

Softwarezitation als Technik der Wissenschaftskultur

Vom Umgang mit Forschungssoftware in den Digital Humanities

Henny-Krahmer, Ulrike

ulrike.henny-krahmer@uni-rostock.de
Universität Rostock, Germany

Jettka, Daniel

daniel.jettka@uni-paderborn.de
Universität Paderborn

Einleitung

Software spielt in der gegenwärtigen geisteswissenschaftlichen Forschung eine zentrale Rolle bei der Gewinnung, Anreicherung, Auswertung und Veröffentlichung von digitalen Daten und hat damit einen wesentlichen Anteil an der Schaffung eines digitalen Gedächtnisses. Doch wie steht es um die Erinnerung an die Software selbst im digitalen Kontext?

Mit diesem Beitrag wird anhand einer Analyse der DHd-Konferenzabstracts der Jahrgänge 2016 bis 2020 untersucht, wie Forschungssoftware in den Digital Humanities zitiert wird. Zunächst wird erläutert, welche Relevanz Softwarezitation für die Anerkennung und Nachhaltigkeit von Forschungssoftware in den Digital Humanities hat. Im Anschluss werden bestehende Empfehlungen für Softwarezitation ausgewertet, um Bestandteile von Zitationen zu identifizieren, zu denen Informationen erhoben werden können. Ausgehend von den DHd-Abstracts wird eine Liste erwähnter Software generiert. Für eine Auswahl der so gefundenen Software wird geprüft, auf welche Weise sie zitiert wird und welche Informationen mit den Zitaten gegeben oder weggelassen werden.

Das Ziel der Analyse ist eine Bestandsaufnahme der Praxis der Softwarezitation in den Digital Humanities, in Anlehnung an eine Studie von Howison/Bullard (2016) zu Softwarezitationen in biologischen Forschungspublikationen. Damit soll das Thema der Zitation von Forschungssoftware in den Digital Humanities in den Fokus gerückt werden, um eine Verbesserung der bestehenden Praxis zu fördern.¹

Relevanz von Softwarezitation

Das Thema Softwarezitation wird im wissenschaftlichen Bereich bisher vor allem aus der Perspektive des Research Software Engineerings (RSE) diskutiert. Es existieren bereits entsprechende Empfehlungen (u. a. Jackson o. D., Smith et al. 2016, Chue Hong et al. 2019a, 2019b, Druskat 2021a, 2021b).

Forschungssoftware zu zitieren hat wichtige Funktionen. Wie bei der Zitation anderer Forschungsergebnisse auch geht es darum, Anerkennung für die wissenschaftliche Leistung auszudrücken, verwendete Quellen offenzulegen und auf sie zu verwei-

sen. In den Digital Humanities gibt es immer wieder eine Diskussion darüber, ob die mit den DH verbundenen Tätigkeiten als Forschung oder als Dienstleistung für Forschung anzusehen sind (Eckhart 2020). Beide Ausprägungen sind möglich,² wodurch es umso wichtiger wird, die Entwicklung von Forschungssoftware in den Digital Humanities als wissenschaftlichen Beitrag zu kennzeichnen und entsprechend zu zitieren.

Dafür bedarf es allerdings einer Definition von Forschungssoftware. Wir verstehen hierunter Software, die für Forschungsfragen, -gegenstände, -daten und -methoden und damit für die Forschungsergebnisse wesentliche Funktionalitäten bereitstellt. Dabei ist unerheblich, ob die Software für Forschungszwecke entwickelt wurde, für diese eingesetzt wird oder z.B. selbst Gegenstand der Forschung ist. Ein Textverarbeitungsprogramm oder ein generischer XML-Editor z. B. fallen in der Regel nicht unter diese Definition, eine für bestimmte Forschungsdaten konfigurierte Datenbank hingegen schon.³

Neben der Anerkennung als wissenschaftliche Leistung sollte Forschungssoftware generell auffindbar sein. Selbst wenn nicht alle Software langfristig lauffähig bleibt, sollten Code oder Beschreibungen verfügbar sein, auf die mit einer Zitation verwiesen werden kann (Smith et al. 2016). So wie Forschungsdaten sollte Software unter Beachtung der FAIR-Prinzipien publiziert werden (Lamprecht et al. 2020). Erst dadurch werden sinnvolle Softwarezitationen möglich und mit Software erzielte Forschungsergebnisse transparenter und nachvollziehbarer.

Empfehlungen und Kriterien für Softwarezitation

In vorhandenen Empfehlungen für Softwarezitation werden Vorschläge gemacht, welche Bestandteile Erwähnungen von Software in wissenschaftlichen Texten, insbesondere in Form von bibliographischen Angaben, haben sollten.

Smith et al. (2016) nennen sechs Prinzipien für die Zitation von Forschungssoftware: *Importance*, *Credit and attribution*, *Unique identification*, *Persistence*, *Accessibility* und *Specificity*. *Importance* meint, dass Software genau wie andere wissenschaftliche Ergebnisse auch in den Metadaten (also der Bibliographie) eines Beitrags aufgeführt werden sollte, wenn sie zitiert wird. Anerkennung (*Credit and attribution*) sollte denjenigen zukommen, die tatsächlich zur Entwicklung der Software beigetragen haben. Dies kann bedeuten, dass man die Entwickler:innen als Autor:innen der Software nennt und z. B. nicht Autor:innen einer Publikation über eine Software, da sich beide Gruppen nicht zwingend entsprechen. Software sollte darüber hinaus einen Identifikator haben, der global eindeutig, interoperabel und sowohl menschen- als auch maschinenlesbar ist. Dieser Identifikator sollte genau wie Metadaten zur Software persistent sein. Zitationen sollten den Zugang zur Software ermöglichen, indem sie auf die Software selbst (binär oder als Code), auf Metadaten zur Software oder auf Dokumentationen der Software verweisen. *Specificity* meint, dass Zitationen die Identifikation und den Zugang zu bestimmten Versionen der Software erlauben sollten (Smith et al. 2016).

Abgesehen von den o. g. Prinzipien und Empfehlungen stammen Vorschläge für Softwarezitationen vor allem aus allgemeinen Zitierstilen (u. a. MLA oder APA) oder von Entwickler:innen selbst. Bestandteile dieser Empfehlungen lassen sich in der Regel auf die oben genannten Prinzipien zurückführen.

Auf der Grundlage der genannten Empfehlungen für Softwarezitationen formulieren wir die nachfolgenden Kriterien um die Erwähnung von Forschungssoftware in den DHd-Abstracts zu cha-

rakterisieren. Die Kriterien sind als TEI-Taxonomie modelliert und verfügbar in Henny-Krahmer/Jettka (2021) sowie auf GitHub⁴.

- **Bibliographieeintrag für Software (Bib.Soft):** Die Bibliographie enthält einen Eintrag für die Software selbst. Dieser kann den Namen der Software selbst enthalten, Namen von Verantwortlichen, eine URL, einen PID, Versionsangaben, usw.
- **Bibliographieeintrag für Referenzpublikation (Bib.Ref):** Die Bibliographie enthält einen Eintrag mit einer Publikation über die Software.
- **Nur namentliche Nennung der Software (Name.Only):** Die Software ist nur namentlich genannt.
- **Namentliche Nennung der Verantwortlichen (Agent):** Personen, Gruppen oder Institutionen, die für die Entwicklung der Software verantwortlich sind, werden namentlich genannt.
- **URL:** Die Zitation enthält eine URL, die auf die Software selbst verweist (z. B. zu einer Webseite über die Software, einem Code-Repository, einem Metadatensatz oder einer ausführbaren Version).
- **Persistenter Identifikator (PID):** Die Zitation enthält einen persistenten Identifikator (PID), z. B. eine DOI, der auf die Software selbst verweist (z. B. zu einer Webseite über die Software, einem Code-Repository, einem Metadatensatz oder einer ausführbaren Version).
- **Versio(n) (Ver):** Die Zitation enthält die Angabe einer bestimmten Softwareversion oder -revision und ggf. anderweitig notwendige Spezifikationen (z. B. eine Version für ein spezifisches Betriebssystem, ein bestimmtes Softwarepaket oder ein Datum).

Bei Empfehlungen wird z. T. zwischen der Perspektive von Software-Anbieter:innen, die Zitiervorschläge machen, und der Perspektive von Nutzer:innen, die Software zitieren, unterschieden. Der Fokus liegt hier auf der Perspektive der Verfasser:innen von wissenschaftlichen Publikationen, in denen Software zitiert wird. Bei der Analyse solcher Zitationen ist zu beachten, dass spezifische Zitiervorschläge von den Anbietenden einen Einfluss darauf haben können, wie die entsprechende Software zitiert wird. Dass solche Vorschläge gemacht werden, ist wichtig und trägt wesentlich dazu bei, dass alle essentiellen Informationen über eine Software verfügbar sind. Insofern stellt eine Analyse von Zitiervorschlägen durch Entwickler:innen eine sinnvolle Folgeuntersuchung dar.

Daten und Methoden

Die Datengrundlage für die Analyse zur Praxis der Zitation von Forschungssoftware in den Digital Humanities bilden die Bände der DHd-Konferenzabstracts aus den Jahren 2016 bis 2020, die vom DHd-Verband auf GitHub in PDF- und TEI-Format zur Verfügung gestellt werden.⁵ Die Jahrgänge 2014 und 2015 wurden von der Untersuchung ausgeschlossen, da die Abstracts für diese Jahre nur im PDF-Format verfügbar sind und mit den anderen Jahrgängen bereits eine breite Datenbasis bestehend aus insgesamt 686 Abstracts für Panels, Workshops, Poster und Vorträge mit insgesamt ca. 55.000 Sätzen und 1,2 Mio. Tokens verfügbar ist.⁶

Erschließung von Softwareentitäten

Für die Erhebung der Softwarezitation in den Konferenzabstracts wurde zunächst eine Liste von in den Digital Humanities häufig verwendeter Software erstellt. Diese wurde zum Auffinden konkreter Nennungen in den DHd-Abstracts genutzt. Neben der Auflistung uns bereits bekannter einschlägiger Software⁷ wurden auch automatische Methoden evaluiert, um weitere Benennungen von Software aus den DHd-Abstracts zu erschließen, ohne dass dieser Beitrag damit einen Schwerpunkt auf die Entwicklung automatisierter Verfahren zur Erkennung von Softwareentitäten legt. Vielmehr ging es darum, praktikable Ansätze zur Gewinnung einer Datenbasis zu entwickeln, mit der Softwarezitationen untersucht werden können.

Da existierende Ansätze zu Software Entity Recognition aus der Bioinformatik (Duck et al. 2015) und Biomedizin (Wei et al. 2020) nicht ohne Weiteres auf die Domäne Digital Humanities anwendbar sind, wurde ein Ansatz evaluiert, der allgemeine Named Entity Recognition (NER)⁸ nutzt, um Kandidaten von Software-Benennungen zu ermitteln und im Nachgang auszuwerten. Hierzu wurden die Abstracts mit Hilfe von WebLicht (CLARIN-D/SfS-Uni. Tübingen 2012; Hinrichs et al. 2010) bzw. WebLicht as a Service⁹, automatisch mit Named Entities annotiert. Da der NER-Service¹⁰ nicht auf die Erkennung von Software als Named Entity trainiert ist, bietet sein Einsatz zwar eine Möglichkeit der Annäherung an weitere Kandidaten, allerdings nur in sehr eingeschränktem Maß. Durch manuelle Nachbearbeitung der Liste aller ermittelten Named Entities (insgesamt 29.028), bei der nur Einträge mit mindestens 10 Vorkommen in den DHd-Abstracts betrachtet wurden (910 Named Entities), wurden lediglich 10 Namen von Software ermittelt.

Obwohl der Einsatz allgemeiner NER für die Ermittlung von in den DHd-Abstracts genannter Forschungssoftware keinen großen Mehrwert bieten konnte, steht durch die Kombination der händisch erstellten Auswahl mit den automatisch erzielten Ergebnissen schließlich eine Liste von 138 Softwarenamen zur Verfügung, die für die Ermittlung von Nennungen und Zitationen in den DHd-Abstracts genutzt werden kann.

Erfassung von Softwarezitationen

Auf Basis der ermittelten Softwarenamen können nun Zitationen (und Nicht-Zitationen) von Forschungssoftware in den DHd-Konferenzabstracts erschlossen und klassifiziert werden. Im Weiteren wird der Begriff Zitation auch für reine Namensnennungen (also im engeren Sinn Nicht-Zitationen) verwendet, da diese ebenfalls ausgewertet werden sollen. Die Softwareliste wurde zunächst anhand der Anzahl vorkommender Instanzen der Software in den Abstracts sortiert, und somit die Analyse häufig genannter Software höher priorisiert, da einerseits möglichst viele Varianten von Zitationen abgedeckt werden sollten und andererseits aufgrund des großen Aufwands nur eine Auswahl von Software erfasst werden konnte.

Mit Hilfe der o. g. TEI-Taxonomie wurden 995 Vorkommen von 32 Softwarenamen in Kombination mit evtl. vorhandenen Zitationen manuell in den TEI-Dateien der DHd-Abstracts annotiert.¹¹ Die verwendete Softwareliste, die TEI-Taxonomie, die annotierten TEI-Dokumente und die Daten zur Auswertung sind verfügbar in Henny-Krahmer/Jettka (2021)¹².

Ergebnisse und Diskussion

Um einen Eindruck von der Zitationspraxis für Forschungssoftware in den DHD-Konferenzabstracts der Jahrgänge 2016 bis 2020 zu erlangen, wurden die manuellen Annotationen der TEI-Dokumente ausgewertet. Da Software häufig mehrfach in einem Abstract genannt, aber sinnvollerweise nicht bei jeder Nennung vollständig zitiert wird, wird das Vorhandensein von Zitationsbestandteilen für jede Software einmal pro Beitrag gezählt und nicht pro Nennung (wenn z. B. einmal im Beitrag eine URL genannt wird, zählt dieses Kriterium als erfüllt). Hierbei ist zu beachten, dass die Zitationsarten (bis auf Name.Only) nicht exklusiv sind, und nicht selten mehrere Zitationsarten mit einer Software verbunden sind. So kann eine Software in einem Beitrag beispielsweise sowohl mit Bib.Soft als auch Bib.Ref zitiert worden sein. Die Verteilung (vgl. Abbildung 1) basiert auf einer Gesamtzahl von $n=218$, welche die einfach gezählten Nennungen einer bestimmten Software in einem bestimmten Beitrag repräsentiert. Die einzelnen absoluten Werte zeigen die Häufigkeit einer Zitationsart einer bestimmten Software, jeweils einfach gezählt pro Abstract, an.

Zitationstyp	Abs. Häufigkeit (n = 218)	Rel. Häufigkeit (%)
Bib.Soft	46	21
Bib.Ref	84	39
Name.Only	46	21
Agent	45	21
URL	114	52
PID	1	0,5
Ver	16	7

Abb. 1: Häufigkeitsverteilung von Softwarezitationen

Die Verteilung der verschiedenen Zitationsarten und -bestandteile zeigt, dass in ca. der Hälfte der erfassten Fälle eine URL für eine Software angegeben wurde, während in nur einem von 218 Fällen ein persistenter Identifikator bereitgestellt wurde. Wenn eine Softwarezitation über einen bibliographischen Eintrag erfolgt, dann zumeist über eine Referenzpublikation (in 39% der Fälle), seltener über einen Bibliographieeintrag für die Software selbst (21%). Ebenfalls für ca. ein Fünftel der betrachteten Softwarenennungen erfolgte eine Nennung von verantwortlichen Institutionen oder Personen, allerdings in ähnlichem Maß auch gar keine Zitation. In 7% der Fälle wurde eine Version der Software genannt.

Die Ergebnisse der Erhebung zeigen, dass Forschungssoftware einerseits zwar in der Regel in Verbindung mit einer der Zitationsarten genannt wird, andererseits jedoch relativ selten eine direkte, langfristige Zitation über bibliographische Einträge für die Software selbst oder unter Verwendung eines persistenten Identifikators erfolgt. Zu prüfen wäre nun, ob sich das Bewusstsein für die Notwendigkeit der nachhaltigen Zitation von Forschungssoftware seit dem Jahr 2016 gewandelt haben könnte und die Berücksichtigung des gesamten Fünf-Jahres-Zeitraums möglicherweise kein adäquates Bild des aktuellen Stands zeichnet. Betrachtet man

allerdings die Verteilung der relativen Häufigkeiten der Zitationsarten pro Jahr, ließe sich sogar eine abnehmende Tendenz der Häufigkeit direkter Softwarezitationen in Bibliographien vermuten (vgl. Abbildung 2). Signifikante Aussagen bedürfen jedoch einer größer angelegten Studie.

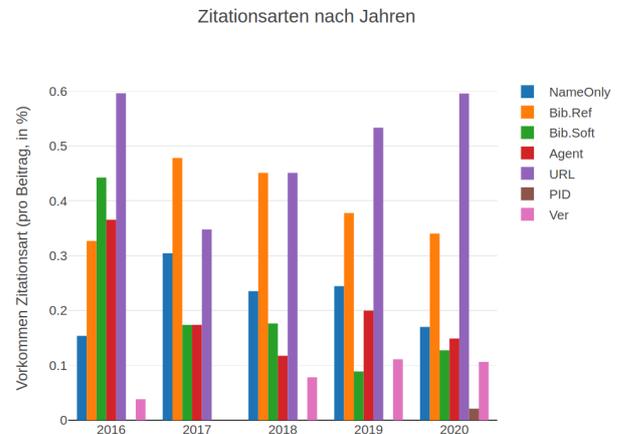


Abb. 2: Relative Häufigkeiten von Zitationsarten pro Jahrgang

Fazit

Ausgehend von der Darstellung der Rolle von Forschungssoftware für den wissenschaftlichen Erkenntnisprozess und den Aufbau eines digitalen Gedächtnisses in den Digital Humanities wurden im vorliegenden Beitrag Empfehlungen für Softwarezitation vorgestellt, die darauf abzielen, Software als wissenschaftliches Werkzeug und Ergebnis adäquat zu identifizieren, anzuerkennen und nachzuhalten.

Um sich ein erstes Bild vom aktuellen Status von Forschungssoftware in den Digital Humanities zu machen, wurden Kriterien für ihre Erwähnung in den DHD-Abstracts der Jahrgänge 2016 bis 2020 formuliert und in Form einer TEI-Taxonomie repräsentiert. Anhand der Kriterien sowie einer Liste von Software, die in den Digital Humanities verwendet wird, erfolgte eine manuelle Annotation ausgewählter Software und ihrer Zitationen in den DHD-Abstracts.

Die vorliegende Bestandsaufnahme der Praxis der Softwarezitation in den Digital Humanities weist deutlich auf Verbesserungsbedarf hin, sowohl im Hinblick auf die Verwendung bibliographischer Einträge für Forschungssoftware selbst und den Einsatz von persistenten Identifikatoren als auch mit Blick auf die Nennung von verantwortlichen Personen und Institutionen, deren Leistungen entsprechend anerkannt werden sollten. Sowohl auf Seiten der Zitationspraxis, also bei Anwender:innen, als auch auf Seiten der Zitationsempfehlungen von Softwareprojekten, also bei Entwickler:innen und Betreiber:innen, besteht in diesem Zusammenhang noch Handlungsbedarf.

Fußnoten

1. Der Hintergrund dieser Einreichung ist das Projekt NF-DI4Culture, in dessen Arbeitsbereich zu Forschungstools und Datendiensten ("Research Tools and Data Services") Überlegungen, Empfehlungen, Beratung und Angebote für die nachhaltige

Entwicklung von Forschungssoftware im Kulturerbe fallen (s. a. <https://nfdi4culture.de/de/aufgaben/aufgabenbereiche/aufgabenbereich-3.html>).

2. Siehe z. B. die Unterscheidung zwischen Forschung und Dienstleistung bei den DH an der Uni Bern: <https://www.dh.unibe.ch>, das Service Center Digital Humanities in Münster (<https://www.uni-muenster.de/EScience/schwerpunkte/dh.html>) oder den Lehrstuhl für Digital Humanities in Trier (<https://www.uni-trier.de/universitaet/fachbereiche-faecher/fachbereich-ii/faecher/computerlinguistik-und-digital-humanities/digital-humanities>).

3. Für andere Definitionsvorschläge siehe Hettrick et al. 2014 und Homburg et al. 2020.

4. <https://github.com/daniel-jettka/software-citation-dhd/blob/main/conf/citation-taxonomy.xml>.

5. <https://github.com/DHd-Verband>.

6. Die Satz- und Tokenanzahl wurde ermittelt mit Hilfe des BBAW Tokenizer and Sentence Splitters (WebLicht Service Handle PID: <https://hdl.handle.net/21.11120/0000-0008-3183-C>), der in WebLicht (CLARIN-D/SfS-Uni. Tübingen, 2012) als Webservice zur Verfügung steht.

7. Für eine erste Annäherung wurden neben einer selbst erstellten Liste und Ergebnissen aus einer ersten Sichtung der DHd-Abstracts folgende Quellen herangezogen: die Webseite des Projekts fürTEXT (<https://fortext.net/>), in der Zeitschrift RIDE rezensierte Tools (<https://ride.i-d-e.de/issues/issue-11/>) sowie Software, die im Anhang es Projektantrags von NFDI4Culture genannt ist (öffentliche Fassung: <https://riojournal.com/article/57036/>).

8. Sticker Named Entity Recognizer, <https://github.com/stickeritis/sticker/>.

9. Die Definition der Webservice-Chain und ein Beispielskript für deren Aufruf sind verfügbar in Henny-Krahmer/Jettka (2021).

10. WebLicht Service Handle PID: <http://hdl.handle.net/11022/0000-0007-DA29-6>.

11. Direkte Anwendung finden alle Kategorien aus der TEI-Taxonomie außer Name.Only, welches sich implizit aus dem Fehlen der anderen Kategorien ergibt.

12. Verschiedene Entwicklungsstände der Datensätze sind ebenfalls abrufbar über <https://github.com/daniel-jettka/software-citation-dhd>.

Bibliographie

CLARIN-D/SfS-Uni. Tübingen (2012): "WebLicht: Web-Based Linguistic Chaining Tool" Online. <https://weblicht.sfs.uni-tuebingen.de/> [Letzter Zugriff: 13.07.2021].

Chue Hong, Neil (ed.) (2019a): "Software Citation Checklist for Authors" (Version 0.9.0). *Zenodo*. <http://doi.org/10.5281/zenodo.3479199>.

Chue Hong, Neil (ed.) (2019b): "Software Citation Checklist for Developers" (Version 0.9.0). *Zenodo*. <http://doi.org/10.5281/zenodo.3482769>.

Druskat, Stephan (2021a): "Research software citation for researchers" *Research Software Citation. Cite and Make Citable!* (Version 1.1). <https://cite.research-software.org/researchers/> [Letzter Zugriff: 13.07.2021].

Druskat, Stephan (2021b): "Research software citation for developers" *Research Software Citation. Cite and Make Citable!* (Version 1.1). <https://cite.research-software.org/developers/> [Letzter Zugriff: 13.07.2021].

Duck, Geraint / Kovacevic, Aleksandar / Robertson, David L. / Stevens, Robert / Nenadic, Goran (2015). "Ambiguity and variability of database and software names in bioinformatics", in: *Journal of Biomedical Semantics* 6 (29). <https://doi.org/10.1186/s13326-015-0026-0>.

Eckhart, Arnold (2020): "Digital Humanities: Is it Research or is it Service?" *dhmuc. Digital Humanities München*. Blog post. 26.7.2020. <https://dhmuc.hypotheses.org/2834> [Letzter Zugriff: 13.07.2021].

Henny-Krahmer, Ulrike / Jettka, Daniel (2021): "Software-citation in den Digital Humanities" (Version 0.1). *Zenodo*. <http://doi.org/10.5281/zenodo.5106391>.

Hettrick, Simon / Antonioletti, Mario / Carr, Les / Chue Hong, Neil / Crouch, Stephen / De Roure, David / Emsley, Iain et al. (2014, dec): "UK Research Software Survey 2014" *Zenodo*. <https://doi.org/10.5281/zenodo.14809>.

Hinrichs, Erhard W. / Hinrichs, Marie / Zastrow, Thomas (2010): "WebLicht: Web-Based LRT Services for German", in: *Proceedings of the ACL 2010 System Demonstrations* 25–29.

Homburg, Timo / Klammt, Anne / Mara, Hubert / Schmid, Clemens / Schmidt, Sophie C. / Thiery, Florian / Trognitz, Martina (2020): "Diskussionsbeitrag - Handreichung zur Rezension von Forschungssoftware in den Altertumswissenschaften / Impulse - Recommendations for the review of archaeological research software" *GitHub*. https://research-squirrel-engineers.github.io/Impuls_SoftwareRezensionen_DGUF/Draft.htm.

Howison, James / Bullard, Julia (2016): "Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature", in: *Journal of the Association for Information Science and Technology (JASIST)* 67 (9): 2137–2155. <https://doi.org/10.1002/asi.23538>.

Jackson, Mike (o. D.): „How to cite and describe software.“ *Software and research: The Software Sustainability Institute's Blog*. Blog post . <https://www.software.ac.uk/how-cite-and-describe-software> [Letzter Zugriff: 13.07.2021].

Lamprecht, Anna-Lena / Garcia, Leyla / Kuzak, Mateusz / Martinez, Carlos / Arcila, Ricardo / Del Pico, Eva M. / Dominguez Del Angel, Victoria et al. (2020): "Towards FAIR principles for research software", in: *Data Science* 3 (1): 37–59. <https://doi.org/10.3233/DS-190026>.

Smith, Arfon M. / Katz, Daniel S. / Niemeyer, Kyle E. / FORCE11 Software Citation Working Group (2016): "Software citation principles", in: *PeerJ Computer Science* 2:e86. <http://dx.doi.org/10.7717/peerj-cs.86>.

Wei, Qiang / Zhang, Yaoyun / Amith, Muhammad / Lin, Rebecca / Laypeyrolerie, Jenay / Tao, Cui / Xu, Hua (2020): "Recognizing software names in biomedical literature using machine learning", in: *Health Informatics Journal* 26 (1): 21–33. <https://doi.org/10.1177/1460458219869490>.