



Nachhaltigkeit von Forschungssoftware

Forschungsdaten in den Geisteswissenschaften

FORGE 2016: Jenseits der Daten

Dr. Matthias Katerbow (DFG III-LIS), Twitter: @katerbow

14. September 2016

Inhalt

1. Einführung
2. Forschungssoftware
3. Nachhaltigkeit
4. Förderangebote der DFG
5. Hinweise zu Workshops



1. **Einführung**
2. Forschungssoftware
3. Nachhaltigkeit
4. Förderangebote der DFG
5. Hinweise zu Workshops



Relevanz zur Förderung von Forschungssoftware

Aus Sicht der DFG als Forschungsförderer

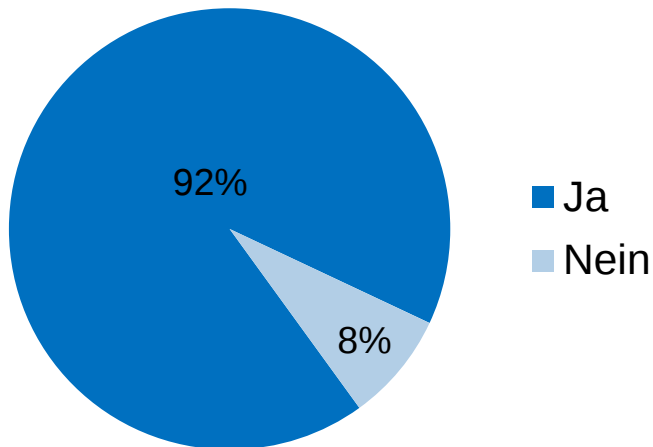
- ▶ Nachnutzung
- ▶ Nachhaltigkeit
- ▶ Reproduzierbarkeit

- ▶ Folge: Entwicklung von infrastrukturellen Rahmenbedingungen

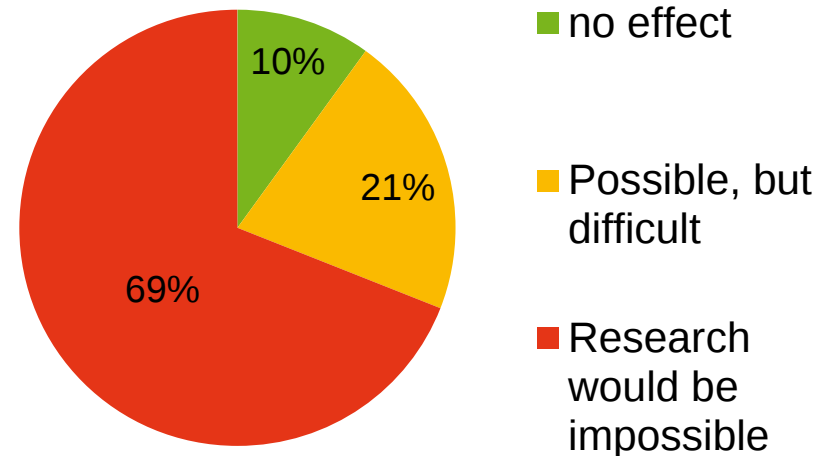
Relevanz zur Förderung von Forschungssoftware

UK Research Software Survey 2014 (response of 417 researchers)

**Do you use
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>

Bedarfe in den Geisteswissenschaften?

Sammlung von Diskussionspunkten

- ▶ Was würde in den Geisteswissenschaften geschehen, wenn Forschungssoftware nicht länger verfügbar wäre?

Inhalt

1. Einführung
2. **Forschungssoftware**
3. Nachhaltigkeit
4. Förderangebote der DFG
5. Hinweise zu Workshops



Was ist Forschungssoftware?

Allgemeine Definition der DFG

- ▶ Funktional: Forschungssoftware wird zum Beispiel zur Generierung, Verarbeitung, Analyse und Visualisierung von Forschungsdaten genutzt
- ▶ Zweck: Forschungssoftware sind die eigens zum wissenschaftlichen Erkenntnisgewinn erstellten Software-Anwendungen und Software-Bibliotheken
- ▶ Abgrenzung: kommerzielle oder (im Open Source) verfügbare Standard-Software ist nicht gemeint (z.B. Office)

Was ist Forschungssoftware?

Abgrenzung zu Forschungsdaten

- ▶ Software is executable, unlike papers or data
- ▶ Data provides evidence, software provides a tool
- ▶ Software is a creative work, data are facts
- ▶ Software can be used to express or explain concepts, unlike data
- ▶ Software is active, data is passive
- ▶ Software is updated more frequently than papers or data
- ▶ The lifetime of software is generally not as long as that of data

Daniel S. Katz et al. <https://github.com/danielskatz/software-vs-data>

Forschungssoftware in den Geisteswissenschaften

Welche Forschungssoftware gibt es? Wie wird sie „betrieben“?

- ▶ <http://digitalhumanities.org/dhq/vol/10/1/000235/000235.html#d52216e653>
- ▶ <http://software.esciencecenter.nl/software/>
- ▶ <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2014-4.pdf>
- ▶ <http://wikis.fu-berlin.de/display/tdr/Tools>
- ▶ <https://de.dariah.eu/fachwissenschaftliche-dienste>
- ▶ <https://www.hastac.org/blogs/melody-dworak/2012/04/23/digital-humanities-resource-guide>
- ▶ https://www.zotero.org/groups/doing_digital_humanities_-_a_dariah_bibliography



► Research Activities

- Capture (74)
- Creation (54)
- Enrichment (49)
- Analysis (86)
- Interpretation (20)
- Storage (43)
- Dissemination (95)
- Uncategorized (19)



- TaDiRAH breaks down the research lifecycle into high-level "goals", each with a set of "methods". The goals and methods are the core of TaDiRAH, and are a controlled vocabulary, albeit one open to periodic revision.

TaDiRAH: a Case Study in Pragmatic Classification
Luise Borek, Technical University of Darmstadt
Quinn Dombrowski, Miami University
Jody Perkins, Miami University
Christof Schöch, University of Würzburg
<http://digitalhumanities.org/dhq/vol/10/1/000235/000235.html#d52216e653>

Nachhaltigkeit von Forschungssoftware / Dr. Matthias Katerbow

Bedarfe in den Geisteswissenschaften?

Sammlung von Diskussionspunkten

- ▶ Was würde in den Geisteswissenschaften geschehen, wenn Forschungssoftware nicht länger verfügbar wäre?
- ▶ Welche Forschungssoftware gibt es? Wie wird sie „betrieben“?
- ▶ Welche Forschungssoftware ist unerlässlich?
- ▶ Wie können Community-Angebote langfristig gewährleistet werden?

Inhalt

1. Einführung
2. Forschungssoftware
- 3. Nachhaltigkeit**
4. Förderangebote der DFG
5. Hinweise zu Workshops



Grundlegende Wissenschaftsprinzipien

Transparenz, Nachvollziehbarkeit und freier Diskurs

- ▶ 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?

Grundlegende Wissenschaftsprinzipien

Veränderung des wissenschaftlichen Arbeitens

- Technische Entwicklungen sind der Anlass für Veränderungen nahezu jeden Schrittes des wissenschaftlichen Arbeitens

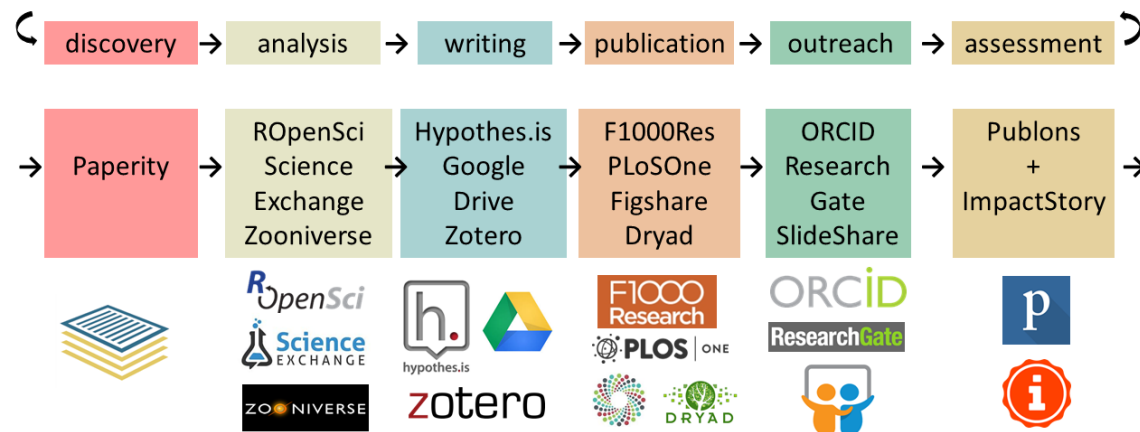


Bild-Quelle:
<https://101innovations.wordpress.com/>,
<http://innoscholcomm.silk.co/>

- Auf politischer und organisatorischer Ebene muss dafür Sorge getragen werden, dass die grundlegenden Wissenschaftsprinzipien – **Offenheit, Transparenz, Nachvollziehbarkeit und freier Diskurs** – unter den sich verändernden Bedingungen weiterhin aufrecht erhalten werden

Reproduzierbarkeit / Nachvollziehbarkeit

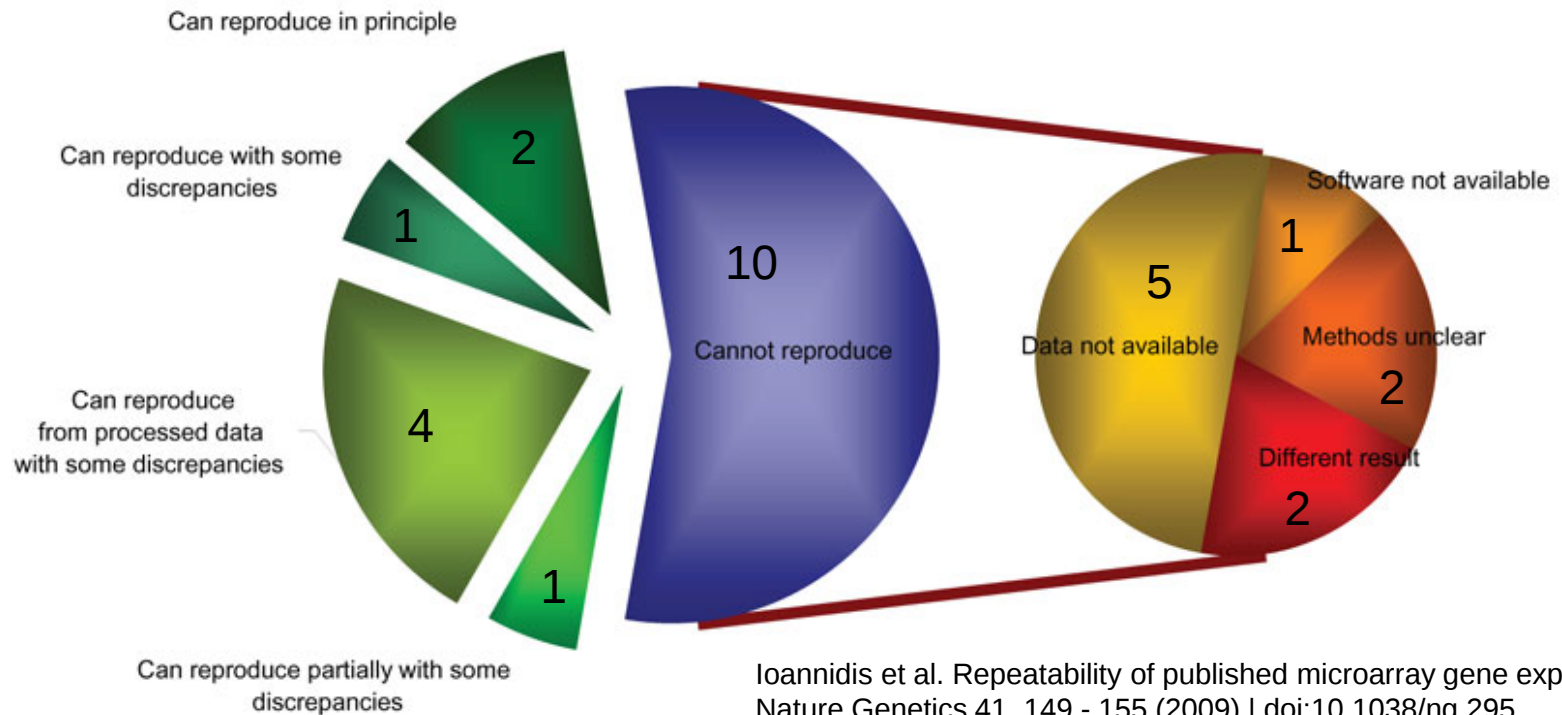
Research Software: A missing link in the chain of Open Science



Reproduzierbarkeit

Beispiel 1: Biomedizin

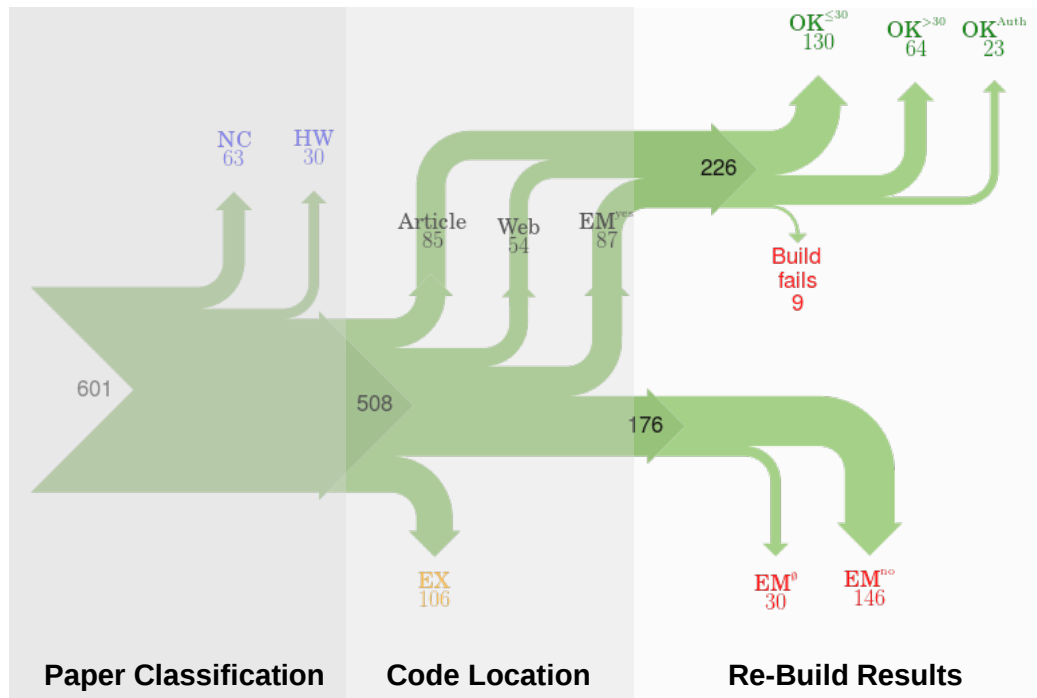
Evaluation der Replikation der Datenanalyse von 18 Artikeln zum Thema *microarray-based gene expression*, die *Nature Genetics* von 2005 bis 2006 publizierte.



Ioannidis et al. Repeatability of published microarray gene expression analyses
Nature Genetics 41, 149 - 155 (2009) | doi:10.1038/ng.295

Reproduzierbarkeit

Beispiel 2: Repeatability in Computer Science



Re-Build results:
Failed = 185
Built = 217

Repeatability and Benefaction in Computer Systems Research. A Study and a Modest Proposal
Christian Collberg, Todd Proebsting, Alex M Warren February 27, 2015
<http://repeatability.cs.arizona.edu/index.html>

Bedarfe in den Geisteswissenschaften?

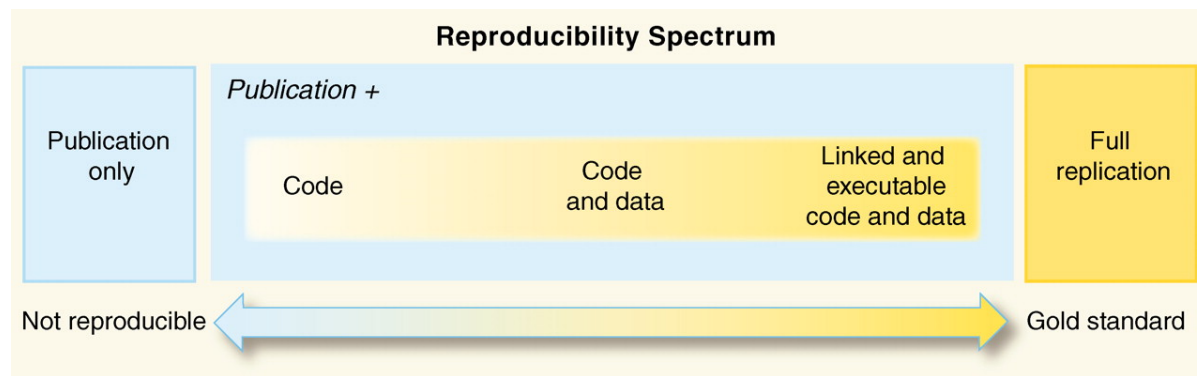
Sammlung von Diskussionspunkte

- ▶ Was würde in den Geisteswissenschaften geschehen, wenn Forschungssoftware nicht länger verfügbar wäre?
- ▶ Welche Forschungssoftware gibt es? Wie wird sie „betrieben“?
- ▶ Welche Forschungssoftware ist unerlässlich?
- ▶ Wie können Community-Angebote langfristig gewährleistet werden?
- ▶ Gibt es Reproduktionsstudien in den Geisteswissenschaften?

Reproduzierbarkeit

Reproducibility Spectrum

“Replication is the ultimate standard by which scientific claims are judged. With replication, independent investigators address a scientific hypothesis and build up evidence for or against it.”



Roger D. Peng: Reproducible Research in Computational Science
Science 02 Dec 2011: Vol. 334, Issue 6060, pp. 1226-1227
DOI: 10.1126/science.1213847

Nachhaltigkeit

Kriterien für nachhaltige Software-Entwicklung

	Top 10 Ranked Metrics	Avg Importance ^a	Avg Implementability ^b	Avg Sum Priority Score ^c	Manual Priority Evaluation ^d
1. Version control	Is version control used?	5	4.6	9.6	1
2. Code reviews	Is the software discoverable?	4.1	5	9.1	2
3. Automated testing	Is an automated build system used?	4.6	3.9	8.4	3
4. Not reinventing the wheel	Are test data available?	3.8	4	7.8	4
5. Discoverability	Does software contain parts that reimplement existing technology?	4.4	2.9	7.3	5
6. Reusability of source code	Is the software compliant with community standards?	4.1	2.5	6.6	6
7. Reusability of software	Are code reviews performed?	3.4	2.8	6.1	7
8. Licensing	Is automated testing performed?	3.5	3.1	6.6	8
9. Issue tracking / bug tracking	Is the code documented?	2.4	4.3	6.6	9
10. Support process	How high is the code complexity?	3.5	2.9	6.4	10
11. Compliance with community standards					
12. Buildable code					
13. Open development					
14. Making data available					
15. Documentation					
16. Simplicity					
17. Dependency management					

Artaza H, Chue Hong N, Corpas M *et al.* Top 10 metrics for life science software good practices [version 1; referees: awaiting peer review]. *F1000Research* 2016, **5**(ELIXIR):2000 (doi: [10.12688/f1000research.9206.1](https://doi.org/10.12688/f1000research.9206.1))

Zitierbarkeit und Reproduzierbarkeit

Was sind die Herausforderungen?

- ▶ „The challenge is not just the textual citation of software in a paper, but the more general identification of software used within the research process“ (Software Citation Principles, S. 2)
- ▶ „Citation of specific software used is necessary for reproducibility, but is not sufficient. Additional information such as configurations and platform issues are also needed“ (Software Citation Principles, S. 3)

Smith AM, Katz DS, Niemeyer KE, FORCE11 Software Citation Working Group. (2016) Software Citation Principles. PeerJ Preprints 4:e2169v2 <https://doi.org/10.7287/peerj.preprints.2169v2>

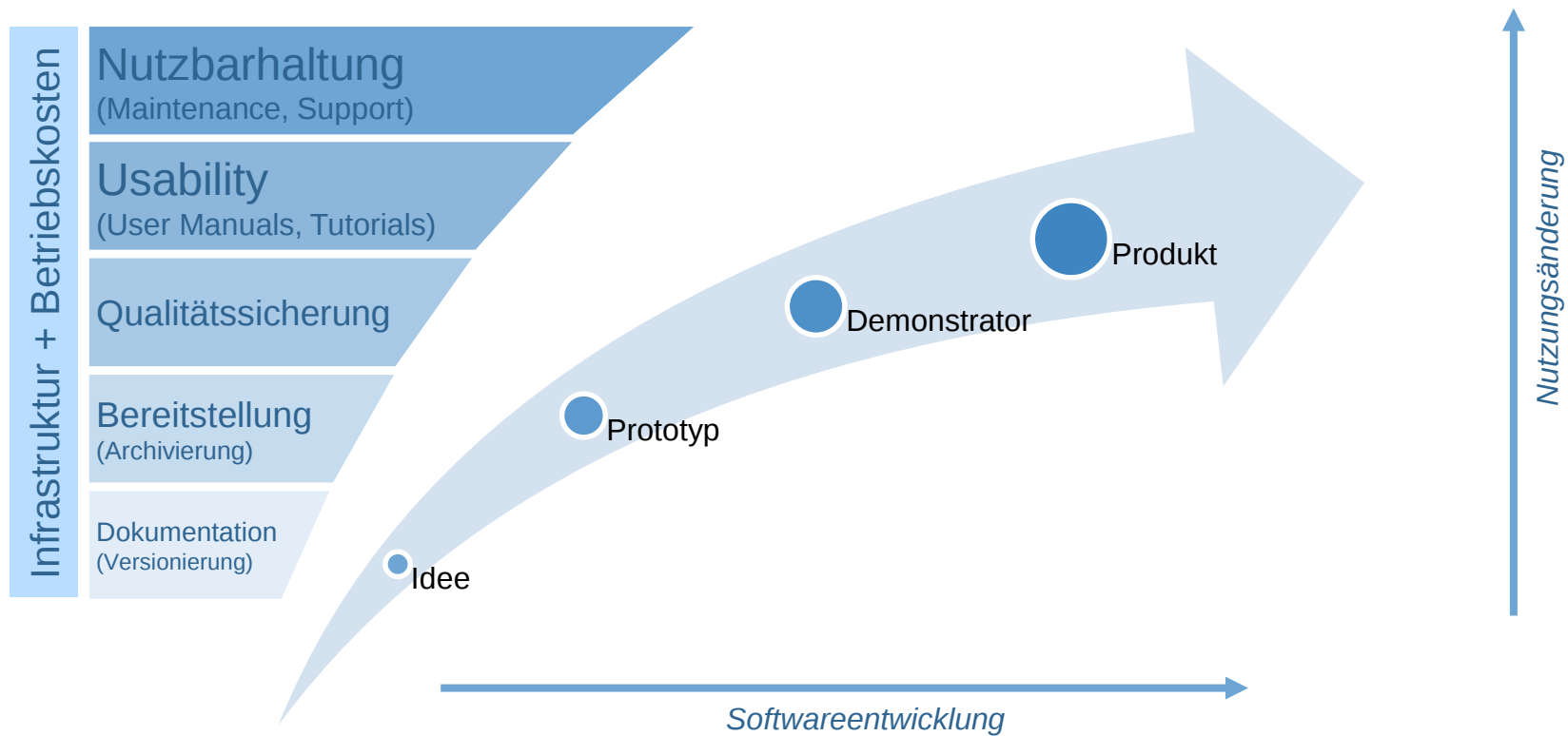
Bereitstellung

Software Citation Principles (by FORCE11, 2016)

- ▶ **Importance:** Software should be considered a legitimate and citable product of research
- ▶ **Credit and Attribution:** Software citations should facilitate giving scholarly credit and normative and legal attribution to all contributors
- ▶ **Unique Identification:** A software citation should include a method for identification that is machine actionable, globally unique, interoperable, and recognized
- ▶ **Persistence:** Unique identifiers and metadata describing the software and its disposition should persist
- ▶ **Accessibility:** Software citations should facilitate access to the software itself and to its associated metadata, documentation, data, and other materials necessary for both humans and machines
- ▶ **Specificity:** Software citations should facilitate identification of, and access to, the specific version of software that was used. Software identification should be as specific as necessary, such as using version numbers, revision numbers, or variants such as platforms

Smith AM, Katz DS, Niemeyer KE, FORCE11 Software Citation Working Group. (2016) Software Citation Principles. PeerJ Preprints 4:e2169v2 <https://doi.org/10.7287/peerj.preprints.2169v2>.

Forschungssoftware als Infrastruktur



Inhalt

1. Einführung
2. Forschungssoftware
3. Nachhaltigkeit
4. **Förderangebote der DFG**
5. Hinweise zu Workshops



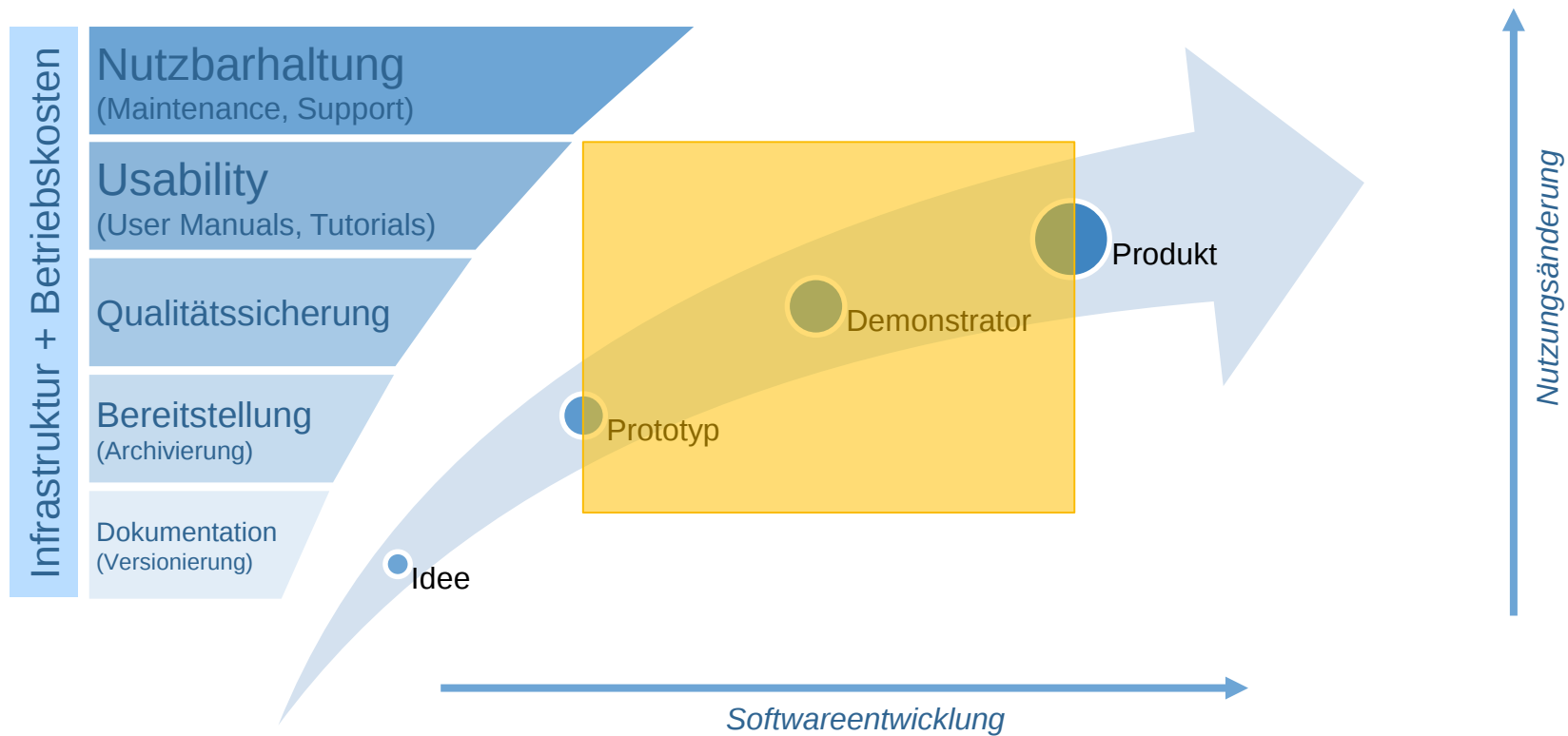
e-Research-Technologien (DFG-Vordruck 12.19)

Förderprogramm im Bereich LIS

- ▶ Anträge zum Auf- und Ausbau überregionaler, digitaler Informationsinfrastrukturen
- ▶ Zur Unterstützung der auf digitale Informationen und Daten angewiesenen Forschung durch webbasierte Techniken und Verfahren
- ▶ Entwicklung und Ausgestaltung von
 - Technologien, Werkzeugen, Verfahren oder Anwendungen für die Beschaffung, für die Zugänglich- und Nutzbarmachung, für die Bearbeitung und Auswertung sowie für die Sicherung von wissenschaftlich relevanten Informationen
 - den Einsatz von e-Research-Technologien nötigen Organisationsformen und von Modellen, mit denen der langfristige Betrieb von Informationsinfrastrukturen gesichert wird

Nachhaltigkeit von Forschungssoftware

Ausschreibung ca. November 2016



Inhalt

1. Einführung
2. Forschungssoftware
3. Nachhaltigkeit
4. Förderangebote der DFG
- 5. Hinweise zu Workshops**



Aktuelle Aktivitäten

- ▶ 2011 Science Code Manifesto (<http://sciencecodemanifesto.org/>)
- ▶ 2015 Research Software Sustainability, Knowledge Exchange
- ▶ 2015 DFG Expertengespräch
- ▶ 2016 Engineering Academic Software at Schloss Dagstuhl
- ▶ 2016 Software Citation Principles (FORCE11 Software Citation Working Group)
- ▶ 2016 WSSSPE4 workshop (12.-14. September)
- ▶ 2016 Helmholtz “Zugang zu und Nachnutzung von wiss. Software” (22./23. November)
- ▶ 2016 Allianz Initiative: Ad-hoc AG Wissenschaftliche Software

Bedarfe in den Geisteswissenschaften?

Sammlung von Diskussionspunkte

- ▶ Was würde in den Geisteswissenschaften geschehen, wenn Forschungssoftware nicht länger verfügbar wäre?
- ▶ Welche Forschungssoftware gibt es? Wie wird sie „betrieben“?
- ▶ Welche Forschungssoftware ist unerlässlich?
- ▶ Wie können Community-Angebote langfristig gewährleistet werden?
- ▶ Gibt es Reproduktionsstudien in den Geisteswissenschaften?



Vielen Dank für Ihre Aufmerksamkeit

Weitere Informationen

- ▶ zur DFG: <http://www.dfg.de>
- ▶ zum Förderatlas: <http://www.dfg.de/foerderatlas>
- ▶ zu allen geförderten Projekten: <http://www.dfg.de/gepris>
- ▶ zu den deutschen Forschungseinrichtungen: <http://research-explorer.de>