

Kerstin Bischoff (Forschungszentrum L3S Hannover), Heidemarie Hanekop (SOFI Göttingen), Kerstin Brückweh (Neuere Geschichte, Universität Trier) und Nicole Mayer-Ahuja (SOFI Göttingen und Fachbereich Sozialökonomie, Universität Hamburg)

Herausforderungen und Best Practices interdisziplinärer Kooperation in den eHumanities – Erfahrungen aus dem BMBF-Verbund „Gute Arbeit“ nach dem Boom

Die Chancen interdisziplinärer Kooperationen zwischen Informatik, Sozial- und Geisteswissenschaften sind bestechend. Der offene Austausch von Wissen und Methoden und die Bündelung komplementärer Kompetenzen können neue Forschungsansätze beflügeln; bspw. können mit neuen IT-Werkzeugen ganz neue Fragestellungen angegangen werden. In der Praxis allerdings sehen sich interdisziplinäre Projektverbünde, wie sie im Rahmen der Förderung der Digital Humanities entstehen, auch neuen Herausforderungen gegenübergestellt. Neben den durchaus erwarteten Verständigungsproblemen lauern schwerwiegende, oft auch zunächst unentdeckte Probleme, die zu Ängsten und Missverständnissen und folglich zu wechselseitigen Blockaden führen können.

Der vorgeschlagene Beitrag diskutiert Chancen und Herausforderungen der interdisziplinären Kooperation am Beispiel des BMBF-Projektes „Gute Arbeit“ nach dem Boom (Re-SozIT), an dem ArbeitssoziologInnen, ZeithistorikerInnen und InformatikerInnen beteiligt sind. Zunächst wird in Kürze das Verbundprojekt und Ergebnisse aus den ersten 1,5 Jahren vorgestellt. Der Verbund betritt gemeinsam Neuland bei der Entwicklung neuer Methoden der Längsschnittanalyse von qualitativen soziologischen Studien mit neuen IT-Werkzeugen. Im zweiten Teil werden die hierbei aufgetretenen Kooperationsprobleme und Herausforderungen analysiert. Gerade in der ersten Phase eines solchen Verbundprojektes werden Weichen gestellt, die über das spätere Scheitern oder den Erfolg entscheiden. Verbreitete Risiken sind z.B. das Aneinander-vorbei-entwickeln trotz bester Vorsätze oder die Nicht-Anwendung der von der Informatik entwickelten Tools durch die Sozial- und GeisteswissenschaftlerInnen oder wechselseitige Blockaden und Ineffizienz. Im dritten Teil werden erste Erfahrungen und Best Practices zur Erkennung und Überwindung solcher Hemmnisse vorgestellt.

"Gute Arbeit" nach dem Boom (ReSozIT) – ein Pilotprojekt zur Längsschnittanalyse arbeitssoziologischer Betriebsfallstudien mit neuen e-Humanities-Werkzeugen

Das Ziel des Verbundvorhabens "Re-SozIT" besteht in der IT-basierten Erschließung und Analyse qualitativer Forschungsdaten in einer Längsschnittperspektive: über 50 seit den 1970er Jahren durchgeführte arbeits- und industriesoziologische Forschungsprojekte des Soziologischen Forschungsinstituts Göttingen (SOFI) stehen zur Verfügung. Um das qualitative Material dieser Primärprojekte für themen- und zeitübergreifende Sekundäranalysen aktueller Fragestellungen aus Arbeitssoziologie und Zeitgeschichte auszuwerten, werden neuartige IT-Werkzeuge und Analyseverfahren entwickelt. Fragestellungen sind z.B. die Subjektivierung von Arbeit, „Gute Arbeit“ als Alltagspraxis (Soziologie) oder die Wahrnehmung von Arbeit im Zeichen von Arbeitslosigkeit (Geschichte).

Im Zentrum der interdisziplinären Arbeit standen bis dato das Datenmanagement, die prototypische Entwicklung von innovativen Suchwerkzeugen sowie die Anonymisierung.

- Eine zentrale Herausforderung ist der Aufbau eines geeigneten *Datenmanagements* inklusive *Metadaten*, die die Logik von heterogenen Primärprojekten hinreichend genau abbilden, um projektübergreifende Sekundäranalysen zu ermöglichen. Denn diese erfordern weitreichende Kenntnisse über Methoden und Design aber auch soziale Rahmenbedingungen der Primärprojekte. Ein wichtiges Ergebnis ist, dass cleveres Datenmanagement den Aufwand

für die Metadatenerfassung erheblich verringern kann (bspw. durch Vererbung von Informationen). Allerdings erweist sich die Festlegung von konkreten (Meta-)Datenstrukturen als sehr kritischer Punkt zwischen den Forschern aus den unterschiedlichen Disziplinen.

- Die Sekundäranalyse qualitativer Daten kann gerade im Hinblick auf das Finden von interessanten Primärdaten von modernen IT-basierten *Such- und Erschließungswerkzeugen* profitieren. Unser Material von 350.000 Seiten digitalisiertem Material aus 6000 qualitativen Interviews ist eine manuell nicht zu bewältigende Menge. Mittels informationstechnologischer Ansätze wie Topic Modelling, Sentiment-Analyse, etc. können (semi-)automatisch Themen, Terme und Meinungen extrahiert und u.a. zur Metadatenanreicherung genutzt werden, um über die Volltextsuche hinaus eine bessere Durchsuchbarkeit zu gewährleisten.
- Um das Spannungsverhältnis zwischen der notwendigen *Anonymisierung* personenbezogener Daten und dem Interesse an aussagekräftigen Forschungsdaten zu bewältigen wurde ein Konzept erstellt, welches auf den Bausteinen Risikoklassifizierung, IT-gestützte Anonymisierungsmaßnahmen und infrastrukturelle Rahmenbedingungen mit abgestuften Zugangsrechten beruht.

Für die Sekundäranalyse der soziologischen und zeitgeschichtlichen Teilprojekte ergeben sich völlig neue Analysemöglichkeiten und für die Informatik ergeben sich neuen Möglichkeiten des Zugangs zu wertvollen Datenmaterialien zur Erprobung ihrer Algorithmen.

Herausforderungen der interdisziplinären Kooperation

Zusätzlich zu den durchaus erwarteten Kommunikations- und Verständigungsproblemen ergaben sich einige unerwartete, zunächst verdeckte Herausforderungen, die u.a. mit wechselseitiger Rollenzuschreibung zu tun hatten oder mit unterschiedlichen und unbekanntem Agenden und Interessen – bei gleichzeitiger gegenseitiger Abhängigkeit bei notwendigen Arbeitsschritten. Diese führen zu Koordinierungsproblemen, Blockaden und Spannungen. Dabei bilden sich je nach Thematik unterschiedliche ‚Fronten‘.

- Typisch für *Kommunikations- und Verständigungsprobleme* sind begriffliche Differenzen, die sogar verschärft werden, wenn teils gleiche Begriffe mit unterschiedlicher Bedeutung verwendet werden. Dies betrifft u.a. die Benennung von Schaltflächen in den Tools.
- *Mangelndes Verständnis* der unterschiedlichen Arbeitsweise und Methoden, des Selbstverständnisses und der Forschungstradition der anderen Disziplin(en) erschwert die Kooperation. Es treffen hier teils gänzlich verschiedene Denkschulen aufeinander. Während die Geschichts- und qualitativen SozialwissenschaftlerInnen eher Wert auf die Entfaltung von Argumentationslinien und elaborierten Diskurs legen, werden InformatikerInnen auf klare, prüfbar, validierbare und messbare Festlegungen drängen. Es soll eindeutig spezifiziert werden, unter welchen Ausgangsbedingungen sich das System bei bestimmten Benutzerinteraktionen wie verhalten soll.
- Die Gleichzeitigkeit von methodischer und inhaltlicher Innovation (Soziologie und Zeitgeschichte) und Entwicklung neuer Tools und Werkzeuge (Informatik) führt zu *wechselseitiger zeitlicher Abhängigkeit* und nur schwer zu erfüllenden gegenseitigen Erwartungen. Strukturell problematisch ist nun, dass die Anforderungen an IT-Tools für die Sekundäranalyse daher noch teils unbekannt sind. Ein zu starkes Drängen oder „Erzwingen“ wollen einer deutlichen Strukturierung mit festhaltbaren Ergebnissen kann u.U. dann zu Übervereinfachung führen. Dabei befindet sich in unserem Projekt die Informatik in einer Sandwich-Position zwischen den Anforderungen der Herkunftswissenschaft und potentiellen Sekundärwissenschaften, die wiederum recht anders gelagerte Informationsbedürfnisse haben können.
- *Gegensätzliche Anforderungen* der Disziplinen zeigen sich auch in punkto Anonymisierung: Während die ArbeitssoziologInnen als Datengeber weitreichende Anonymisierungsmaßnahmen zusammen mit Auflagen für die Nutzung befürworten, ist für die

Geschichte die absolute Nachvollziehbarkeit der Quelle unabdingbar. Die Informatik als dritter Spieler, sieht hier die schwierige Machbarkeit von zuverlässigen IT-basierten Anonymisierungsmaßnahmen. Auch kann die Offenlegung von Forschungsdaten auf Bedenken der Datenerheber hinsichtlich möglicher Kritik stoßen. Bspw. gibt es aktuell rege zeithistorische Auseinandersetzungen, die sich um die Nutzbarkeit von sozialwissenschaftlichen Daten als Quelle oder/und als Darstellung bzw. Fakten drehen.

- Versteckte oder zumindest unbekannte Agenden und mangelndes Verständnis bzw. Wertschätzung (Stichwort Fachwissenschaftler vs. Programmierer) zwischen den unterschiedlichen Disziplinen behindern die effiziente Kooperation, solange sie nicht aufgedeckt, diskutiert und ausgehandelt werden. Hier geht es letztlich um die Hegemonie zwischen den Disziplinen: wer setzt die *Forschungsagenda* bzw. -vorgehensweise fest.
- Die Befürchtung einer Art feindlicher Übernahme durch die Informatik bei Sozial- und Geisteswissenschaftlern spiegelt sich u.a. in einer latenten *Angst vor der Automatisierung*, dem nur bedingt vorhandenen Vertrauen in die Technologien und dem Unbehagen aufgrund einer potentiellen Entmündigung durch die IT. Diese Skepsis zeigt sich insbesondere anhand der Frage, ob Strategien wie *text mining* oder *opinion mining* Such- oder Analysefunktionen beinhalten – eine Frage, die nicht zuletzt die Selbstverständnisse der beteiligten Geistes- und Sozialwissenschaften berührt: welche Rolle spielen aggregierte Daten in vorliegendem Material, können sie als „repräsentativ“ oder auch nur „exemplarisch“ gelten? Inwiefern wird die Forschungsstrategie der Primärstudien in diesen Strategien „aufgehoben“ oder „gelöscht“?
- *Divergierende Interessen* speisen sich aus dem Spannungsverhältnis zwischen den jeweils eigenen Aufgaben und Zielen der Fachwissenschaften und Kooperationsanforderungen. Zumeist sind hier Doppelanforderungen zu erfüllen, denn zeitaufwändige Arbeiten ‚nur‘ für das Projekt bzw. die andere Disziplin, für die es aber keine Meriten in der Herkunftsdisziplin gibt, können zu Störfaktoren werden. So bedeutet bspw. die Evaluierung von Software-Prototypen Aufwand für die GeisteswissenschaftlerInnen und SoziologInnen; das Nachprogrammieren von ‚trivialen‘ Features, u.U. inspiriert durch kommerzielle Programme, ist für die Informatikforschung uninteressant.

Best Practices – Erste Erfahrungen im Zueinanderfinden

Im Laufe der ersten Projekthälfte haben wir erste Lernprozesse durchlaufen. Grundsätzlich sollte die interdisziplinäre Zusammenarbeit als eigenständige Aufgabe ernstgenommen und regelmäßig evaluiert werden. Dafür ist es empfehlenswert, sich viel Zeit zum gegenseitigen Kennenlernen zu nehmen sowie eine Kultur der Offenheit und Wertschätzung zu etablieren. Das betrifft die Vorstellung und kritische Reflektion der eigenen Arbeiten, Forschungsinteressen und Projektziele, der Forschungstradition und des Selbstverständnisses der eigenen Disziplin inklusive Spielregeln innerhalb der Community sowie zentrale (aktuelle) Methoden und Arbeitsweisen. Nur auf diese Weise scheint es möglich, besser die jeweiligen Motive nachzuvollziehen, Arbeit anzuerkennen, etc.

Kompromisse und Mittelwege zu finden ist gerade in den frühen Phasen der Zusammenarbeit wichtig. Statt dass also jede Disziplin forschungsmäßig auf ihrem (hegemonialen) Maximalanspruch besteht und somit ein effektives gemeinsames Vorankommen verhindert, können erste – wenn auch u.U. technisch, sozialwissenschaftlich oder geschichtswissenschaftlich etwas weniger herausfordernde – Fragestellungen ein guter Ausgangspunkt sein, von dem aus sich dann neuer, tatsächlich interdisziplinärer Forschungsbedarf aufzeigt. In unserem Falle hat sich so eine interessante Fragestellung zum Zeitvergleich via Topic Modelling herauskristallisiert, welche konzeptionell und auch im Hinblick auf die Umsetzung weder für die ZeithistorikerInnen noch für die Informatik trivial ist.

Die Auswahl und *Anpassung der jeweils eigenen disziplinspezifischen Methoden und Vorgehensweisen* sollte soweit als möglich an den Kooperationspartnern und dem Kontext ausgerichtet

werden – mit dem Ziel gemeinsam eine neue Vorgehensweise zu erproben und ggf. neue interdisziplinäre Methoden abzuleiten. So verfolgen wir bspw. eine sehr agile Vorgehensweise der Software-Entwicklung mit vielen Iterationen, in der durchaus noch unvollständige und fehlerhafte Forschungsprototypen schnell ‚bespielbar‘ sind. Hier besteht die Gefahr, zwischen den Zyklen ‚umsonst‘ vorgearbeitet zu haben. Gerade unter dem Gesichtspunkt der teils unbekanntem Anforderungen an die mit IT-Tools zu bewältigende Aufgabe bietet ein derartiges Vorgehen jedoch die Möglichkeit, nach und nach besser die Potentiale und Unzulänglichkeiten von Techniken kennenzulernen und die ‚Fantasie‘ für weitere Unterstützungsmöglichkeiten und Anforderungen anzuregen. Ausführliche Diskussionen zeigten dabei die Notwendigkeit der Anpassung von generellen IT-Methoden an die Bedarfe und Arbeitsweisen der jeweiligen Nutzungsdisziplin.

Statt Angst vor feindlicher Übernahme durch die Informatik als schwarzer Ritter auf der einen Seite bzw. die Sorge vor Giftpillen als Versuch der (vermeintlich notwendigen) Gegenwehr auf der anderen Seite sollten sich die Digitalen Geisteswissenschaften also besser als ein gemeinsames Wagnis verstehen, das für alle beteiligten Disziplinen eine Bereicherung bieten und neue Forschungsansätze hervorbringen kann.