachtungsprozess
  - Für Auffindbarkeit und Zugänglichkeit leicht umsetzbar
    - Nicht möglich, wenn Daten von Industriepartnern
  - Bewertbarkeit?
  - Nachnutzbarkeit?
- Nachvollziehbarkeit von wiss. Ergebnissen
  - Wie genaue Konfiguration dokumentieren?
  - Inwieweit unterstützen die Artefakte die Aussagen der Publikation?



# Aktuelle Forschung: Validitäts-Aussagen in der SWT-Forschung



Data Item (Artefact)	Data Description
<b>Research Object</b>	Investigated object of research (dt. „Untersuchungsgegenstand“)
<b>Statement</b>	Kind of statement: (i) One paper can be related to many statements. (ii) Each statement may have given different kinds and strength of evidence. (iii) Papers may add evidence to a statement of other papers with additional data or arguments. (iv) Papers may contain statements that refine statements of other papers or (v) even contradict them with some evidence.
<b>Evidence</b>	Research method design and evidence of statement validity
<b>Validation Question (VQ)</b>	Template for validation question: * e.g., „How <i>effective</i> is Tool A in comparison to Tool B?“

Gemeinsame Arbeit mit Angelika Kaplan und Ralf Reussner, Folie von Angelika Kaplan

# Überblick

- Welche Forschungsdaten?
- Wo stehen wir heute?
  - ACM Badges
  - Einschätzung zum Stand der FAIR Prinzipien
- Was sind Herausforderungen?
  - Anreizsysteme
  - Begutachtungsprozesse
  - Nachvollziehbarkeit von wissenschaftlichen Ergebnissen

Dabei meine Perspektive: Software-Technik-Forschung

Für die Diskussion später: Passt es zu anderen Gebieten,  
in denen Software Forschungsdaten sind?