

Kontext, Kontext, Kontext

Der methodologische Beitrag geistes- und sozialwissenschaftlicher Forschung zu iterativer Entwicklung von Forschungssoftware

Sabina Mollenhauer

Open Research im interdisziplinären Kontext

Im Zuge der zunehmenden Forderung nach offenen Daten stehen die Geistes- und Sozialwissenschaften (GSW) unter Druck, ihre Daten digital teilbar zu gestalten. Dies ist in anderen wissenschaftlichen Disziplinen bereits länger und in größerem Umfang Praxis (Longley und Hearn 2021: 829) und wird in Deutschland vor allem durch die Nationale Forschungsdateninfrastruktur (NFDI) vorangetrieben (RfII 2023). Es handelt sich bei diesen Forderungen um ein Paradigma der offenen Forschung (Vicente-Saez und Martinez-Fuentes 2018: 435) und bis heute dominiert der Positivismus naturwissenschaftlicher Fächer auch die derzeitigen technischen Entwicklungen rund um offene Forschung (Poljak und Posavec 2024: 4). Forschende der GSW arbeiten weitgehend mit interpretativen und kritischen Methoden, die stark auf Kontext und Bedeutungsanalyse fokussiert sind (Longley und Hearn 2021: 830). Die daraus resultierende Sichtweise auf Forschungsdaten und Datenpraktiken birgt wichtige Beiträge für die gesamte Entwicklung von Open Research. In diesem Kontext zielt mein Dissertationsprojekt darauf ab, Software zu entwickeln, welche die Aufbereitung für Publikation und mögliche Wiederverwendung qualitativer Interviewtranskript-Daten unterstützt, wie sie in den GSW entstehen. Wenn geisteswissenschaftliche Datenpraxis bezüglich qualitativer Daten über die Entwicklung von Forschungssoftware in den Gesamtkontext der eingangs erwähnten Forderungen nach offener Forschung einfließt, kann sie einen grundlegenden Beitrag zu diesem Paradigma leisten.

In meiner Forschung stelle ich Forschende der GSW nicht nur als Nutzende in den Fokus, indem ich deren Anforderungen an die Entwicklung einer Software und den darin eingeschriebenen Umgang mit digitalen Forschungsdaten in Kollaboration mit ihnen erarbeite. Meine Forschung basiert darüber hinaus auf theoretischen und methodologischen Grundlagen, die in diesen Disziplinen ent-

wickelt werden. In Forschungskontexten, die grundsätzlich mehr als digital sind, stellt Software etwas dar, dass inhärent digital ist und deswegen diese Kontexte nur begrenzt berücksichtigen kann. Meine interdisziplinäre Forschung schlägt eine Brücke zwischen den mehr-als-digitalen Kontexten der Forschenden und der Forschungssoftwareentwicklung selbst, indem ich die Entwicklung der Software auch durch Begrifflichkeiten und Prozesse des Software Engineering (SE) darstelle. Offene Forschung im Kontext der wissenschaftlichen Softwareentwicklung, dem Research Software Engineering (RSE), braucht GSW-Methodologie, um Software für Forschung gezielt und passgenau zu realisieren. Im Folgenden steht deshalb die theoretische und methodologische Einordnung dieser Perspektive zentral.

Research Software Engineering als Brücke zwischen Informatik und Geistes- und Sozialwissenschaften

Die Aufbereitung qualitativer Interviewtranskripte dient als naheliegender Anwendungsfall, da sie die am häufigsten produzierten Datentypen in den GSW darstellen (Imeri and Danciu [Mitarb.] 2017: 10, Strunk 2018: 21f.), aber in Bezug auf die FAIR-Datenpraktiken immer noch weitgehend unverfügbar sind (Ambrasat und Heger 2019: 33). Im Gegensatz zu vielen anderen Forschungsdaten sind Interviewdaten nicht nur besonders sensibel, sondern auch die Frage der Aufbereitung ist besonders relevant. Unstrukturierte Textdaten, die potenziell sensible Daten enthalten, spielen in verschiedenen Disziplinen eine große Rolle (Simukovic et al. 2013: 44), wobei sie vor allem in den GSW dominieren (Schöpfel und Prost 2016: 100). Gleichzeitig ist das Verhältnis zu Forschungsdaten und der Umgang mit diesen in vielen Feldern der GSW besonders kritisch (Knöchelmann 2019: 2).

Dabei leiten zwei zentrale Fragen mein Forschungsvorhaben: Erstens wird herausgearbeitet, was die Anforderungen jener GSW-Forschenden sind, die bereit sind, ihre qualitativen Interviewdaten zu teilen. Zweitens untersuche ich, welche dieser Anforderungen durch eine gezielt entwickelte Software unterstützt und erfüllt werden.

Damit ist das Projekt übergreifend dem RSE zuzuordnen. In dem relativ neuen Forschungsfeld werden seit den 2000er Jahren die besonderen Charakteristika von Softwareentwicklung für Forschungszwecke im akademischen Umfeld thematisiert (Li et al. 2008). Vor allem im Rahmen der NFDI stellt RSE ein Schnittstellenthema dar (Felderer et al. 2023: 1).

SE beinhaltet alle Teilbereiche, die im Softwareingenieurwesen zur Entwicklung von Software beitragen, wie beispielsweise Architektur, Design, Modellierung und Tests. Darüber hinaus wird es auch in die Felder spezifischer Anwendung unterteilt, wobei die Besonderheiten von Forschungssoftware lange unbeachtet blieben (ebd.: 2). RSE beschäftigt sich mit Software, die Forschung unterstützt und

der Sammlung, Verarbeitung, Analyse, Visualisierung, Modellierung oder Simulation von Forschungsdaten dient. Ähnlich wie für Software in anderen Feldern haben sich die Anforderungen an Forschungssoftware weiterentwickelt, nachdem diese lange individuellen Zwecken dienten: es dominierte Software, die vor allem nebenbei ohne längeren Prozess schnell geschrieben wurde und Einzellösungen für eine kleine Gruppe Forschender bot. Heute soll Forschungssoftware langlebig und deren Anwendung nach den FAIR-Prinzipien übertragbar sein: „[D]ocumentation, testing, software design techniques, and code management are key here. They help to ensure that software is reliable, easier to maintain, and that the results it outputs are reproducible“ (Cohen et al. 2020: 2).

GSW-Disziplinen sind in der Entwicklung von Forschungssoftware stark unterrepräsentiert (Philippe et al. 2019). Wie ich im Folgenden zeige, fließen ihre Konzepte und Methoden jedoch seit langem in spezifische Schritte der SE als dem RSE vorausgegangene Disziplin ein, insbesondere wenn es um die Ermittlung von Anforderungen an Software – das Requirement Engineering – geht. Da qualitative Forschungsmethoden bereits im SE in Gebrauch sind, sind sie auch für die zukünftige Entwicklung von RSE äußerst relevant.

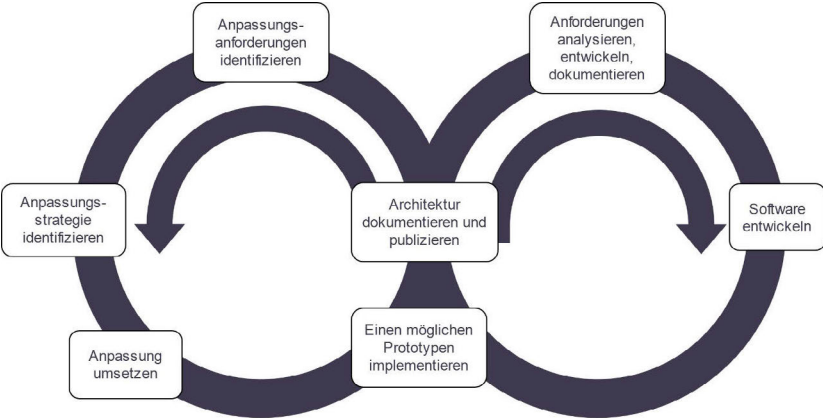
Interdisziplinarität in der iterativen Entwicklung von Forschungssoftware

Als interdisziplinäre Forscherin wechsele ich ständig die Sichtweise und den Kontext. Dabei stoße ich immer wieder auf Begriffe und Konzepte, die in den scheinbar gegensätzlichen Disziplinen der GSW und den Softwareingenieurwissenschaften identisch, ähnlich oder überschneidend sind. Konzepte wie Iteration und Kontextabhängigkeit scheinen hier vor allem relevant und dienen deshalb als Ausgangspunkt meines methodologischen Vorgehens. In der ersten Analyse zur Generierung von Anforderungen wurde klar, dass theoretische Grundkonzepte zum Tragen kommen, vor allem solche wie Multi-Sited Ethnography (Marcus 2000) und Contextual Inquiry (CI, Beyer und Holtzblatt 1995), die der ethnografischen Forschung zugeordnet werden (McGraw und Harbison 1997: 7f.; Kotonya und Sommerville 1998: 53). Im Folgenden wird daher zuerst mein übergeordneter, iterativer Prozess dargestellt, worauf eine Erläuterung der theoretischen und methodischen Grundlagen folgt. Diese dienen der Darstellung der Durchführung im nächsten Abschnitt, wo das methodische Vorgehen ausgeführt wird.

Die der SE zugrundeliegenden Schritte von Problemanalyse hin zur Pflege von Software werden in der Regel iterativ wiederholt (Royce 1987: 330). Seit etwa 30 Jahren werden Ansätze immer feldspezifischer und nutzer:innenorientierter (Raccoon 1997: 91). Qualitative, nutzer:innenzentrierte Methoden sind in Problem- und Anforderungsanalyse, im Testen und in der Anforderungsanpassung relevant (McGraw und Harbison 1997: 29).

Wie in Abbildung 1 dargestellt, erlaubt der iterative Prozess einen Entwicklungsansatz, der flexibel auf Kontexte reagiert. Ähnlich stellen es Siadat und Song (2012) in ihrem zyklischen Entwicklungsprozess dar, in welchem die Anpassung in Reaktion auf Kontexte von Software integriert ist.

Abb. 1: Kontext-abhängiger, iterativer Software-Lebenszyklus in Anlehnung an Siadat und Song (2012: 538)



Wie der Grafik zu entnehmen ist, spielen Anforderungen und deren Anpassung eine wiederkehrende Rolle. Beginnend mit der Anforderungsentwicklung durch explorative, qualitative Methoden („Anforderungen analysieren, entwickeln, dokumentieren“) führt der Weg über die technischen Entwicklungsschritte hin zu einem ersten prototypischen Werkzeug, das dazu eingesetzt werden kann, ethnografisch Anforderungsanpassungen zu identifizieren. Diese Anforderungen können zuletzt umgesetzt werden. Der Prozess ist jedoch, beliebig wiederholbar, so dass die Softwarelösung einem sich wandelnden Kontext angepasst werden kann.

RSE ist ein Feld, in dem Anforderungsanalysen noch wenig Anwendung finden, unter Anderem, da der Prozess als forschungsbehindernd oder nicht auf Forschungssoftware zutreffend wahrgenommen wird, weil diese – im Gegensatz zu SE – keine Kund:innen hat, die beforscht werden müssen, wenn deren Nachfrage als Nutzende auch die Existenz der Software beeinflusst (Bajraktari et al. 2024: 1). Einige grundlegende Ideen und Perspektiven werden aber innerhalb der RSE entwickelt. Bajraktari et al. etwa weisen auf Reproduzierbarkeit als Qualitätskriterium hin und schlagen Softwareentwicklungsprozesse vor, die explorative Arbeit und organisches Wachstum ermöglichen, welches zu Anforderungen führt, denen Forschungskonzepte zugrunde liegen (ebd.: 8). Auf diesen RSE-Kontext beziehe ich mich sowohl in meiner Perspektive der ethnografischen Feldforschung als auch in

der Anforderungsanalyse und Softwareentwicklung. Denn meine Arbeit soll diesen interdisziplinären Zugang nicht nur in der initialen Anforderungsanalyse, sondern auch in weiteren Schritten des Softwareentwicklungsprozesses abdecken, da sie so weit wie möglich dem Feld und den Nutzenden nahe und kollaborativ angelegt ist.

Konzepte und Methoden der Geistes- und Sozialwissenschaften im Software Engineering

Softwareentwicklungsprozesse mit Fokus auf Nutzende werden im SE auch als „scenario-based engineering process“ (SEP, McGraw und Harbison 1997) betrachtet. Das Szenario der digitalen Aufbereitung von Interviewtranskripten in der qualitativen Forschung wurde, wie eingangs dargestellt, aufgrund seiner Relevanz gewählt. Aus Sicht der GSW trägt die Auswahl dieses Szenarios zur Bildung eines temporären Gebildes bei, einer Verflechtung von menschlichen und nicht-menschlichen Akteur:innen und deren Einflüsse aufeinander: eine sozio-technische Assemblage (Vepřek 2024: 15). Von der Einordnung des Forschungsgegenstandes über die Erhebungsmethoden hin zur Analyse, die zur Generierung von Anforderungen und deren Anpassung führt, ist SE von Konzepten aus den GSW durchdrungen, auch wenn sie mit unterschiedlichen Begriffen bezeichnet werden.

In der klassischen Problem- und Anforderungsanalyse des SE kann auf Grundlage von dokumentierten Informationen über bestehenden Systemen, erhobenen Bedürfnissen der Interessengruppen, feststehenden Organisationsstandards, Vorschriften und Domäneninformationen (Kotonya und Sommerville 1998; Shams-ul-Arif et al. 2009: 42) ein detaillierter Prozess umgesetzt werden, aus dem kollaborativ entwickelte und vereinbarte Anforderungen der Nutzenden, Systemspezifikationen und Systemmodelle hervorgehen. Shams-ul-Arif et al. stellen diesen Prozess als im Idealfall iterativ dar. Sie stellen die Erhebungsmethoden wie Interviewformen, Umfragemethoden und Beobachtungen der Bearbeitung gestellter Aufgaben zwar nicht detailliert dar, werten diese jedoch in einem Überblick aus und zeigen eindrücklich, welchen Stellenwert qualitative Forschungsmethoden im SE besitzen.

Diese immanente Anwendung qualitativer Forschungsmethoden wird auch in den grundlegenden Lehrwerken für Softwareingenieur:innen (Kotonya und Sommerville 1998: Kapitel 3; McGraw und Harbison 1997: 150–325) dargestellt. Interessant ist jedoch, dass die Liste von Shams-ul-Arif et al. von Erhebungsmethoden zu Analysemethoden, wie beispielsweise der Introspektion oder der Anforderungsmatrize, übergeht und im Zentrum die von ihnen als relevanteste und komplexeste vorgestellte Methode der Domain Analysis steht. Im SE führte James Neighbors 1980 die Domain Analysis als Möglichkeit ein, ein Feld (z. B. ein Reisebüro) in seiner Komplexität darzustellen, ohne es vorher in ein System zu zwingen, dabei jedoch für die Beschreibung dieses Feldes eine gewisse technische Sprache zu nutzen (Neighbors 1980: 5f.). Neighbors erläutert die Domain Analysis

als eine Tätigkeit, die dazu dient, die Objekte und Vorgänge, die einer Art von ähnlichen Systemen angehören, innerhalb eines bestimmten Problembereichs zu beschreiben (ebd.: 1). Obwohl sie den Begriff nicht direkt definieren, beschreiben McGraw und Harbison die Domain als etwas, das gemeinsame Konzepte, Organisation und einen spezifischen Wortschatz besitzt (McGraw und Harbison 1997: 70). Eine Domain Analysis beinhaltet also die genaue Untersuchung von und ein tiefes Verständnis für die spezielle Umgebung oder den Bereich (die Domäne), in dem die Software verwendet wird bzw. werden soll. Ziel ist es, die relevanten Prozesse, Regeln und Anforderungen der Domäne zu erfassen, damit – bei einem Ansatz, der Nutzer:innen in den Mittelpunkt stellt – die Software optimal an die Bedürfnisse der Nutzenden angepasst werden kann (ebd.). So dient die Domain Analysis dazu, die Besonderheiten eines Fachgebiets so gut wie möglich zu verstehen, um Software exakt auf reale Anforderungen zuzuschneiden. Obwohl das Ziel keine Theoriebildung, sondern die Entwicklung von Werkzeugen und Softwaresystemen ist, erinnert dieser Ansatz an theoriegenerierende Ansätze der qualitativen Sozialforschung wie etwa die Grounded Theory (GT).

Neighbors geht nicht auf Erhebungsmethoden ein. Spätere Softwareingenieur:innen beziehen sich explizit auf qualitative Sozialforschung und nicht näher spezifizierte ethnografische Theorie und Methodik (McGraw und Harbison 1997: 7f.; Kotonya und Sommerville 1998: 53). Auch implizit wird auf den direkten Austausch zwischen Softwareingenieur:in und nutzender Person hingewiesen, da Kotonya und Sommerville in ihrem Erhebungsprozess den Computer als störend empfinden (ebd.: 41). Diese Form der Erhebung, aus der eine Modellierung hervorgeht, die in ihrer Begrifflichkeit eine sozio-technische Assemblage evoziert, stellt eine spezifische Anwendung ethnografischer Methodologie im SE dar.

In ihrem Lehrbuch für Requirement Engineering stellen Kotonya und Sommerville sämtliche Prozesse vor und diskutieren die methodische Umsetzung kritisch: Im Kontext des SE sind Datenerhebung und -analyse untrennbar miteinander verflochten. Eine Trennung wäre insofern nicht sinnvoll, da das Forschungssubjekt nicht statischer Natur ist (Kotonya und Sommerville 1998: 57). Allerdings bleibt zentral, dass Anforderungen das Resultat fortlaufender Verhandlungen zwischen Beteiligten darstellen, und niemals als endgültig betrachtet werden können (ebd.). Dies führt zu Konzepten, die in den GSW ausführlich, jedoch mittels anderer Begriffe behandelt werden. Wenngleich Kotonya und Sommerville qualitative Forschung referenzieren, so evoziert deren Beschreibung eine „starke Nähe zum Feld“ mit einer „theoretisch und methodisch offenen Verfahrensweise“, die „weitläufig, unstrukturiert, nicht planbar,“ ist, wie es von Irene Götz (2010: 102) als Grundlage arbeitsethnografischer Forschungsmethodologie beschrieben wird. Induktive Vorgehensweisen wurden in den GSW in weitaus größerem Detail beschrieben und entwickelt. Die Verflechtung von Erhebung und Analyse wird bei Kotonya und Sommerville zwar diskutiert, es ist jedoch essenziell, die Erhebungsmethodologie,

wie sie in den Herkunftsdisziplinen entwickelt wurde, als Grundlage für die Anforderungsanalyse zu nehmen. Dies nicht zuletzt, da die Methodologie von und auch für die Expert:innen entwickelt wurde, um die es letztlich bei der Aushandlung von Anforderungen in meinem Projekt als besonderen Fall im RSE geht. Der Umgang dieser Expert:innen mit sensiblen Forschungsdaten kann so nachhaltig in der Entwicklung von digitalen Infrastrukturen verankert werden.

Die Anwendung von GT zur Generierung von Anforderungen beschreibt Mohamad Halaweh kritisch und zeigt, dass GT bisher hauptsächlich in der Beforschung von Informationssystemen und dem Vorgehen in Softwaretechnikprojekten selbst Anwendung findet (Halaweh 2012: 25; Sorrentino und Virili 2005). Dennoch gibt es seit den 2000er Jahren vermehrt Forschungen, bei denen GT direkt für die Modellierung eines Feldes oder die Generierung von Anforderungen im Kontext von SE (Chakraborty und Dehlinger 2009; Würfel et al. 2016) und speziell im Bereich RSE (Kaufmann et al. 2022) genutzt wird. Für eine erfolgreiche Anwendung im Software-Kontext sollten die Analyst:innen nicht nur theoretische Sensibilität für die Bedarfe des Feldes, sondern ebenso für das zu lösende (technische) Problem und seine Grenzen haben (Halaweh 2012: 24, 28). Gleichzeitig sollte die Kommunikation mit den Nutzenden zur Validierung von durch GT erzeugten Anforderungen nicht zu technisch sein (ebd.: 34).

Schlussendlich führt uns Halawehs Kritik der doppelten theoretischen Sensibilität, die sowohl die Bedarfe der Nutzenden als auch der technischen Möglichkeiten abdecken muss, gepaart mit seiner Kritik an der mangelhaften interdisziplinären Kommunikation und Übersetzung von GT-erzeugten Anforderungen zur Lösungsumsetzung, zurück zur Domain Analysis im Review von Shams-ul-Arif et al. Sie halten fest, dass diese komplexe Art der Forschung notwendigerweise durch eine:n Expert:in der Domain selbst durchzuführen ist (2010: 48). Zusammen mit den nicht weiter ausgeführten Konzepten, die mit Begriffen wie „Problembereich“ (Neighbors 1980: 1) oder „parts-and-assemblies concept“ (ebd.: 7) im SE angerissen werden, führte mich diese Forderung zurück zu einer Betrachtung von Feld- und Technikforschung in den GSW, um meine Methodologie weiterzuentwickeln.

Feldbegriff und Technologie in Kulturanthropologie und Soziologie

Konzepte, Begriffe und Herangehensweisen der Ethnografie bilden die Grundlage nutzer:innenzentrierter Anforderungsanalyse (McGraw und Harbison 1997: 7f.; Kotonya und Sommerville 1998: 41, 53) und ihr Verständnis ist zugleich zentral für die kollaborative Beforschung der Arbeitsweisen qualitativ Forschender in der Anforderungsanalyse. Begriffe wie „Feld“ und „Feldforschung“ werden immer wieder verwendet.

Sabine Hess und Maria Schwertl (2013: 25) betonen, dass das Feld nicht nur geografisch verstanden werden kann, sondern als eine Konstruktion betrachtet wird,

die durch die Praxis der Forschung entsteht (ebd.: 32). Deepa S. Reddy beschreibt das Feld als eine fast zufällige Ansammlung von Orten, die durch den Prozess der Feldforschung selbst kohärent werden (Reddy 2009).

Theoretische Konzepte ethnografischer Forschung, die sich mit der Konstruktion, Beschaffenheit und Dynamik komplexer Felder beschäftigt, erstrecken sich über „Akteursnetzwerke, Dispositive und Regime“ (Hess und Schwertl 2013: 30) bis zum Begriff der Assemblage (Welz 2021), der menschliche und nicht-menschliche Akteur:innen, ihre gegenseitige Beeinflussung und ihre miteinander verbundenen Beziehungen als vorübergehende Formationen beschreibt (Vepřek 2024: 15). Beide genannten Konzepte – Assemblage und Akteursnetzwerk – teilen laut Ina Dietzsch grundlegende Gemeinsamkeiten: Sie betonen die Bedeutung von Elementen und ihren relationalen Verbindungen, sowie die Notwendigkeit, von einer radikalen Symmetrie auszugehen, um Hierarchien und Machteffekte sichtbar zu machen. Sie beschreiben die Struktur dieser Verbindungen als beweglich und komplex, um die Entstehungsprozesse und Emergenz von Systemen sichtbar zu machen (Dietzsch 2020: 86). George Marcus thematisiert die Notwendigkeit, diejenigen als Subjekte zu sehen, die in den Zentren der institutionellen Macht agieren, auch wenn sie selbst nicht offenkundig in machtvollen Positionen sind. Er bezeichnet diese Personen als „para-sites“ (oder „Parasiten“ im übertragenen Sinne, Marcus 2000: 1ff.). Ein Para-Site ist jemand, der oder die in einer bestimmten Situation oder Institution agiert, aber nicht unbedingt Teil der traditionellen Machtstruktur ist. Sie sind „moderately empowered“ (ebd.: 4), was bedeutet, dass sie zwar einige Macht haben, aber nicht die absolute Kontrolle über die Situation. Marcus nennt beispielsweise Ärzt:innen und sogar Softwareingenieur:innen (ebd.: 9). Sie sind an der Analyse ihrer Situation interessiert, also an der Beziehung zwischen den verschiedenen Elementen in einer bestimmten Umgebung. Deren „preexisting ethnographic consciousness or curiosity“ (Holmes und Marcus 2008: 82) ermöglicht eine Art der Ethnografie, die Para-Ethnografie, die über Kollaboration hinaus die Analysefähigkeit der beforschten Subjekte einbezieht, um die Probleme und Herausforderungen zu definieren, mit denen diese sich auseinandersetzen (ebd.: 86). Dieses gleichberechtigte Verhältnis und die Ermächtigung der Beforschten sind relevante Aspekte einer nutzer:innenzentrierten, kollaborativen Anforderungsanalyse.

Die Berücksichtigung der Komplexität eines Feldes – einschließlich seiner Machtstrukturen, inhärenten Dynamiken und insbesondere seiner praxeologischen Konstruktion – trägt wesentlich zu einer nachvollziehbaren Anforderungsanalyse bei. Jedoch fordert die Beforschung für die weitere Entwicklung einer digitalen Technologie zusätzliche Ansätze, die das Feld begreifbar machen. Die Multi-Sited Ethnography (Marcus 1995) ist ein Paradigma der Feldforschung, bei dem Forschende nicht nur an einem einzelnen Ort beobachten, sondern verschiedene miteinander verbundene Orte, Menschen, Gegenstände, Geschichten und Konflikte in ihre Analyse einbeziehen, um komplexe, globale Zusammenhänge besser zu

verstehen. Für digitale Technologien ist dieser Zugang besonders relevant, denn sie sind global vernetzt und überschreiten geografische Grenzen: Sie werden an verschiedenen Orten genutzt, beeinflusst und vor allem (weiter)entwickelt. Retrospektiv kann man sagen, dass Marcus' Beispiele bereits in den 1990er Jahren die Anwendung für das Feld Open Data andeuten. Er thematisiert zum einen eine Anwendung auf „electronic communication such as the internet“, zum anderen auf die „emergence of biotechnology and ‚big‘ science projects“ (ebd.: 104). Diese Anwendungsbeispiele beziehen sich sowohl auf das relevante Medium (Internet) als auch auf geteilte wissenschaftliche Daten über dieses Medium. In industriellen Kontexten sind Felder leichter zu definieren, da Software oft für eine spezifische Nutzer:innengruppe entwickelt wird. Im Gegensatz dazu ist Forschungssoftware oft Open Source (Struck et al. 2020) und sollte FAIR sein (auffindbar, zugänglich, interoperabel, wiederverwendbar, Lamprecht et al. 2020), was ein Feld schafft, das (inter-)disziplinär verteilt ist mit verschiedenen Kontexten und Nutzer:innen-gruppen. Dieser Verteilung lässt sich durch Multi-Sited Ethnography analytisch entgegenwirken.

Methodologie iterativer Entwicklung von Forschungssoftware Kontexte im Research Software Engineering

Für die Erforschung von digitalen Technologien im Rahmen von RSE sind Ansätze der Science and Technology Studies (STS) besonders hilfreich, die interdisziplinär (Beck et al. 2012: 11) sozio-technische Verwebungen sowie deren Bedeutungen in den Mittelpunkt des Forschungsinteresses stellen (ebd.: 9) und denen George Marcus selbst auch die Multi-Sited Ethnography zuordnet (Marcus 1995). Denn wenn wir anerkennen, dass Forschungsmethoden sozusagen ein soziales Leben haben – dass sie also nicht nur einfache Techniken sind, sondern dass sie bestimmte soziale Welten eröffnen und ich sie somit durch meine eigene Fragestellung und meine eigene Agenda mitgestalte (Law und Ruppert 2013: 233) – sollten diese Annahmen auch sichtbar gemacht werden, da sie Interpretationen und Erkenntnisse in Bezug auf das Forschungsfeld beeinflussen (Silvast und Virtanen 2019: 461).

Im Kontext von Open Research und der Aushandlung von Nachnutzungsmechanismen erlaubt Multi-Sited Ethnography, die darin involvierten globalen Netzwerke zu untersuchen und Verbindungen zwischen Entwickelnden, Nutzenden („follow the people“, Marcus 1995: 106) und Plattformen an unterschiedlichen Orten („follow the thing“, ebd.) zu erkennen. Anhand dieser Herangehensweise kann ich vielschichtige Perspektiven einfangen: Im Kontext digitaler Daten und Technologien gibt es in der Regel viele Akteur:innen mit unterschiedlichen Rollen wie Programmierung, Design, Nutzung, Regulierung und Diskussion um Ethik und Recht. Multi-Sited Ethnography ermöglicht es, all diese Perspektiven an ver-

schiedenen Orten zu berücksichtigen und die Wechselwirkungen zwischen ihnen zu verstehen („follow the plot“, ebd.: 109). Mit dem Fokus, eine Software für die Bedarfe und Prozesse der Nutzenden zu entwickeln, stehen diese Aspekte der Multi-Sited Ethnografie im Vordergrund im Gegensatz zu Aspekten, die bei anderen Forschungskontexten und Forschungsfragen erkenntnisreicher wären („follow the life“, „follow the conflict“).

Agile Entwicklungsprozesse aus dem RSE und qualitative Forschung basierend auf GT teilen den unabgeschlossenen Charakter des Iterativen. Diese Ansätze werden in der iterativen Entwicklung von Forschungssoftware verknüpft durch die Feldkonstruktion. Dabei stehen Erhebungsmethodik und Analyse nicht linear hintereinander. Zur besseren Nachvollziehbarkeit und der linearen Textform geschuldet erläutere ich sie dennoch nacheinander.

Erhebung und Feldkonstruktion

Den Startpunkt der praxeologischen Konstruktion meines Feldes setze ich mit meiner Auseinandersetzung des zu Beginn genannten Paradigmas von Open Research oder der generellen Openness, die – im Gegensatz zum oft verwendeten Begriff Open Science – geisteswissenschaftliche Forschung explizit einschließt (Longley und Hearn 2021: 829). Dieses Paradigma begegnete mir zuerst in Form des 4. Berichts des Rats für Informationsinfrastrukturen (RfII 2023). Dieser Bericht skizziert pragmatisch den Stand der Dinge, Zielsetzung und geplanten Weg hin zu offener Forschung durch Infrastrukturen, wie der NFDI, bis ins Jahr 2028 (ebd.: 11). In meinen Notizen zum Text wird deutlich, dass ich bereits zu diesem Zeitpunkt selbst nach Unterrepräsentation, zunächst der Digital Humanities, in existierenden Infrastrukturen suchte und diese auch fand (ebd.: 17). Dennoch stellt dieses zentrale Dokument aus Sicht des Rates das Paradigma von Open Research mit einer „Verankerung in den wissenschaftlichen Gemeinschaften“ grundlegend in den Vordergrund (ebd.: 29). Im gesamten Bericht wird immer wieder betont, dass diese Entwicklung anhand der Bedarfe aus den Fachgemeinschaften erfolgen (exemplarisch ebd.: 34) und maßgeblich durch diese geprägt werden soll. Aus diesem Bericht lese ich einen Plot, bei dem sich Menschen an unterschiedlichen Orten mit potenziellen Problemen im Umgang mit dem Open Research Paradigma auseinandersetzen sollen. Hiervon ausgehend folgte ich dem Plot, Menschen, Dingen und Orten, wie es George Marcus (1995) beschreibt. Gleichzeitig spielen – in den Worten des SE – Anforderungen eine entscheidende Rolle in der Ausgestaltung des Feldes, das geprägt ist durch menschliche und nicht-menschliche Akteur:innen, die in temporären Zusammenkünften und Auseinandersetzungen, sowohl online als auch offline, die Datengrundlage meiner Anforderungsanalyse bilden. Dementsprechend besuchte ich digitale und analoge Räume, an denen Menschen sich austauschten über 1. ihre existierenden Datenpraxen in den GSW, 2. den Um-

gang GSW-Forschender mit dem Openness-Paradigma und bereits entwickelten technischen Möglichkeiten, und 3. den Umgang Forschender aus unterschiedlichen Disziplinen mit dem Openness-Paradigma und den bestehenden technischen Möglichkeiten. Das führte mich auch zu bestehenden Bedarfsanalysen aus den Fachgemeinschaften (etwa Ambrasat and Heger 2019; Imeri and Danciu [Mitarb.] 2017; Strunk 2018; Wuttke 2022).

Die Problematik des Openness-Paradigmas eröffnete mir also ein Feld, als ich begann den Fragen nach der Umsetzung innerhalb der Forschungsgemeinschaften zu folgen. Dies führte mich zu Tagungen, Workshops und Einzelgesprächen mit Menschen, die mit Maßnahmen, Fragestellungen und Infrastrukturen hin zu Openness direkt im Zusammenhang stehen. Gleichzeitig kann ich multi-sited den Weg vom 4. Bericht des Rats für Informationsinfrastrukturen zu den Bedarfsanalysen aus den Fachgemeinschaften direkt nachverfolgen, den ich in einer analytisch-iterativen Schleife zu Beginn meiner Forschung vornahm, um meine Forschungsfragen zu präzisieren.

Eindeutig sind diese Forschungsfragen und somit auch das Feld zusätzlich geprägt von Konzepten und Sprache des SE, deren Akteur:innen und Auseinandersetzungen. Denn die Anforderungen GSW-Forschender an das Teilen qualitativer Interviewdaten und deren Unterstützung durch gezielt entwickelte Software stellen die Fragen dar, die aus den Disziplinen selbst gestellt wurden und die ich aus den bestehenden Bedarfsanalysen heraus nicht beantworten konnte. Als Forschende, die sich sowohl den GSW als auch dem SE zuordnet, relationiere ich mein Umfeld, meine Informationsquellen und die Veranstaltungen und Diskussionen, an denen ich beobachtend teilnehme, interdisziplinär und temporär zur Repräsentation einer Assemblage, also zu meinem Feld, in dem ich mich para-ethnografisch bewege. In meiner Interdisziplinarität bin ich auch Teil der qualitativ Forschenden. Hier orientiere ich mich an CI, die bereits seit 1986 von Karen Holtzblatt entwickelt wird. Kollaborativ wird dabei mit dem Meister:in-Schüler:in-Prinzip geforscht: Die qualitativ Forschenden sind für mich die Expert:innen ihrer eigenen Tätigkeit, während ich in meiner Beobachtung die Rolle der Lernenden einnehme, um deren Arbeitsprozesse, Entscheidungen und Denkweisen im Umgang mit ihren Daten zu verstehen (Beyer und Holtzblatt 1995: 48).

Der in den oben genannten digitalen und analogen Räumen immer wieder thematisierte Konflikt, eine qualitative Datenpraxis mit existierenden technischen Ansätzen des Openness-Paradigmas zusammenzuführen, brachte mich zu den in diesem Zusammenhang oft genannten und offenbar besonders relevanten Data Stewards. Diese Rolle ist nicht klar definiert, wurde aber in ihrer Vielfalt von den Informationswissenschaftler:innen Ana Carballo-Garcia und Juan-José Boté-Vericad so beschrieben, wie ich es auch im Feld wiederfand: Data Stewards sind demnach spezialisierte Fachkräfte, die Forschende bei der Verwaltung, Wiederverwendung, Speicherung und langfristigen Sicherung von Forschungsdaten unterstützen.

Sie vermitteln Datenrichtlinien und Standards an Forschende, unterstützen bei Datenmanagementplänen (DMPs) und beraten Forschende zu Aspekten wie Datenaufbereitung, Dokumentation und Schutz (Carballo-Garcia und Boté-Vericad 2022: 50f.). Es handelt sich um Expert:innen der Datenkuration an Forschungsdatenzentren und Expert:innen des Forschungsdatenmanagements an Universitäten. Diese Menschen bilden in ihrer Arbeit eine Schnittstelle zwischen Praktiken, Erwartungen, Bedürfnissen und Möglichkeiten der Nutzenden in den GSW selbst, rechtlichen Rahmenbedingungen und technischen Gegebenheiten. Sie sind also eng an den Konflikten des Plots beteiligt, berichten über diese aus der Perspektive der Forschenden, der Technologieentwicklung und ihrer eigenen Arbeit. Sie bilden ein Querschnittswissen des Feldes ab, dass ich in meiner Ergebnisdarstellung (Mollenhauer 2025) unterteilt habe in die Bereiche „Forschungsdatenmanagement“, „Softwaretechnik und Data Engineering“ und „Geistes- und Sozialwissenschaften“.

Ich führte Expert:inneninterviews mit Data Stewards, um meine Feldnotizen anzureichern. Die Auswahl und Anzahl der Data Stewards war dabei angelehnt an verschiedene Positionen, die diese rund um offene Forschungsdaten einnehmen: Disziplinen-spezifische Data Stewards der GSW, nicht disziplinär-verankerte Data Stewards mit technischem Hintergrund und Data Stewards mit unterschiedlichen institutionellen Anbindungen. Mein Auffinden von Menschen in solch vermittelnden Positionen war unterschiedlich: Manchmal begegneten sie mir durch Zufall bei einer Veranstaltung, ich schrieb aber auch systematisch direkt Universitäten und Forschungsdatenzentren an, die in ihren Selbstbeschreibungen vorgaben, qualitative Daten zu verarbeiten. Außerdem ging ich aktiv auf die Suche, wenn ich auf öffentlich zugänglichen Publikationsservern Transkripte qualitativer Interviews fand. Das Sample entwickelte sich somit iterativ aus meiner zunehmenden Feldkenntnis. Ich führte Interviews mit diesen Expert:innen sehr offen, indem ich meine Gesprächspartner:innen darum bat, den Umgang mit dem Konflikt um Forschungsdatenmanagement ihrerseits und seitens der Forschenden zu schildern.

Feldkonstruktion und Analyse

Der 4. Bericht des Rats für Informationsinfrastrukturen (RfII 2023) stellt neben der Feldkonstruktion auch den Ausgangspunkt meines analytischen Prozesses dar. Ganz im Sinne der GT entwickelte ich hiervon ausgehend eine theoretische Sensibilität (Truschkat et al. 2011: 358). Auch im weiteren Verlauf setzt sich meine Herangehensweise aus den Begegnungen im Feld und Interviews mit Data Stewards sowie gezielten, systematischen Analysen dieser Daten im Zusammenspiel mit der Analyse der Bedarfsanalysen aus den Fachgemeinschaften zusammen, die in der Präzisierung der Frage nach Anforderungen an den Aufbereitungsprozess qualitativer Interviewtranskripte als digitale Daten mündet (ebd.: 363).

Das erste Material, das ich mit dem Ziel der Identifikation genereller Anforderungen analysierte, waren die Transkripte der Expert:inneninterviews. Durch Annotation einzelner Stellen, konnte ich das Material entlang der Konfliktlinien der Plots aufbrechen und im Zuge des offenen und axialen Kodierens Anforderungen herausarbeiten. In iterativen Schleifen, in denen die weitere Sammlung von Daten und die Erweiterung des Feldes Abstand zu dieser Analyse gab, ergänzte ich kodierte Elemente und setzte diese schließlich in Kategorien miteinander in Beziehung. In der Folge kamen zu den Interviewtranskripten auch meine para-ethnografischen Feldnotizen hinzu.

Spätestens ab diesem Punkt unterscheidet sich der gemeinsame methodologische Weg von GSW Forschung und (R)SE. Wo GSW-Forschende den letzten Schritt der Ausarbeitung einer Kernkategorie durch selektives Kodieren nutzen, um eine „gegenstandsbezogene Theorie“ (ebd.: 369) zu bilden, zielt meine RSE-Analyse darauf ab, Artefakte oder Zwischenergebnisse zu bilden, die mich im Vorhaben, eine Software nutzer:innenzentriert zu entwickeln, weiterbringen. Ein Prototyp dieser Software wiederum dient iterativ dazu, die Anforderungen weiterzuentwickeln. Das bedeutet also, dass meine GT-Analyse zwar zuerst offen und dann axial verlief, sie jedoch von Beginn an sehr thematisch fokussiert war. Fokus der Interviewfragen und Auswahl von Expert:innen spiegelten sich im Ergebnis des ersten offenen Kodierens wider, das sehr konzentriert war auf meine Zielsetzung, eine machbare, für die Nutzenden sinnvolle Software zu entwerfen. Die Kondensierung der offenen Kodierung in bestimmte Kategorien als axiales Kodieren folgte daher recht schnell und wurde unterbrochen durch iteratives, weiteres offenes Kodieren.

Abb. 2: Schematische Darstellung einer Tabelle zur Anforderungsermittlung durch GT

Traceability Mapping		
Kodiertes Segment	Kategorie 1	Anforderung 1 Anforderung 2
Kodiertes Segment		
Kodiertes Segment	Kategorie 2	Anforderung 3

Im Anschluss nutzte ich das daraus entstandene Kategoriensystem, um aus jeder Kategorie anhand der Aussagen in den annotierten Segmenten Anforderungen zu formulieren, wie in Abbildung 2 schematisch dargestellt.

Ich ordne diesen Schritt als den ersten in einer Reihe ein, die ich dem axialen Kodieren zurechne, da diese Anforderungen über die einzelnen Kategorien hinaus „empirische Beziehungen“ (ebd.: 368) herstellen. Meine Anforderungsformulierungen entsprechen ganz bewusst dem ISO/IEC/IEEE „International Standard – Systems and software engineering – Life cycle processes – Requirements engineering“, der sehr technisch vorgibt, dass Anforderungen klar ausformulieren, welche Funktionalität im Einzelnen zu welchem Grad erfüllt werden soll (ISO/IEC/IEEE 2018: 15f.).

GT-Analyse wird von unterschiedlichen Softwareingenieur:innen genutzt, um im Entwicklungsprozess Entscheidungen zu Funktionalitäten und Design von Software zu fällen (Mitchell et al. 2022; Halaweh 2012; Würfel 2016; Kaufmann et al. 2020; Mucha 2023). Die Weiternutzung der axialen Kodierung ist jedoch unterschiedlich. Beispielsweise folgt bei Würfel et al. die Ableitung einer analytischen Geschichte zum Ablauf, den eine nutzende Person durchläuft, um ein bestimmtes Ziel zu erreichen. Eine analytische Geschichte ist vergleichbar mit, aber ausführlicher als eine User Story, der am häufigsten genutzten Softwarefunktionsbeschreibung im agilen SE (Raharjana et al. 2021: 53811). Da User Stories jedoch sehr kurze, persönlich wirkende Beschreibungen sind, entschied ich mich, Anforderungen generisch zu formulieren. Denn der Rückbezug zum Kontext, zu den dichten Beschreibungen der Interviews, soll notwendigerweise erhalten bleiben und möchte ich mir selbst (und anderen) keine Möglichkeit geben, diese durch eine scheinbar vollständige, jedoch verkürzte Beschreibung von „Als <Rolle> möchte ich <Ziel/Wunsch> um <Nutzen>“ der klassischen User Story (Scrum Alliance o. J.) zu ersetzen. Außerdem steht bei mir im Gegensatz zu den oben genannten GT-Analysen im SE, die Formulierung konkreter Anforderungen nicht am Ende des offenen, axialen und selektiven Kodierens, sondern findet innerhalb dieser Schritte iterativ statt.

Die konkrete Fallbeschreibung im Zusammenhang mit Anforderungen für eine Softwareentwicklung macht das Vorgehen vergleichbar mit dem „selective coding step“ bei Würfel et al. (2016: 652), wo eine zentrale Kategorie in Beziehung zu anderen Kategorien gesetzt wird. In meiner Umsetzung habe ich zu diesem Zweck weitere Artefakte erarbeitet, die mir dabei halfen: Zuerst modellierte ich den Datenverarbeitungsprozess, der generell für alle Forschungsdaten generisch bestimmte Schritte enthält, und erweiterte diesen in einem Diagramm mit allen Schritten, die ich in meinen Anforderungen wiederfand. Dies half mir, besser abzuschätzen, ob eine Softwareentwicklung für einen bestimmten Schritt machbar sein würde. Über die Nutzbarkeit für Anwender:innen hinaus auch die technische Machbarkeit abzuschätzen, ist ein wichtiger Schritt im fundierten SE (Cooper 2015; Mitchell et al. 2022). Die Modellierung half mir daher nicht nur für die Relevanz

der Anforderungen für meine nächsten Arbeitsschritte. Durch die Zuordnung zu den Anforderungen im Kategoriensystem konnte ich zurück zu den ursprünglichen Kontexten in den Interviews gehen, um parallel verschiedene, mögliche Lösung zu entwickeln – also in der Entwicklung des Softwaredesigns.

Da ich meine Fragen zu Anfang offen gestellt hatte, boten mir die Beschreibungen der Expert:innen im Zusammenhang mit Feldnotizen, verschiedene existierende oder mögliche Technologien in Bezug zur Praxis, zu den Möglichkeiten und Fähigkeiten der Nutzenden, zu den infrastrukturellen Gegebenheiten sowie zum Anspruch an Open Research. Im Rahmen konkreter Anwendungsfälle kodierte ich entsprechend der Datenverarbeitungsschritte erneut selektiv in der Tiefe der dichten Beschreibungen mit stärker fokussierter Sichtweise.

Statt die Einschätzung der Expert:innen zu bestimmten Technologien quantifizierend einem Scoring zu unterziehen, wie es gängige Praxis im SE ist (Mitchell 2022; Parolin 2024: 8f.), eröffnet die Analyse ihrer ausführlichen Erzählungen andere Möglichkeiten zur iterativen Analyse. Im Gegensatz zu den Zahlen einer Scoring-Karte sind meine Transkripte und Feldnotizen dichter („thick descriptions“, Geertz 2008) und ich kann immer wieder zu ihnen zurückkehren, um auf verschiedenen Ebenen zu analysieren. Dieser iterative Umgang mit dem Kontext von Expert:innenmeinungen erspart mir den Einbau weiterer Schritte im SE, beispielsweise um Anforderungen priorisieren zu können, indem quantitative Daten qualitativ rekontextualisiert werden. Dies wird im SE verwendet, wenn der Ursprung von Anforderungen unbekannt ist oder ursprüngliche Kontexte nicht dokumentiert wurden (Rojas et al. 2022: 131772f.). SE wird häufig ausschließlich von Softwareingenieur:innen betrieben, ethnografische Erhebungsmethodologien und qualitative Analysemethoden werden nicht reflektiert und weiterentwickelt. Die hier entworfene RSE entsteht hingegen aus einem interdisziplinären Kontext. Durch die Methodologie meines Projekts erhebe und analysiere ich qualitativ und kontextgebunden. Hierdurch werden Ansätze, die im SE seit langem adaptiert wurden, wesentlich vertieft und fundiert. Das ist möglich, weil es nicht im industriellen Sektor des SE, sondern in der Wissenschaft, dem RSE, stattfindet und den Ansprüchen an FAIR Softwareentwicklung entgegenkommt.

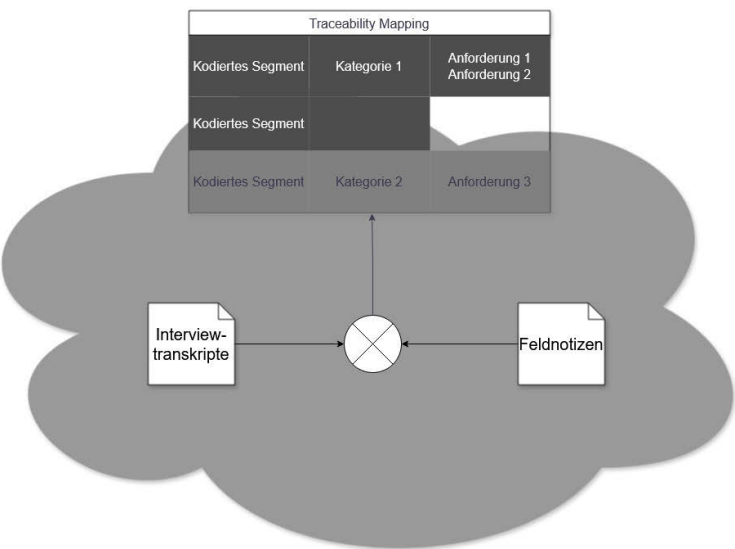
Zukünftige Schritte im iterativen Vorgehen mittels Traceability Map

Die Nachvollziehbarkeit der Entwicklung einer Anforderung spielt im weiteren Kontext der SE eine grundlegende Rolle, dort werden für Anforderungen Qualitätskriterien aufgestellt: unter anderem Vollständigkeit, Konsistenz und Nachvollziehbarkeit, die mit dem englischen Begriff *traceability* besetzt ist. *Requirement Traceability* wird definiert als die Identifikation und Dokumentation der Anforderungsherleitung (nach oben) und deren Umsetzung im weiteren Verlauf der Entwicklung (nach unten) (ISO/IEC/IEEE 2018: 5). Wenngleich es derzeit Ansätze

gibt, die Nachvollziehbarkeit GT-basierter Softwareanforderungsanalysen zu verbessern (Mucha 2023), merkte Halaweh bereits 2012 an, dass eine Repräsentation der Anforderungen in den klassischen Modellen der SE nicht in der Lage ist, die Lücke zwischen qualitativen Daten und den ermittelten Anforderungen zu schließen (Halaweh 2012: 28).

Grundlegend für das iterative Vorgehen dieses Projekts ist das Traceability Mapping, wie es in Abbildung 3 dargestellt wird.

Abb. 3: Schema des Traceability Mapping von Feld, Daten und Anforderungen



Die prototypische Umsetzung ausgewählter Anforderungen oder Lösungsansätze in einer Software dient dabei der iterativen Weiterentwicklung der Anforderungen unter Einbeziehungen der Nutzenden. Dies geschieht anhand weiterer ethnografischer Forschungsmethodik.

Diskussion der Erhebungs- und Analysemethoden im Kontext des Research Software Engineering

Obwohl SE seit Jahrzehnten in Forschungskontexten betrieben wird, wurde RSE als Konzept im Rahmen eines Workshops erst 2012 (Software Sustainability Institute o. J.) diskutiert, um Sichtbarkeit und Gemeinschaft zu schaffen. Im Rahmen dieser Gemeinschaft ist interdisziplinärer Austausch wichtig, da die meisten Research

Software Engineers Autodidakt:innen sind (Cohen et al. 2021: 3) und eine Herkunftsdisziplin abseits des SE haben. Meine Forschung steht prinzipiell in dieser Tradition, da sie in einer Disziplin außerhalb des SE verankert ist, ich jedoch fundierte SE-Kenntnisse über Autodidaktik hinaus besitze. RSE stellt, wie die Digital Humanities¹ selbst, eine „trading zone“ (Galison 1997: 781ff.) dar (DH-RSE o. J.), in der getrennte Disziplinen verbunden werden durch „experimenting, theorizing, and instrument building“ (Galison 1997: 803), und in der ich mich mit anderen Research Software Engineers über Methoden, Ansätze, Theorien und Datenpraktiken austauschen und diese gemeinsam weiterentwickeln kann. Im Kontext der RSE bieten Multi-Sited Ethnography und para-ethnografische Erhebung, GT-Analyse sowie die Problemstellung der digitalen Aufbereitung von sensiblen Interviewdaten in qualitativer Forschung mehrere Vorteile. Auf diese möchte ich in Folge eingehen.

Qualitative, iterative Erhebungs- und Analysemethoden sind im Bereich des SE zwar selten, aber nicht neu (Halaweh 2012: 27; Sorrentino und Virili 2005) und werden zunehmend zur Anforderungsanalyse angewendet (Kaufmann et al. 2022). Dennoch ist der Anwendungsfall von Software, die der Aufbereitung von Forschungsdaten in den GSW dienen soll, ein besonderer, mit spezifischem Potential für das RSE. Halaweh schlägt vor, GT als Analysemethode zu nutzen, um im SE über technische Aspekte hinaus einen Fokus auf soziale und organisatorische Aspekte zu legen (Halaweh 2012: 1). Dieser Aspekt muss im RSE nicht extra berücksichtigt werden, da die Entwickelnden selbst meist Teil der Nutzenden-Community sind. In meinem Projekt erhebe ich Anforderungen, denen ich selbst durch die Verarbeitung sensibler, unstrukturierter Textdaten unterliege, in einem GSW-Umfeld, das sich mit den Implikationen dieser Daten beschäftigt und deren organisatorischer und technischer Grenzen ich in meiner Arbeit direkt ausgesetzt bin. Einen weiteren kritischen Aspekt bei der Umsetzung stellt für Halaweh das Scheitern der Kommunikation in einem interdisziplinären Team dar, dass dazu führt, dass die durch GT gewonnenen Erkenntnisse sich nicht in der Entwicklung niederschlagen (2012: 24). Mein interdisziplinäres Feld verstehe ich mit Hilfe des Konzepts der Assemblage als eine sich ständig wandelnde Repräsentation, die durch meine Forschung und ihre schriftliche Darstellung erst festgehalten wird. Dabei sind permanente Übersetzungsleistungen verschiedener Sprachen im Feld, bei der Erhebung, während der Analyse und zur Kommunikation aller Ergebnisse notwendig und hilfreich, und erst durch einen Multi-Sited-Ansatz möglich. Laut Marcus ist die Übersetzung in der Multi-Sited-Forschung nicht nur eine technische Aufgabe, sondern auch eine kulturelle und soziale Praxis, welche die verschiedenen Sites verbindet und soziale Brüche zwischen ihnen aufdeckt (Marcus 1995: 100). In Verbindung mit einer Multi-Sited ethnografischen Erhebung, bietet die GT daher vor allem im RSE über die

1 Die DH-RSE ist den Digital Humanities zugeordnet als AG im Verband Digital Humanities im deutschsprachigen Raum (DHd).

bisher genutzten Elemente der GT hinaus einen Lösungsansatz für das Problem der notwendigen Übersetzungs- und Kommunikationsleistung zwischen den Nutzenden, den Anforderungsanalyt:innen und letztlich denjenigen, die einzelne Anforderungen technisch implementieren.

Ethnografische Erhebungsmethoden wie Multi-Sited Ethnography oder CI, die über die bekannten Standards hinaus gehen, sind im RSE noch nicht angekommen. Methodologien aus Herkunftswissenschaften vermehrt zu integrieren, ist das besondere Potential von RSE, da es im Gegensatz zum SE bereits eine grundlegende Interdisziplinarität besitzt. Wie die meisten Research Software Engineers (Damerow 2025: 121) bin ich in meiner Forschung allein und führe sowohl die Anforderungsanalyse als auch die technische Entwicklung einer potenziellen Problemlösung durch. Dies stellt in mancherlei Hinsicht Schwierigkeiten dar, etwa einen Mangel an Infrastruktur für den Code-Review-Prozess (ebd.). Für mich ist es allerdings ein Vorteil. Denn nicht nur die Anforderungen, sondern deren Datenbasis, die Probleme, Konflikte, menschliche und nicht-menschliche Akteur:innen fließen dauerhaft in den iterativen Entwicklungsprozess mit ein, wenn ich diese Rolle in einer Person erfülle und die verschiedenen Aspekte immer wieder miteinander verweben kann und muss, da sie mir alle bekannt sind.

Es gibt im RSE jedoch Aspekte über die Einzelforschung hinaus, die der Multi-Sited Ethnography einen passenden Anwendungsraum bieten. Cohen et al. stellen den Aspekt der Community als Notwendigkeit im RSE dar, um interdisziplinären Austausch zwischen einzeln arbeitenden Research Software Engineers zu fördern (Cohen et al. 2021: 2). Nicht zuletzt schreibt die NFDI dem RSE eben diese Querschnittsrolle der interdisziplinären Kommunikation aktiv zu (Felderer et al. 2023: 1). Wo RSE mir einen grenzüberschreitenden interdisziplinären Raum bietet, kann ich durch Multi-Sited Ethnography ein Bewusstsein für mein Vorgehen entwickeln, meine Feldkonstruktion reflektieren und folglich auch meine GT-Analyse nachvollziehbar und theoretisch fundiert darstellen.

Hervorheben möchte ich vor allem eine Konfliktlinie, die sich durch sämtliche Bestandteile der von mir beforschten Assemblage zieht: der Umgang mit sensiblen, personenbezogenen Daten. Ich möchte diesen Plot als zentral und exemplarisch hervorheben, da er sowohl die Dreiteilung meines Feldes, beziehungsweise die Verbundenheit dieser Teile und aller Akteur:innen darin darstellt, als auch die Notwendigkeit einer Multi-Sited Ethnography, die von einer para-ethnografisch Forschenden durchgeführt wird. Betrachten wir die Anforderungen ohne das von mir erzeugte Feld dahinter, brachte dieser Plot analytisch scheinbar einzig die banale Anforderung „The tool shall encrypt sensitive data“ hervor. In meiner Analyse habe ich jedoch diesen Plot durch die Daten meines Feldes verfolgbarm gemacht. Dadurch kann ich nachvollziehen, dass im speziellen Fall des Umgangs mit sensiblen, personenbezogenen Daten, die sich in unstrukturierten Textdaten befinden, ein Konflikt besteht, der über die GSW hinaus in anderen Feldern, wie

der Medizininformatik (Lin 2009), durchaus bearbeitet und diskutiert wird. Der von mir verfolgte interdisziplinäre Ansatz ermöglicht es, durch die praxeologische Einordnung des Feldes und dessen Repräsentation neue Perspektiven zu eröffnen. Im Kontext der RSE als Trading Zone, ermöglicht mir der Ansatz Plotlinien, die über Disziplingrenzen hinaus gehen, zu verbinden und somit auch Lösungsansätze verschiedener Disziplinen miteinander abzugleichen, sich darüber auszutauschen und diese weiterzuentwickeln. In diesem Sinn bietet Multi-Sited Ethnography wichtiges Potential für innovative, nutzer:innenzentrierte Entwicklungen im RSE.

Relevanz und Kontext des Ansatzes

Wenngleich in diesem Beitrag die methodologische Grundlage meiner Forschung erläutert wird, stellen die Ergebnisse zugleich einen Beitrag zur Entwicklung von FAIR und Open Research dar (Mollenhauer 2025). Diese können erst durch die spezifische Methodenkombination erlangt werden. Ganz grundlegend ermöglichen GSW-Konzepte, Ansätze und Methoden das seitens der SE eingeforderte Qualitätskriterium der Requirement Traceability anhand des so bewusst und durch Traceability Mapping sichtbar gemachten Feldes.

Wie der Anthropologe Michael M. J. Fischer schreibt, ist Kollaboration für die erfolgreiche Umsetzung von wissenschaftlicher, offener technologischer Entwicklung, die über para-ethnografische Forschung hinaus geht, essenziell. Er beschreibt dazu den Versuch des Integrative Weaving. Dadurch sollen MINT-Fächer, GSW und Kunst miteinander verbunden werden und der zunehmenden Spezialisierung, Abschottung und dem Mangel an Kommunikation zwischen den Disziplinen im Kontext großer Projektstrukturen entgegengewirkt werden. So sollen einerseits Kontrollmechanismen möglich und andererseits verhindert werden, dass moralische Urteile individualisiert stattfinden, die öffentliche Diskussion geschwächt oder der Zugang zu Wissen und Macht diskriminierend eingeschränkt wird (Fischer 2006: 173). Demzufolge geht es nicht nur darum, Ethnografie in den Dienst guter nutzer:innenzentrierter Softwareentwicklung zu stellen, sondern Wissen und dementsprechend Macht soll über Fächergrenzen hinaus geteilt werden. Dies führt im besten Fall dazu, dass Infrastrukturen und Institutionen – auch im Kontext geteilter Daten – kollaborativ entwickelt werden. Interdisziplinäre Trading Zones, wie DH und RSE, bieten Räume für dieses Integrative Weaving.

In diesem Sinne sind nicht nur die methodologischen und theoretischen Ansätze der GSW wichtig für die weitere Entwicklung des jungen Bereichs RSE, sondern die kritischen Datenpraktiken innerhalb der GSW können auch in die übergreifende Weiterentwicklung von Ansätzen und Infrastrukturen des Open Research Paradigmas einfließen. Die hier vorgestellte Forschung ist für ein Integrative Weaving, bei dem Methoden interdisziplinär und produktiv angewendet und weiterentwickelt

werden, besonders gut geeignet. Denn beim Szenario unstrukturierter, sensibler Textdaten im dynamischen Kontext von Open Research handelt es sich selbst um eines von interdisziplinärer Relevanz.

Literatur

- AG Research Software Engineering in den Digital Humanities (DH-RSE) – digital humanities im deutschsprachigen raum. o. J. Zuletzt abgerufen am 9. März 2026. <https://digitalhumanities.de/ag-research-software-engineering-in-den-digital-humanities-dh-rse/>.
- Ambrasat, Jens, und Christophe Heger. 2019. „*Ergebnisse der Wissenschaftsbefragung 2019/20 (DZHW Monitoringbericht)*“. Berlin: Deutsches Zentrum für Hochschulinformation und Wissenschaftsforschung (DZHW), https://www.wb.dzhw.eu/downloads/wibef_barometer2020.pdf.
- Beck, Stefan, Jörg Niewöhner, und Estrid Sørensen. 2012. „Einleitung. Science and Technology Studies–Wissenschafts- und Technikforschung aus sozial- und kulturanthropologischer Perspektive“. In *Science and Technology Studies. Eine sozialanthropologische Einführung*, herausgegeben von Stefan Beck, Jörg Niewöhner und Estrid Sørensen: 9–48. Transcript.
- Beyer, Hugh R., und Karen Holtzblatt. 1995. „Apprenticing with the Customer“. In *Communications of the ACM* 38 (5): 45–52. <https://doi.org/10.1145/203356.203365>.
- Carballo-Garcia, Ana, und Juan-José Boté-Vericad. 2022. „*Fair Data: History and Present Context.*“ <https://hdl.handle.net/2437/345997>.
- Carvalho, Lucila, Louise Scott, und Ross Jeffery. 2005. „An exploratory study into the use of qualitative research methods in descriptive process modelling“. *Information and Software Technology* 47 (2): 113–127. <https://doi.org/10.1016/j.infsof.2004.06.005>.
- Chakraborty, Suranjan, und Josh Dehlinger. 2009. „Applying the Grounded Theory Method to Derive Enterprise System Requirements“. In *2009 10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing*: 333–338. <https://doi.org/10.1109/SNPD.2009.102>.
- Cohen, Jeremy, Daniel S. Katz, Michelle Barker, Neil Chue Hong, Robert Haines, und Caroline Jay. 2021. „The Four Pillars of Research Software Engineering“. *IEEE Software* 38 (1): 97–105. <https://doi.org/10.1109/MS.2020.2973362>.
- „Collaborations Workshop 2012 (CW12) | Software Sustainability Institute“. o. J. Zuletzt abgerufen am 9. März 2026. <https://www.software.ac.uk/workshop/collaborations-workshop-2012-cw12>.
- Cooper, Robert G. 2006. „Managing Technology Development Projects“. *Research-Technology Management* 49 (6): 23–31. <https://doi.org/10.1080/08956308.2006.11657405>.

- Czerniak, Andreas, Lena Dreher, Emilie Hermans, und Jochen Schirrwagen. 2020. „Wie befördern NFDI und EOSC einen kulturellen Wandel für eine offene Wissenschaft?“ <https://doi.org/10.5281/zenodo.4062376>.
- Damerow, Julia, Rebecca Sutton Koeser, Jeffrey C Carver, und Malte Vogl. 2025. „Code Review in Digital Humanities“. *Digital Scholarship in the Humanities* 40, Nr. Supplement_1: 118–26. <https://doi.org/10.1093/llc/fqae052>.
- Davis, Alan M. 1995. „Object-oriented requirements to object-oriented design: An easy transition?“ *Journal of Systems and Software* 30 (1): 151–159. [https://doi.org/10.1016/0164-1212\(94\)00123-5](https://doi.org/10.1016/0164-1212(94)00123-5).
- Deleuze, Gilles, Félix Guattari, und Charles J. Stivale. 1984. „Concrete Rules and Abstract Machines“. *SubStance* 13 (3/4): 7–19. <https://doi.org/10.2307/3684771>.
- Dietzsch, Ina. „Akteur-Netzwerk-Theorie“. In *Kulturtheoretisch argumentieren. Ein Arbeitsbuch*, herausgegeben von Timo Heimerdinger und Markus Tauschek: 77–99. Waxmann.
- „FAIR Data Spaces | NFDI“. Zuletzt abgerufen am 5. Mai 2025. <https://www.nfdi.de/fair-data-spaces/>.
- Felderer, Michael, Michael Goedicke, Lars Grunske, Wilhelm Hasselbring, Anna-Lena Lamprecht, und Bernhard Rumpe. 2023. „Toward Research Software Engineering Research“. <https://doi.org/10.5281/zenodo.8020525>.
- Fischer, Michael M. J. 2006. „Science, Technology and Society“. *Theory, Culture & Society* 23 (2–3): 172–174. <https://doi.org/10.1177/026327640602300228>.
- Galison, Peter. 1997. „Image and Logic: A Material Culture of Microphysics.“ University of Chicago Press.
- Geertz, Clifford. 2008. „Thick Description: Toward an Interpretive Theory of Culture“. *The Cultural Geography Reader*. Routledge.
- Götz, Irene. 2010. „Ethnografien der Nähe – Anmerkungen zum methodologischen Potenzial neuerer arbeitsethnografischer Forschungen der Europäischen Ethnologie“. *AIS-Studien* (1): 101–117. <https://doi.org/10.21241/ssoar.64752>.
- Halaweh, Mohanad. 2012. „Using Grounded Theory as a Method for System Requirements Analysis“. *Journal of Information Systems and Technology Management* 9 (1): 23–38. <https://doi.org/10.4301/s1807-17752012000100002>.
- Hess, Sabine, und Maria Schwertl. 2013. „Vom ‚Feld‘ zur ‚Assemblage‘? Perspektiven europäisch-ethnologischer Methodenentwicklungen – eine Hinleitung.“ In *Europäisch-ethnologisches Forschen. Neue Methoden und Konzepte*, herausgegeben von Sabine Hess, Johannes Moser und Maria Schwertl: 13–37. Reimer Verlag.
- Holmes, Douglas R., und George E. Marcus. 2008. „Collaboration Today and the Re-Imagination of the Classic Scene of Fieldwork Encounter“. *Collaborative Anthropologies* 1 (1): 81–101.
- Imeri, Sabine, und Ida Danciu [Mitarb.]. 2017. „Open Data. Forschungsdatenmanagement in den ethnologischen Fächern. Auswertung einer Umfrage des Fachinformationsdienstes Sozial- und Kulturanthropologie an der Universitätsbibliothek

- der Humboldt-Universität zu Berlin 2016. Teil I: Statistiken“, https://www.evifa.de/d/e/assets/dokumente/umfrage_bericht_statistiken_1-o_14-06-2017.pdf.
- „ISO/IEC/IEEE International Standard – Systems and software engineering – Life cycle processes – Requirements engineering“. 2018. *ISO/IEC/IEEE 29148:2018(E)*: 1–104. <https://doi.org/10.1109/IEEESTD.2018.8559686>.
- Kaufmann, Andreas, Julia Krause, Nikolay Harutyunyan, Ann Barcomb, und Dirk Riehle. 2022. „A Validation of QDAcity-RE for Domain Modeling Using Qualitative Data Analysis“. *Requirements Engineering* 27 (1): 31–51. <https://doi.org/10.1007/s00766-021-00360-6>.
- Kotonya, Gerald, und Ian Sommerville. 1998. „*Requirements Engineering: Processes and Techniques*.“ Wiley Publishing.
- Lamprecht, Anna-Lena, Leyla Garcia, Mateusz Kuzak, et al. 2020. „Towards FAIR Principles for Research Software“. *Data Science* 3 (1): 37–59. <https://doi.org/10.3233/DS-190026>.
- Latour, Bruno. 1993. „*We Have Never Been Modern*.“ Harvard University Press.
- Law, John, und Evelyn Ruppert. 2013. „The Social Life of Methods: Devices“. *Journal of Cultural Economy* 6 (3): 229–240. <https://doi.org/10.1080/17530350.2013.812042>.
- Li, Hai-tao, Ai-song Shi, und You-xin Meng. 2008. „'Parallel Development and Independent Integration Model' for Scientific Research Software Engineering“. In *2008 International Conference on Computer Science and Software Engineering*, 2: 244–246. <https://doi.org/10.1109/CSSE.2008.1372>.
- Lin, Li-Chen. 2009. „Data Management and Security in Qualitative Research“. *Dimensions of critical care nursing : DCCN* 28: 132–137. <https://doi.org/10.1097/DCC.0b013e31819aeff6>.
- Longley Arthur, Paul, und Lydia Hearn. 2021. „Toward Open Research: A Narrative Review of the Challenges and Opportunities for Open Humanities“. *Journal of Communication: jqabo28*. <https://doi.org/10.1093/joc/jqabo28>.
- Marcus, George E. 1995. „Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography“. *Annual Review of Anthropology* 24: 95–117.
- Marcus, George E., Hg. 2000. „*Para-Sites: A Casebook against Cynical Reason*.“ Bd. 7. Late Editions: Cultural Studies for the End of the Century. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/P/b03628789.html>.
- McGraw, Karen L., und Karan Harbison. 2020. „*User-centered Requirements: The Scenario-based Engineering Process*.“ CRC Press. <https://doi.org/10.1201/9781003064138>.
- Mitchell, Rick, Rob Phaal, Nikoletta Athanassopoulou, Clare Farrukh, und Christian Rassmussen. 2022. „How to Build a Customized Scoring Tool to Evaluate and Select Early-stage Projects“. *Research-Technology Management* 65 (3): 27–38. <https://doi.org/10.1080/08956308.2022.2026185>.
- Mollenhauer, Sabina. 2025. „*Software Requirements for Qualitative Interview Data: A Participatory Approach – From qualitative interview transcripts to FAIR research*“

- data through collaborative software engineering*“. Gehalten auf der Konferenz Forschungsdaten in den Geisteswissenschaften (FORGE), Rostock.
- Mucha, Julia. 2023. „*The QDAcity-RE-RS Method for Creating Complete, Consistent, and Traceable Requirements Specifications*“. <https://open.fau.de/handle/openfau/30075>.
- Neighbors, James M. 1980. „*Software Construction Using Components*“. Ph. D. Thesis, University of California.
- Parolin, Giacomo, Tim C. McAlloone, und Daniela C. A. Pigosso. 2024. „How can technology assessment tools support sustainable innovation? A systematic literature review and synthesis“. *Technovation* 129: 102881. <https://doi.org/10.1016/j.technovation.2023.102881>.
- Philippe, Olivier, Martin Hammitzsch, Stephan Janosch, Anelda van der Walt, Ben van Werkhoven, Simon Hettrick, Daniel S. Katz, et al. 2019. „*softwaresaved/international-survey: Public release for 2018 results*“. <https://doi.org/10.5281/zenodo.2585783>.
- Raccoon, L. B. S. 1997. „Fifty Years of Progress in Software Engineering“. *ACM SIGSOFT Software Engineering Notes* 22 (1): 88–104. <https://doi.org/10.1145/251759.251878>.
- Raharjana, Indra Kharisma, Daniel Siahaan, und Chastine Fatichah. 2021. „User Stories and Natural Language Processing: A Systematic Literature Review“. *IEEE Access* 9: 53811–53826. <https://doi.org/10.1109/ACCESS.2021.3070606>.
- Reddy, Deepa S. 2011. „Caught! The Predicaments of Ethnography in Collaboration“. In *Fieldwork is not what it used to be. Learning Anthropology's Method in a Time of Transition*, herausgegeben von James D. Faubion und George E. Marcus: 89–112. Cornell University Press. <https://doi.org/doi:10.7591/9780801463594-006>.
- RfII – Rat für Informationsinfrastrukturen. 2023. „Föderierte Dateninfrastrukturen für die wissenschaftliche Nutzung“. *RfII Berichte*. <https://rfii.de/download/rfii-bericht-foederierte-dateninfrastrukturen-fuer-die-wissenschaftliche-nutzung-maerz-2023/>.
- Robey, Daniel, Richard Welke, und Daniel Turk. 2001. „Traditional, Iterative, and Component-Based Development: A Social Analysis of Software Development Paradigms“. *Information Technology and Management* 2 (1): 53–70. <https://doi.org/10.1023/A:1009982704160>.
- Rojas, Luis, Cristian Olivares-Rodríguez, Claudio Alvarez, und Pedro G. Campos. 2022. „OurRank: A Software Requirements Prioritization Method Based on Qualitative Assessment and Cost-Benefit Prediction“. *IEEE Access* 10: 131772–131787. <https://doi.org/10.1109/ACCESS.2022.3230152>.
- Schöpfel, Joachim, und Hélène Prost. 2016. „Research Data Management in Social Sciences and Humanities: A Survey at the University of Lille (France)“. *Libreas: Library Ideas* 29: 98–112. <https://doi.org/10.18452/9097>.

- Shams-Ul-Arif, Qadeem Khan, und S. A. K. Gahyyur. 2010. „Requirements Engineering Processes, Tools/Technologies, & Methodologies“. *International Journal of Reviews in Computing*.
- Siadat, Seyed Hossein, und Minseok Song. 2012. „Understanding Requirement Engineering for Context-Aware Service-Based Applications“. *Journal of Software Engineering and Applications* 5: 536–544. <https://doi.org/10.4236/jsea.2012.58062>.
- Silvast, Antti, und Mikko J. Virtanen. 2019. „An assemblage of framings and tamings: multi-sited analysis of infrastructures as a methodology“. *Journal of Cultural Economy* 1 (6): 461–477. <https://doi.org/10.1080/17530350.2019.1646156>.
- Simukovic, Elena, Maxi Kindling, und Peter Schirmbacher. 2013. „Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin“. <http://dx.doi.org/10.18452/13568>.
- Sorrentino, Maddalena, und Francesco Virili. 2005. „Web Services System Development: A Grounded Theory Study“. *BLD 2005 Proceedings*. <https://aisel.aisnet.org/bled2005/51>.
- Struck, Alexander, Axel Loewe, Elke Achhammer, et al. 2020. „A Guide for Publishing, Using, and Licensing Research Software in Germany.“ <https://doi.org/10.5281/zenodo.4327148>.
- Strunk, Sonja. Evaluationsbericht 2018. „Bedarfserhebung Fachinformationsdienst (FID) Soziologie“. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-74925-5>.
- „The Anatomy of a User Story | Scrum Alliance | Includes Template“. o. J. Zuletzt abgerufen am 19. Juni 2025. <https://resources.scrumalliance.org/Article/anatomy-user-story>.
- Truschkat, Inga, Manuela Kaiser-Belz, und Vera Volkmann. 2011. „Theoretisches Sampling in Qualifikationsarbeiten: Die Grounded-Theory-Methodologie zwischen Programmatik und Forschungspraxis“. In *Grounded Theory Reader*, herausgegeben von Günter Mey und Katja Mruck: 353–379. VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93318-4_16.
- Vepřek, Libuše Hannah. 2024. „At the Edge of AI: Human Computation Systems and Their Intraverting Relations.“ transcript Verlag. <https://doi.org/10.14361/9783839472286>.
- Vicente-Saez, Ruben, und Clara Martinez-Fuentes. 2018. „Open Science now: A systematic literature review for an integrated definition“. *Journal of Business Research* 88: 428–436. <https://doi.org/10.1016/j.jbusres.2017.12.043>.
- Welz, Gisela. 2021. „Assemblage“. In *Theoretische Reflexionen. Perspektiven der Europäischen Ethnologie*, herausgegeben von Peter Hinrichs, Martina Röthl, und Manfred Seifert: 161–176. Dietrich Reimer Verlag.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. „The FAIR Guiding Principles for scientific data management and stewardship“. *Scientific Data* 3: 160018. <https://doi.org/10.1038/sdata.2016.18>.

Würfel, David, Rainer Lutz, und Stephan Diehl. 2016. „Grounded requirements engineering: An approach to use case driven requirements engineering“. *Journal of Systems and Software* 117: 645–657. <https://doi.org/10.1016/j.jss.2015.10.024>.