



SoNAR (IDH) AP4-4 Evaluierung III: Analyse des Forschungsprozesses von HNA-Expert:innen und sich daraus ergebende Bedürfnisse an eine Infrastrukturlösung

Sandra Balck, Sina Menzel und Vivien Petras

Humboldt-Universität zu Berlin

- Version 2.0: Juli 2021-

Inhalt

1 Einleitung	3
1.1 Fragestellungen der Interviewstudie	3
1.2 Methodik	3
2 Historische Netzwerkanalyse	4
2.1 Definition	4
2.2 Community	5
2.2.1 HNA Infrastruktur	5
2.2.2 Positionierung in der historischen Forschungslandschaft	6
3 Forschungsprozess	7
3.1 Entwicklung von Forschungsfragen	8
3.2 Datenauswahl und -erhebung	8
3.2.1 Historisches Quellenmaterial	9
3.2.2 Datentypen	9
3.2.2.1 Bibliographische Daten	9
3.2.2.2 Personendaten im Kontext	10
3.2.2.3 Informelle Daten	10
3.2.3 Erhebungsmethoden	11
3.3 Qualitätssicherung	12
3.3.1 Statistische Berechnung	12
3.3.2 Intellektuelle Durchsicht	13
3.3.3 Visualisierung	13
3.4 Datenanalyse	13
3.4.1 Statistische Methoden	13
3.4.2 Visualisierung als Datenanalyse-Instrument	14
3.5 Dokumentation	15
3.5.1 Dokumentation des Forschungsprozesses	15
3.5.2 Dokumentation der Qualitätssicherung	15
3.6 Datenaustausch und Datenpublikation	16
4 Werkzeuge	16
4.1 Verwendete Werkzeuge und Funktionen	16
4.2 Anforderungen an Funktionalitäten	18
4.2.1 Datenexport und -import	18
4.2.2 Datenaufbereitung	18
4.2.3 Visualisierung und Netzwerkerkundungen	18
4.2.4 Dokumentation und Datenpublikation	19

5 Problematik Datenqualität	19
5.1 Probleme mit der Datenqualität	19
3.1.1 Ungenauigkeit und Ambiguität	19
5.1.2 Informationslücken und Unvollständigkeit	20
5.2 Anforderungen an die Datenqualität	20
5.2.1 Transparenz durch Dokumentation	20
5.2.2 Standards und Identifikatoren	21
5.3 Historische Daten als Herausforderung für Visualisierungen	21
6 Erkenntnisse	22
6.1 Anforderungen an die Infrastruktur	22
6.2 Validierung des modellhaften Forschungsdesigns	23
6.3 Fazit	23
Anhang	24
Interviewleitfaden	24
Interviewteilnehmer:innen	29
Codebuch	30

1 Einleitung

Ziel der vorliegenden Studie ist es, die Ergebnisse des Arbeitspaketes AP2-1 zu evaluieren. Das durch die historisch arbeitenden Fachwissenschaftler:innen entwickelte modellhafte Forschungsdesign¹ zur Nutzung der Forschungstechnologie SoNAR (IDH) beschreibt Forschungsprozesse und -fragen der Historischen Netzwerkanalyse, wie diese methodisch bearbeitet werden und welche speziellen Anforderungen sich daraus ergeben. Die dazu durchgeführte qualitative Erhebung beantwortet Fragen nach dessen Validität und ermöglicht Rückschlüsse auf Bedarf und Umfeld für die Forschungstechnologie. Dafür wurden Fachwissenschaftler:innen zum Vorgehen bei eigenen Forschungsarbeiten im Bereich der HNA in Form qualitativer Leitfadeninterviews befragt. Die Auswertung der Interviews gibt einen Überblick über exemplarische Forschungsprozesse und deren Aufbereitung in einzelne Teilschritte sowie erste Erkenntnisse zu Anforderungen an Visualisierungen für die Historische Netzwerkanalyse (HNA) in verschiedenen Disziplinen. Sie validiert das vorliegende modellhafte Forschungsdesign und konkretisiert Anforderungen an die Infrastruktur.

1.1 Fragestellungen der Interviewstudie

1. Wie sieht der Forschungsprozess von Expert:innen auf dem Gebiet der HNA aus, insbesondere in Bezug auf die Datenerhebung und -analyse?
2. Welche Konsequenzen für SoNAR ergeben sich daraus?

1.2 Methodik

Es wurden insgesamt 7 Interviews mit Expert:innen der HNA durchgeführt. Von den sieben Interviewten waren zwei männlich und fünf weiblich. Die Interviewten kamen aus den Bereichen Sozialwissenschaft, Geschichtswissenschaft und Wirtschaftswissenschaft (siehe Anhang Interviewteilnehmer:innen). Die Interviews wurden im Zeitraum 03.07.-05.08.20 durchgeführt. Das kürzeste Interview dauerte 37:08 Min., das längste 51:28 Min.

Als Expert:innen für diesen Kontext gelten Personen, die wissenschaftliche Arbeiten im Bereich HNA vorweisen. Die Akquise erfolgte über Kontaktvermittlung durch die Projektpartner (SBB, HHU, FHP), über das eigene Netzwerk sowie durch die Sichtung geeigneter HNA-Publikationen und die damit verknüpften Publikationsnetzwerke.

Im Vorfeld wurden Erhebungen des Visualisierungsworkshops von AP3 systematisch ausgewertet. Zusätzlich fand ein engmaschiger Austausch mit den Projektpartner:innen statt; hierbei wurden geäußerte fachwissenschaftliche Anforderungen festgehalten und zur konkreten Studienvorbereitung genutzt. Die Ergebnisse dienten der Konzipierung des Leitfadens (siehe Anhang Interviewleitfaden). Um die Validität des Leitfadens sicherzustellen, fand ein Pretest statt, welcher ohne Auffälligkeiten verlief.

Die Befragungen wurden über die Videokonferenzlösung Zoom durchgeführt, mit Einverständnis der Proband:innen aufgezeichnet und anschließend mittels MAXQDA 2020 transkribiert und ausgewertet.

¹ Link zum Dokument:

<https://onedrive.live.com/edit.aspx?cid=3b9129e4b3c7c3c9&page=view&resid=3B9129E4B3C7C3C9!1029&parId=3B9129E4B3C7C3C9!986&app=Word>

Zusätzlich fand eine Kurzprotokollierung der Sitzungen statt, in denen Hinweise der Interviewerin aufgenommen wurden.

Die Codierung erfolgte ebenfalls mit MAXQDA 2020. Es folgte eine thematische Analyse, deren Ziel es war, induktiv Themen und Muster aus dem Datenmaterial zu identifizieren. Gebildete Kategorien dienten als Grundlage für die Auswertung sowie in Teilen der Strukturierung des vorliegenden Dokuments. Insgesamt wurden 126 Kategorien gebildet, die in 6 Oberkategorien zusammengefasst wurden (die Unterkategorien können dem Codebuch entnommen werden - siehe Anhang):

- Community: Disziplinäres Selbstverständnis bezogen auf wissenschaftliche und disziplinäre Infrastrukturen, Wissenschaftskommunikation sowie die Einbindung digitaler Methoden in historische Curricula.
- Netzwerke: Einschätzungen über die Verbindung der Netzwerkanalyse mit der historischen Forschung im allgemeinen und Definitionsversuche/Einordnungen der HNA durch die Interviewten.
- Forschungsprozess: Äußerungen zum Vorgehen bei der Bearbeitung eines Forschungsvorhabens mittels historischer Netzwerkanalyse.
- Werkzeuge: Für die Datenaufbereitung- und -speicherung verwendete Software, ihre Vor- und Nachteile sowie Anforderungen an zukünftige Systeme.
- Datenqualität: Datenqualität ist als besondere Herausforderung der HNA zu verstehen, der Fokus liegt hier auf Anforderungen und Probleme bezogen auf historische Daten.
- Visualisierung: Bezieht sich auf die Anwendung von Visualisierungen und dem daraus resultierenden Erkenntnisgewinn sowie Kritik und Herausforderungen im Umgang mit historischen Daten.

2 Historische Netzwerkanalyse

2.1 Definition

„Die Geschichtswissenschaft beschäftigt sich mit der Analyse des menschlichen Zusammenlebens in der Vergangenheit und bemüht sich, Ereignisse und Entwicklungen aus der jeweiligen Zeit heraus zu verstehen und zu deuten. Gefragt wird nicht nur nach den Handlungen einer Person, sondern auch nach deren Handlungsspielraum unter Berücksichtigung der jeweiligen politischen, ökonomischen und sozialen Umstände. Historische Akteure werden demnach immer auch als Kontext der sie umgebenden Strukturen betrachtet.“² Düring und Keyserlingk argumentieren weiter, dass sich viele dieser „Strukturen in sozialen Beziehungen manifestieren.“³

An dieser Stelle setzt die historische Netzwerkanalyse an, sie wird verwendet, um durch die Gewichtung von Beziehungen, Strukturen innerhalb eines Netzwerks sichtbar zu machen. In den Interviews wird es beschrieben als „Relationen zwischen Leuten zu verstehen, zu visualisieren, zu analysieren, die vor langer Zeit gelebt haben.“ (Person 4 Abs. 4)

Lemercier betont zudem den unbewussten Charakter dieser Strukturen und die Bedeutung der Netzwerkanalyse bei der Erforschung dieser: „Formale Netzwerkanalyse erlaubt uns, Strukturen zu entdecken,

² Düring, Marten & von Keyserlingk, Linda: Netzwerkanalyse in den Geschichtswissenschaften. Historische Netzwerkanalyse als Methode für die Erforschung von historischen Prozessen. In: Schützeichel, Rainer & Jordan, Stefan (Hrsg.): Prozesse - Formen, Dynamiken, Erklärungen. Springer, 2015, S. 337

³ ebenda

die nicht von allen betroffenen Akteuren erkannt werden, aber deren Form uns über zugrundeliegende soziale Mechanismen unterrichtet. [...] Der Punkt formaler Netzwerkanalyse ist es [...] ihre Muster präzise zu beschreiben, zu verstehen, wie sie geschaffen wurden und welche Folgen sie haben. Netzwerke sind nirgends und überall, da ja praktisch alles als eine Menge von Beziehungen beschrieben werden kann.”⁴

Die Netzwerkanalyse dient dabei als “[...] das formale Instrumentarium, das einem dann erlaubt Schlüsse zu ziehen, die eben nicht vielleicht offensichtlich sind [...]” (Person 5). Auch Düring und von Keyserlingk beschreiben die HNA als eine Methode, die bei der Interpretation hilft – verschiedene Hinweise und Ansatzpunkte liefert, Ergebnisse aber nicht „sui generis“ produziert.⁵ Es ist demnach vielmehr ein Zusammenspiel aus formaler Netzwerkanalyse und historischer Erzählung.

Die Netzwerkanalyse dient also der Sichtbarmachung von nicht immer offensichtlichen sozialen Strukturen und Verbindungen mit Hilfe einer systematischen Auswertung von verschiedenen Quellentypen, wobei besonders die Aufbereitung der Daten für die systematische Auswertung – die Erhebung und Modellierung diverser Datenbestände (vgl. Person 2 Abs. 2) von Bedeutung ist. Die Interpretation oder Analyse der Daten ist dabei immer eine historische und weist demnach auch Lücken auf: “[...] allgemein kann man ja nur einen Bruchteil von dem erfassen in der Datenbank, im Netzwerk, was in der Realität gewesen ist.” (Person 6 Abs. 67) Es entsteht ein Abbild der Vergangenheit auf Grundlage der erhobenen und zur Verfügung stehenden Daten, welches Verbindungen sichtbar macht und die Analyse unterstützt, jedoch keinen Anspruch auf Vollständigkeit haben kann.

Die Darstellung der Daten in ihrem jeweiligen historischen Kontext ist besonders wichtig, um die Interpretation dieser zu ermöglichen - Kontextualisierung ist Grundlage der historischen Netzwerkanalyse.

2.2 Community

Folgende Ausführungen beziehen sich auf die Community der historischen Netzwerkforschung und der Digital Humanities. Beschrieben werden disziplinäre Infrastrukturaspekte, wie Publikation, Wissenskommunikation und der Ausbau technischer Kenntnisse. Dies wiederum geht einher mit der Positionierung der HNA in der historischen Forschungslandschaft, welche im zweiten Abschnitt thematisiert wird.

2.2.1 HNA Infrastruktur

Von den Interviewten wird eine Stärkung der HNA Infrastruktur gefordert. Gemeint ist die wissenschaftliche/disziplinäre Infrastruktur, die die Kommunikation von Forschungsergebnissen innerhalb der HNA bzw. den Geschichtswissenschaften erleichtert. Die Transparenz der eigenen Arbeit, die Publikation von Daten, aber auch grundlegende Fragen wie ein einheitliches Verständnis der Methode sind Themen, die erwähnt werden. Als Hilfestellung können dafür spezifische Werkzeuge oder methodische Tutorials dienen. (vgl. Person 2 Abs. 18) Auch Veranstaltungen sind Teil einer wissenschaftlichen Infrastruktur und werden als wichtige Plattform für den Austausch über die eigene Arbeit und zum Entdecken neuer Werkzeuge genannt. Dazu gehören Workshops, Tagungen und Summer Schools.

⁴ Lemercier, Claire: Formale Methoden der Netzwerkanalyse in den Geschichtswissenschaften: Warum und Wie?. *Österreichische Zeitschrift für Geschichtswissenschaften*, 23(1), 2012, S. 21

⁵ vgl. Düring, Marten & von Keyserlingk, Linda: Netzwerkanalyse in den Geschichtswissenschaften. Historische Netzwerkanalyse als Methode für die Erforschung von historischen Prozessen. In: Schützeichel, Rainer & Jordan, Stefan (Hrsg.): Prozesse - Formen, Dynamiken, Erklärungen. Springer, 2015, S. 347-348

In Studiengängen der klassischen Geschichtswissenschaft werden keine technischen Kenntnisse vermittelt und damit der Zugang zu digitalen Methoden erschwert. Diese gewinnen jedoch immer mehr an Bedeutung: „[...] da liegt auch gerade noch die große Diskrepanz eigentlich dazwischen, dass man irgendwie sobald es ins Doktorat geht oder auch sonst irgendwelches wissenschaftliche Arbeiten müssen die digitalen Kompetenzen da sein, aber im Studium ist man - zumindest hier noch - relativ traditionell unterwegs, das bricht erst sehr sehr langsam auf.“ (Person 7 Abs. 24) Es ist jedoch ein steigendes Interesse zu erkennen. So berichten drei der Interviewten, selbst Kurse mit diesem Schwerpunkt anzubieten: „Also je mehr Leuten ich natürlich beibringe, wie es funktioniert und was man damit machen kann, hoffe ich halt auch, dass, ja, diese ganz klassische Geschichtswissenschaft doch vielleicht irgendwann ein bisschen offener wird für neue Ansätze [...].“ (Person 3 Abs. 39) Besonders eine bessere Verfügbarkeit von Daten in Bibliothekskatalogen oder Datenbanken könnte die Einbindung digitaler Methoden in die Lehre erleichtern. Das Fehlen von guten Einstiegspublikationen wird zudem als Problem genannt. (vgl. Person 3 Abs. 37)

Es wird ein allgemeines Interesse bei Kolleg:innen der historischen Netzwerkforschung beschrieben, sich in Themen der „digital history“ einzuarbeiten, was jedoch häufig an fehlenden Strukturen scheitert – personelle sowie finanzielle Ressourcen in kleineren Projekten werden als Hauptproblem genannt. (vgl. Person 2 Abs. 18 und Person 7 Abs. 28) „Und deswegen [...] find ich das sehr spannend und sehr wichtig, dass da auch so eine Infrastruktur überhaupt geschaffen wird erstmal, um dann weitere Forschungen überhaupt möglich zu machen.“ (Person 7 Abs. 28)

2.2.2 Positionierung in der historischen Forschungslandschaft

Aufgrund des fehlenden technischen Verständnisses kommt es zu Problemen bei der Wissensvermittlung von Forschungsergebnissen der HNA. In kunsthistorischen Zeitschriften ist es demnach schwieriger, Analyseschritte zu beschreiben als beispielsweise in Zeitschriften der historischen Netzwerkforschung, da die interdisziplinäre Übersetzung schwierig ist. Es wird sogar von einem disziplinären Konkurrenzempfinden gesprochen „[...] ich glaube auch, dass die historische Forschung da noch einiges dazulernen muss, auch ihre Themen und Fragestellungen zu erweitern. Und ich glaube auch dafür ist eben gerade die Netzwerkanalyse sehr sehr hilfreich, um auch klar zu machen, dass es über das traditionelle Arbeiten hinaus sehr sehr spannende Ansätze gibt, die eben keine Konkurrenz darstellen, sondern eine Weiterentwicklung, Ergänzung und das sicherlich gut ist.“ Kritisiert wird in diesem Zusammenhang eine fehlende Offenheit der Geschichtswissenschaft: „[...] was noch so ein bisschen fehlt meiner Meinung nach, was die historische Netzwerkanalyse angeht, ist so ein Verständnis der Forscher, also sich auch mal auf Zahlen einzulassen, eben auch mal sich auf diese statistischen Auswertungen eben auch mal einzulassen.“ Diese fehlende Offenheit führt teilweise dazu, dass der Arbeitsaufwand nicht anerkannt wird: „Und eben auch auf der anderen Seite [...] gibt es immer sehr viele, die sagen "oh so ein schönes Bild, ist ja toll", aber die Arbeit, die dahinter steht überhaupt nicht nachvollziehen können.“ (Person 3 Abs. 37) Betont wird deshalb, wie wichtig es ist, den wissenschaftlichen Nachwuchs mit den digitalen Methoden vertraut zu machen.

3 Forschungsprozess

Beschrieben wird das Vorgehen bei der Bearbeitung eines Forschungsvorhabens, bezogen auf die historische Netzwerkanalyse. Die historische Netzwerkanalyse setzt sich, wie zuvor beschrieben, aus Komponenten der Sozialwissenschaft und Geschichtswissenschaft und damit verbunden den Digital Humanities zusammen. Der

Forschungsprozess dieser Disziplinen ähnelt sich dabei in vielen Bereichen. Bhattacharjee (2012)⁶ unterteilt den sozialwissenschaftlichen Forschungsprozess in drei Phasen, welche wiederum in neun Unterphasen aufgeteilt werden. (siehe Tab. 1 Bhattacharjee) Zu ähnlichen Schlüssen kommen Bron et al.⁷ in ihrer 2016 veröffentlichten Studie zum Forschungsverhalten von Medienwissenschaftler:innen - sie unterteilen ebenfalls in drei Phasen. (siehe Tab. 1 Bron et al.) Im Gegensatz zu Bhattacharjee fassen Bron et al. die Theorieentwicklung und das Forschungsdesign unter der Kategorie “initial exploration of a topic” zusammen und fokussieren stärker auf Datenerhebung und -analyse und den Prozess des Schreibens und Publizierens. Im Vergleich dazu nennen Burghardt et al.⁸ für die Geisteswissenschaften acht Phasen. (siehe Tab. 1 Burghardt et al.) Ausgenommen von “Kommunikation und Kollaboration”, was im weitesten Sinne auch unter Theoriebildung gefasst werden kann, sind keine Unterschiede zum sozial- oder medienwissenschaftlichen Forschungsprozess zu erkennen.

Bhattacharjee (2012) Sozialwissenschaften	Bron et al. (2016) Medienwissenschaft	Burghardt et al. (2015) Geisteswissenschaften
<ul style="list-style-type: none"> • exploration (research question, literature review, theory) • research design (operationalization, research method, sampling strategy) • research execution (pilot testing, data collection, data analysis) + research report 	<ul style="list-style-type: none"> • Initial exploration of a topic • targeted information gathering and analysis • writing and reporting 	<ul style="list-style-type: none"> • Kommunikation und Kollaboration • Recherche • Konzeptualisierung • Datenerhebung • Datenaufbereitung • Datenauswertung • Verschriftlichung • Veröffentlichung

Tab. 1: Modelle des Forschungsprozesses in unterschiedlichen Disziplinen

Der im Folgenden beschriebene Forschungsprozess der historischen Netzwerkforscher:innen kann insgesamt in sechs Phasen unterteilt werden, wobei Datenauswahl- und erhebung sowie Datenaustausch- und publikation zusammengefasst wurden:

1. Entwicklung von Forschungsfragen / Theoriebildung
2. Datenauswahl und -erhebung
3. Qualitätssicherung
4. Datenanalyse
5. Dokumentation
6. Datenaustausch und -publikation

⁶ vgl. Bhattacharjee, Anol (2012). "Social Science Research: Principles, Methods, and Practices". *Textbooks Collection*. 3.

⁷ vgl. Bron, Marc, van Gorp, Jasmijn & de Rijke, Maarten: Media Studies Research in the Data-Driven Age: How Research Questions Evolve." *Journal of the Association for Information Science and Technology*, 2016, <https://doi.org/10.1002/asi.23458>

⁸ vgl. Burghardt, Manuel, Wolff, Christian & Womser-Hacker, Christa: Informationsinfrastruktur und informationswissenschaftliche Methoden in den digitalen Geisteswissenschaften. *Information - Wissenschaft & Praxis* (66), 2015, 285-286. 10.1515/iwp-2015-0052.

Es lassen sich viele Parallelen zu den schon beschriebenen allgemeinen Forschungsprozessen der Geistes- sowie Sozialwissenschaften erkennen. Im Unterschied zu den anderen Studien sind zwei zusätzliche Prozessabschnitte auszumachen: Qualitätssicherung und Dokumentation. Diese zwei Phasen des Forschungsprozesses sind stark miteinander verbunden und weisen auf die Wichtigkeit von Transparenz innerhalb des Forschungsprozesses der HNA sowie die Schwierigkeiten im Umgang mit historischen Daten bezogen auf die Datenqualität hin.

3.1 Entwicklung von Forschungsfragen

Der Forschungsprozess beginnt mit der Theorieentwicklung. Die Forschungsfragen werden sowohl theorie- als auch datengeleitet entwickelt. Am verbreitetsten ist das theoriegeleitete Vorgehen, bei welchem die Datenerhebung an einer bestehenden Forschungsfrage ausgerichtet wird: „[...] ich habe gewisse Hypothesen und alle Paper sind so aufgebaut, dass ich aus Theorien - bestehenden Theorien, die ich vielleicht etwas weiterentwickelt oder modifiziert habe - Hypothesen ableite und dann auch theoriegeleitet meine statistischen Analysen aufbaue.“ (Person 5 Abs. 8) Im Laufe des Forschungsprozesses können sich Forschungsfragen durch neue Impulse durch die Datenanalyse verändern: „Also es ist durchaus ein Wechselspiel, sodass nicht alle Fragen von vornherein feststehen, sondern dass die Netzwerkanalyse oder die Ergebnisse von ersten Netzwerkanalysen - das ist genauso mit dem Topic Modeling, da ist es fast noch extremer - natürlich auch neue Fragen generieren.“ (Person 3 Abs. 24) Es ist also eine Offenheit für die prozesshafte Weiterentwicklung von Forschungsfragen zu erkennen, welche sogar als eine der Stärken der historischen Netzwerkanalyse beschrieben wird.

Forschungsfragen können auch komplett im Prozess entstehen, wobei eine anfängliche Idee durch methodische Überlegungen und verschiedene Werkzeuge kontinuierlich weiterentwickelt wird. (vgl. Person 2 Abs. 10) Beim datengeleiteten Vorgehen wird ausgehend von einem bestehenden Datenbestand eine darauf passende Forschungsfrage entwickelt. Die Daten wurden in den jeweiligen Beispielen vor Projektbeginn erhoben und sollten für die Beantwortung beliebiger Fragen verwendet werden. „Also so habe ich angefangen, praktisch mit dem Preprocessing die Daten vorzubereiten, hab mir dann natürlich aber auch parallel überlegt: Ja, was will ich mit den Daten eigentlich - was für Forschungsfragen will ich beantworten?“ (Person 4 Abs. 8)

3.2 Datenauswahl und -erhebung

Die Datenauswahl ist abhängig von der Fragestellung sowie von der Verfügbarkeit der Daten. Die Suche wird entsprechend der Forschungsfrage thematisch eingegrenzt, wobei teilweise auch abweichende Daten mit aufgenommen werden, so zum Beispiel Personen, die immer wieder auftauchen, aber keinen direkten Bezug zum Thema haben. (vgl. Person 1) Da die Themen der historischen Netzwerkanalyse in einigen Fällen sehr spezifisch sind, wird dazu tendiert, auch Daten zu angrenzenden Themen mit aufzunehmen: „Ich glaube, wenn man nur die auswählen würde, wo es wirklich ums Thema geht, wär es recht wenig; deswegen würde es schon Sinn machen, wenn man einfach das, sagen wir, das ganze Netz - das Wort Gesamtnetzwerk ist immer schlecht zu gebrauchen, aber möglichst viele - erfasst und sich dann anschaut, was man überhaupt damit machen kann.“ (Person 6 Abs. 57)

Die Datenerhebung ist immer auch abhängig vom Grad der Digitalisierung: „Zum einen - und ich glaube deswegen begrüße ich auch Ihr Projekt sehr, weil ich genau merke, dass genau dieser Punkt fehlt - in den österreichischen Archiven ist quasi nichts digitalisiert, Großteil der Archive ist nicht mal der Datenbestand online, also dass man irgendwelche Findbücher hat, sondern man muss vor Ort per Zettelkatalog [suchen] und das ist dann natürlich ein wahnsinniger Aufwand sich überhaupt einen Datensatz zu bauen, wenn man das so möchte.“ (Person 7 Abs. 4) Kataloge von deutschen und schweizerischen Archiven werden hingegen als sehr gut

beschrieben, konkret genannt werden Staatsarchive und das Bundesarchiv. (vgl. Person 6 Abs. 49)

3.2.1 Historisches Quellenmaterial

Für die Erhebung konsultierte Institutionen sind Archive, Bibliotheken, Verlage und Vereine (sowie deren Archive). Dementsprechend ist ein Schwerpunkt in der Datenerhebung die Arbeit mit analogen Quellen, wobei vorrangig folgende von Relevanz sind: Protokolle, Tagesordnungen, Personenlisten, Biografien und Zeitschriften. Bei der Datenauswahl werden verschiedene Quellentypen zu unterschiedlichen Zwecken berücksichtigt. „Grundinformationen stammen immer aus den Zeitschriften selbst, [...] alles was bibliographische Recherche ist, ist dann mit den Quellen verzeichnet, also gerade zu den Personen, wenn wir Lebensdaten, Lebenswege, persönliche Beziehungen rekonstruieren, das kriegen wir natürlich nicht aus unseren Zeitschriften, da ist das mit den Quellen belegt.“ (Person 3 Abs. 20)

„Also als Historiker geht man ins Archiv. In ein physisches Archiv.“ (Person 6 Abs. 45) Die Recherche und Datenerhebung von archivischen Quellen ist ein langwieriger Prozess, der von verschiedenen Faktoren abhängt: Grad der Digitalisierung (siehe oben) und Erschließung im Allgemeinen, Zugänglichkeit von Katalogen sowie individuelle Sperrfristen. Die Erhebung der Daten findet häufig noch auf dem klassischen Weg statt: Recherche und Bestellung von Akten über Findbücher oder Inventarlisten, Durchsehen der Archivalien, abfotografieren und händisches Erfassen von relevanten Daten. Aufgesucht wurden die Staats- und Bundesarchive Österreichs, Deutschlands und der Schweiz und darüber hinaus Privat- und Vereinsarchive. Letztere wurden besonders für die Erforschung informeller Beziehungen genutzt. (siehe 3.2.2.3 Informelle Daten)

Die Recherche von bibliothekarischen Beständen erfolgt über Bibliothekskataloge und andere bibliographische Datenbanken, namentlich erwähnt wurde die ZDB. Die Recherche konzentriert sich dabei auf die im folgenden Kapitel behandelten Datentypen. Ein weiterer Anlaufpunkt sind domänenspezifische Datenbanken, Personendatenbanken oder Projektdatenbanken wie beispielsweise CorrespSearch, welches im Kontext von Briefnetzwerken erwähnt wurde. (vgl. Person 4 Abs. 48)

3.2.2 Datentypen

Die Netzwerkanalyse versucht Relationen darzustellen und zu analysieren. Die Interviewten haben verschiedene Forschungsschwerpunkte und demnach unterschiedliche Schwerpunkte bezogen auf die zu betrachtenden Datentypen.

3.2.2.1 Bibliographische Daten

Metadaten zu bspw. Verlagen oder Zeitschriften werden u.a. zur Analyse von Publikationsnetzwerken verwendet. Daten zu Zeitschriften können dabei in sozial und inhaltlich unterteilt werden: „[...] welche Zeitschriften repräsentieren zu welchem Zeitpunkt welche Inhalte, dienen welchen Topics, wo sind da Konjunkturen - als auch die Person, das heißt selbst wenn ich keine Person habe, habe ich noch ein Topic, habe noch einen Herausgeber, habe ich einen Zeitpunkt.“ (Person 3 Abs. 12) Wichtig ist in diesem Zusammenhang auch die „Nicht-Information“: „[...] wer schreibt da, wer schreibt da nicht, also auch Nicht-Information ist ja eine Information oder wer bricht weg oder taucht plötzlich auf.“ (Person 1 Abs. 20)

3.2.2.2 Personendaten im Kontext

Hierbei geht es um das Zusammenspiel aus biographischen, institutionellen, geographischen und zeitlichen Daten und dient zum Beispiel zur Analyse von Daten zu Personen im Kontext von Ereignissen, speziell Veranstaltungen (vgl. Person 1 Abs. 14), oder auch Daten zu Veranstaltungen ergänzt durch Personen- und Vereinsdaten: „Mitgliedschaft in Parteien, Vereinen, Protestbewegungen, Bürgerinitiativen und möglichst viele biografische Daten, also vom Wohnort, Beruf, Teilnahmen an Protestaktionen und umgekehrt, wenn ich [unverständlich] von einer Bewegung oder an einem Verein die Mitglieder, die Aktionen die von denen geplant wurden. Ja, in der Hinsicht so dieses Netzwerk aus Vereinen und Akteuren da herauszuarbeiten.“ (Person 7 Abs. 18) Auch Daten zu Personen im Zusammenhang mit Institutionen spielen immer wieder eine Rolle (vgl. Person 5 Abs. 6): Personen in Verbindung mit Organisation und Zeitraum als auch Daten über die Institution selbst, z.B. die Anzahl der Mitglieder, die Zeit, in der eine Organisation oder auch Bewegung fortbesteht, das Ausmaß geographischer Ausbreitung.

Wie anfänglich beschrieben, spielen nicht nur Einzelpersonen und Ihre Handlungen und Beziehungen eine Rolle, sondern auch die politischen, ökonomischen und sozialen Begleitumstände.⁹ Untersucht werden dementsprechend auch Beziehungen zwischen verschiedenen Systemen und der Einfluss aufeinander und andere: “[...] was motiviert Politik dazu, Impulse aus Wissenschaft aufzugreifen, beziehungsweise Veränderungen vorzuschlagen oder Gesetzesveränderungen zu überlegen und da stellte ich fest, dass es ein sehr komplexer Prozess war, in den viele Systeme eingebunden waren - also ich mein Wissenschaft, Medien, Politik, Kultur und auch andere.” (Person 1 Abs. 2)

3.2.2.3 Informelle Daten

Die auch als “weak ties” bezeichneten informellen Daten sind ein wichtiger Bestandteil der Historischen Netzwerkforschung und stellen eine besondere Herausforderung dar, da diese nur schlecht bis gar nicht dokumentiert sind. Problematisch, neben der schlechten Quellenbasis, ist die Umwandlung der Informationen in einen standardisierten Datensatz: „Ich finde vor allem die informellen Beziehungen sehr interessant, als dass es ja die persönliche Erfahrung ist [...] Gesetze, politische Sachen, Interviews - das ist die eine Ebene, die kommuniziert wird, aber der Weg dahin, da läuft ja auch sehr viel über informelle Wege, die wiederum logischerweise nicht in den Quellen zu finden sind, wenn sie gerade bei einer Versammlung bei einer Tasse Kaffee oder bei einer Blockade bei einer Flasche Bier ausgetauscht werden. [...] Und das war auch so ein bisschen der Punkt für mich, wie bringe ich diese informellen Beziehungen, wo ich von manchen wusste, dass es sie gibt, aber die sehr schwierig bis gar nicht für mich in einen Datensatz wirklich umsetzbar waren, sondern die eben eher durch diese Diskurse, die dann noch außenrum geführt werden, aufgegangen sind. Und das war auch eine der Schwierigkeiten, wo ich mich gefragt habe, wie geht man mit informellen Beziehungen, die vielleicht nicht glasklar zum Ziel führen, aber die trotzdem für dieses Gesamtergebnis sprechen, wie nehme ich die auf, wie bekomme ich die in einen Datensatz umgewandelt.“ (Person 7 Abs. 20) Um Daten zu informellen Beziehungen greifbarer zu machen, wird in einigen Fällen auf Zeitzeugeninterviews und kleinere Privatarchive zurückgegriffen und daraus standardisierte Daten erhoben. Diese Daten werden zusätzlich durch Sekundärquellen - Publikationen die für die Erhebung von Primärdaten mitgenutzt werden - ergänzt. (vgl. Person 1, Person 6, Person 7)

⁹ Vgl. Düring, Marten & Keyserlingk, Linda: Netzwerkanalyse in den Geschichtswissenschaften. Historische Netzwerkanalyse als Methode für die Erforschung von historischen Prozessen. In: Schützeichel, Rainer & Jordan, Stefan (Hrsg.): Prozesse - Formen, Dynamiken, Erklärungen. Springer, 2015, S. 337

3.2.3 Erhebungsmethoden

Die Datengewinnung erfolgt zum Teil über klassische Literaturanalyse oder Annotation von Biografien, Zeitschriften oder anderen Quellentypen - in den Texten werden systematisch Entitäten ausgezeichnet und extrahiert. In diesem Zusammenhang wird auch OCR erwähnt, dieses wird als arbeitsleichternd, aber nicht immer praktikabel beschrieben. (vgl. Person 2 Abs. 8. Person 3 Abs. 33, Person 6 Abs. 51)

In einigen Fällen wird auf Sekundärdaten - bereits existierendes Datenmaterial, das von Dritten für statistische Analysen zur Verfügung gestellt wird - zurückgegriffen. Als Problem sekundärer Daten wird die Trennung der Prozesse - Erhebung und Analyse - genannt: „[...] meistens ist es so, dass die Leute, die die Daten zur Verfügung stellen, letztendlich nicht die Leute sind, die die Daten auch analysieren. Und das ist, würde ich sagen, ein großes Problem, weil die Anwendung und die Erstaufbereitung irgendwie so weit auseinandergehen.“ (Person 4 Abs. 4) Eine Lösung wäre eine individuelle Datenauswahl und -aufbereitung.

Erwähnt wird außerdem der Einsatz von Webcrawlern, um Informationen über neue und relevante Informationen zu bekommen. Im Zusammenhang mit Korrespondenznetzwerken wird auf das Verbundprojekt Kalliope hingewiesen, welches durch ein standardisiertes XML-Format einen unkomplizierten Datenexport von Metadaten und Normdaten ermöglicht – die Exportfunktion standardisierter bibliothekarischer Daten wird dabei als großer Vorteil von Bibliothekskatalogen beschrieben. (vgl. Person 2 Abs. 38) Für den Datenexport werden des Weiteren folgende Formate erwähnt: JSON, SQL-Dump, CSV und Excel. (siehe Tabelle 2) „Ich habe tatsächlich keine Präferenzen, so was das Datenformat angeht, weil ich glaube, dass da viele Datenformate relativ flexibel sind [...] dieselben Informationen kann man häufig auch einfach in einem CSV haben, es kann aber auch eine JSON Datei sein und so weiter.“ (Person 5 Abs. 38) In den meisten Fällen wird jedoch CSV als bevorzugtes Format beschrieben, da es von vielen Werkzeugen unterstützt wird. Für den Export einzelner Ausschnitte werden APIs genannt: „Es gibt bestimmt irgendwie spezielle Anwendungsfälle, in denen ist dann API hilfreicher, weil man vielleicht nicht alles herunterladen möchte, sondern irgendwie einen bestimmten Ausschnitt, das kann man dann direkt bei der Anfrage spezifizieren [...]“ (Person 5 Abs. 38) Über eine Daten-API können aktuelle Daten flexibel ausgewählt und in externen Forschungsumgebungen analysiert werden.

Sprache/Format		Erwähnungen in Interviews	Wird genutzt für
Anfragesprache	SQL	3	Datenexport/-import, Datenspeicherung, Suchanfragen
Dateiformat	Microsoft Excel	5	Datenexport/-import, Datenspeicherung
Dateiformat	CSV	3	Datenexport/-import Metadaten
Dateiformat	RDF	1	Beschreibung und Speicherung von Daten

Dateiformat	XML	2	Datenexport/-import, Datenspeicherung
Dateiformat	JSON	2	Datenexport/-import
Dateiformat	BibTeX	1	Datenexport/-import
Dateiformat	SVG	2	Datenexport/-import Visualisierungen
Programmiersprache	Python	2	Datenexport/-import Visualisierungen
Programmiersprache	D3.js / Javascript	1	Datenexport/-import Visualisierungen
Programmiersprache	R	2	Datenaufbereitung, Netzwerkvisualisierung

Tabelle 2: Verwendete Formate und Sprachen

3.3 Qualitätssicherung

Zu Maßnahmen der Optimierung der Datenqualität zählen die Entfernung und Korrektur von Fehlern oder Uneindeutigkeiten in den erhobenen Daten mit Hilfe der intellektuellen Durchsicht und/oder statistischen Berechnungen. Alle Korrekturen und die Schritte bis zu dieser Entscheidung müssen dabei eindeutig dokumentiert werden. (siehe Kapitel 3.5.2 Dokumentation der Qualitätssicherung) Bei historischen Daten handelt es sich zum Großteil um Bereinigungen von Ambiguitäten oder Informationslücken. (siehe Kapitel 5 Problematik Datenqualität)

Die Qualitätssicherung ist ein Prozess, der begleitend zur Datenerhebung und -analyse mitläuft: “Und den Prozess hab ich, glaub ich, fünf mal von vorne angefangen, (lacht) weil ich immer wieder neue Erkenntnisse hatte, was jetzt gemacht werden muss und der Prozess setzt sich immer weiter fort. Also ich arbeite parallel, ich clean diese Daten, ich mache das preprocessing, dann nehm ich praktisch die Daten wieder in ne Analyse rein, seh, ah das preprocessing hat's doch nicht gebracht und nochmal von vorne und so weiter.” (Person 4 Abs. 6)

3.3.1 Statistische Berechnung

Als Beispiel für die automatisierte Disambiguierung von Personennamen durch statistische Berechnungen wurde die Levenshtein-Distanz genannt. Hierfür wurde ein Programm geschrieben, welches die zu disambiguierenden Namen vergleicht und bestimmt. Namen, die über einem bestimmten Schwellenwert lagen, wurden intellektuell durchgesehen. Weitere statistische Berechnungen bezogen sich auf die Korrektur von Zeiträumen. Genannt werden z.B. die Bildung einer Zufallsstichprobe mit Berechnung der Unsicherheitswerte (vgl. Person 4 Abs. 20), Kodierung des Mittelwertes und Überprüfung der Robustheit der Annahmen sowie komplexere statistische Methoden wie beispielsweise die statistische Modellierung von Unsicherheiten. (vgl. Person 5 Abs. 24)

3.3.2 Intellektuelle Durchsicht

Um Ungenauigkeiten im Datenmaterial zu identifizieren, werden die Daten alternativ oder ergänzend zur statistischen Berechnung einer intellektuellen Durchsicht unterzogen. So werden zur Disambiguierung von Personennamen betroffene Datensätze nach Anhaltspunkten untersucht, die dafür sprechen, dass es sich um dieselbe Person handelt. Zu diesen Kriterien zählen u. a. Zeit (z. B. durch den Abgleich von Lebensdaten und Sendedaten), Informationen aus dem Textinhalt, die auf die Identität der Person hinweisen oder auch der Abgleich mit inhaltlich identischen Dokumenten (z. B. Sitzungsprotokolle mit für gewöhnlich gleichen Teilnehmenden). Diese Arbeit kann durch die Dokumentation von Auffälligkeiten im Datenmaterial erleichtert werden – ähnlich der allgemeinen Dokumentation des Arbeitsprozesses (siehe Kapitel 4.5 Dokumentation): „Ja also ich habe mir bei nodegoat auch ein paar Freitextfelder geschaffen für jede Person. Dann schreibe ich eben von wegen: War da anwesend, aber wurde geschrieben als. Und dann kann ich so oftmals dann doch irgendwie Ambiguitäten auflösen und entscheiden: Ok das war wahrscheinlich dieselbe Person [...]“ (Person 6 Abs. 65) Die im Anschluss an diese Arbeit folgende Zusammenführung der disambiguierten Personendatensätze wird als Herausforderung beschrieben: „Also wenn ich jetzt feststelle, das ist das Pseudonym von einer Person, die ich schon habe, die dann zusammenzufriemeln und aus zweien teilweise fünf Personen wieder eine zu machen, das ist rein datenbanktechnisch schwierig.“ (vgl. Person 3 Abs. 28) Versionierungen zur Sicherung des vorherigen Datenbestands werden in diesem Zusammenhang als hilfreich beschrieben.

3.3.3 Visualisierung

Visualisierungen provozieren einen höheren als auch unerwarteten Erkenntnisgewinn (siehe Kapitel 3.4.3 Visualisierungen als Datenanalyse-Instrument), können aber auch dabei helfen, Fehler im Datenmaterial aufzudecken. So werden einige Fehler erst nach der eigentlichen Qualitätssicherung, im Prozess der Datenanalyse/Visualisierung sichtbar: „[...] es sind immer noch Fehler drinne, das ist das Schöne, das fällt einem dann auf bei der Netzwerkanalyse, wenn man es dann visualisiert teilweise erst auf, dass da ein Knoten absolut unverbunden ist [...]“ (Person 3 Abs. 20)

3.4 Datenanalyse

Die quantitative Netzwerkanalyse macht durch die Gewichtung von Beziehungen Strukturen innerhalb eines Netzwerks sichtbar. Erwähnt werden in diesem Zusammenhang die Identifikation von Zentralität, Dichte und Hubs.

3.4.1 Statistische Methoden

Durch die Anwendung statistischer Methoden können Aussagen über die erhobenen Daten getroffen werden. Allgemein ist bei den Interviewten eine Offenheit gegenüber mathematischen Kennzahlen festzustellen, entscheidend sind jedoch die individuellen technischen Vorkenntnisse. So wird von intensiver Auseinandersetzung mit den Konzepten der HNA und der Anwendung dieser auf den eigenen Themenbereich berichtet, es wird betont, dass Personen, die neu an die Methodik herantreten, die Anwendung und Interpretation schwer fällt. Als Beispiel wird der Knotengrad genannt, der ganz allgemein Aussagen darüber treffen kann, wie viele Kontakte eine Person hatte und daraus zu schließen ist, dass diese bedeutender war. Es können aber auch kompliziertere Sachverhalte dargestellt werden, für die mehr technische Kenntnisse erforderlich sind, wie die Reduktion auf bestimmte Kontakte, z. B. Empfänger und Absender bei Briefen. Aktuell werden deshalb eher einfache statistische Analysen durchgeführt, wobei das Interesse besteht, mehr

Analysemöglichkeiten zu erlernen und anzuwenden. Auch Werkzeuge können bei der Analyse unterstützen: „[...] ich wollte die verschiedenen Maße betrachten und vielleicht ist es auch eher etwas oberflächlich - also ich hätte gern noch mehr gemacht [...] Grundsätzlich finde ich die Art der Unterscheidung von Zentralität unglaublich spannend.. und war da im Nachhinein sehr dankbar, dass es zum einen schon Publikationen gibt, [...] zum anderen, dass die Software mir da auch ein Stück weit entgegengekommen ist bei der Berechnung dieser Zentralität - also selbst hätte ich das nicht gemacht.“ (Person 1 Abs. 40)

Um die statistischen Berechnungen aussagekräftiger zu machen, können Gesamtnetzwerke in einzelne Teilnetzwerke aufgeteilt werden, da sich die Datenqualität teilweise sehr stark von Bestand zu Bestand unterscheidet: „[...] ich habe viele Institutionen, für deren Sitzungen ich eben Netzwerkdaten erhebe [...], aber ich habe zum Beispiel von der einen Organisation viel mehr Datengrundlage und viel mehr Treffen als von der anderen. Und auch die einen listen wirklich jeden Teilnehmer und die anderen die Top drei sozusagen. Und deswegen denke ich, dass es, wenn man alles in einen Topf rein wirft - was ja die Formel, die behandelt ja alle Knoten gleich, ohne zu wissen, dass die Datengrundlage nicht so gut ist bei dem einen Knoten - dass es einfach nicht viel Sinn macht da irgendwelche gesamten Zentralitätsindizes zu erheben - wer ist jetzt der zentralste Knoten im ganzen Netzwerk. Aber innerhalb der einzelnen kleinen Gruppierungen macht das schon- also könnte man schon rechtfertigen das zu machen.“ (Person 6 Abs. 33) Darüber hinaus sind quantitative Erkenntnisse immer durch zusätzliche qualitative Analysen zu ergänzen, wie dieses Beispiel zeigt: „[...]der war am häufigsten auf allen Treffen, ja deswegen ist er vielleicht wichtig - aber da kommt dann auch wieder die qualitative Analyse rein, die einfach sagen kann - ja ok, er war eben der Protokollführer, deswegen war er immer da und nicht, weil er irgendwie die wichtige Person ist.“ (Person 6 Abs. 33)

3.4.2 Visualisierung als Datenanalyse-Instrument

Visualisierungen werden dazu genutzt, übergeordnete Muster zu erkennen. Sie dienen dazu, zeitliche und geografische Entwicklungen sozialer Netzwerke sichtbar zu machen und verschiedene Schichten – betreffend Person, Ort und Zeit – in Verbindung miteinander darzustellen. Als Instrumente zum Erkenntnisgewinn können sie dabei unterstützen, Unerwartetes, das durch einfache Betrachtung der Daten übersehen wird, aufzudecken: „Ich habe durch eine [...] Visualisierung zwei Knoten gefunden, die relevant sind und zwar nur durch die Visualisierung. Also beim einfachen Eintippen ist es mir nicht aufgefallen, weil man tippt ja in einem Tag sehr schnell sehr viele Namen ein. Aber dann durch die Grafik habe ich dann gesehen, dass da eben diese Verbindung existiert.“ (Person 6 Abs. 27) Zudem können Visualisierungen dabei helfen, Daten objektiver zu betrachten: „[...] da hilft das [unverständlich] schon teilweise, dieses wirklich nüchterne, wenn man so möchte ... eine Darstellung zu haben und denkt: Ah! Da sind die doch irgendwie mehr zusammen als die oder die Person ist wichtiger als andere.“ (Person 7 Abs. 30) Als konkretes Beispiel wird die Visualisierung von Teilnehmerlisten genannt, bei der die zuvor unbeachtete Sekretärin als besonders großer Knoten auftauchte.

Flussdiagramme (alluvial diagrams), welche Änderungen in der Netzwerkstruktur in einem zeitlichen Verlauf darstellen - werden als hilfreich beschrieben, um temporale Zusammenhänge darzustellen und den Fluss zwischen verschiedenen Entitäten zu visualisieren, z.B. für Personen in Raum und Zeit. Als konkretes Tool für die Visualisierung geografischer und zeitlicher Verteilung von Veranstaltungen wurde nodegoat genannt. Dieses ermöglicht Georeferenzierung, was positiv aufgefasst wird. Multimodale Visualisierungen, welche zeitliche, geographische und personenbezogene Daten in Verbindung setzen, sind aktuell nicht ohne Aufwand zufriedenstellend umzusetzen, da entsprechende Funktionen/Werkzeuge nicht existieren. (siehe Kapitel 4.2.3 Visualisierung und Netzwerkerkundungen)

Es besteht zudem ein Konflikt zwischen Analysen mittels Visualisierungen und anderen historischen Analysen: „In den ersten Vorträgen habe ich immer davon gesprochen, dass es ein iterativer Prozess ist [unverständlich: aus quantitativer?] Arbeit und qualitativer Arbeit und dass die Visualisierungen - sagen wir mal so - Hypothesen bestätigen, aber auch Irritationspunkt sein [können] und neue Pfade der Fragestellungen eröffnen - also warum ist die Verbindung da und was bedeutet das und so weiter. In der Realität sieht es einfach ein bisschen anders aus, weil man dann immer phasenweise sehr viel Daten eintippt und dann eine Grafik hat, aber dann kommt so ein Vortrag und man soll halt eben was über Inhalte sagen und dann muss man das wieder ad acta legen und einfach halt am Inhalt arbeiten, weil die meisten Historikervorträge sind einfach nicht von wegen 'hier ist meine tolle Grafik', sondern: Prozesse, Akteure, Entwicklungen. Also ich hab da so Probleme, die beiden Sachen zusammen zu bringen.“ (Person 6 Abs. 29) Es stehen demnach Anforderungen an eine schlüssige Visualisierung, die Daten und Informationen auch verschluckt, und die historische Analyse, die jedes Detail sichtbar machen möchte, dadurch aber weniger abdecken kann, im Kontrast zueinander.

3.5 Dokumentation

Um die Nachvollziehbarkeit der Analyse sowie die individuelle Beurteilung der Qualität der Forschungsergebnisse zu ermöglichen, muss der Prozess der Datenerhebung und damit verbundenen Qualitätssicherung ausführlich dokumentiert werden: „Das Wichtigste ist halt immer ... transparent zu sein, auch wieder, was man gemacht hat, dann könnte auch jemand anderes sagen: ja gut. Mittelwert, Sie haben den Mittelwert genommen, ich denke das ist falsch, ich mach das anders. Deswegen ist einfach wichtig transparent zu sein.“ (Person 5 Abs. 25) Die Dokumentation des Forschungsprozesses wird durchgehend als unerlässlich beschrieben, ein einheitliches Vorgehen kann jedoch nicht identifiziert werden.

3.5.1 Dokumentation des Forschungsprozesses

Betont wird, dass alle Schritte der Datenerhebung und des Aufbaus der Datenbasis nachvollziehbar sein müssen. (vgl. Person 2) Die Form der Dokumentation reicht dabei von einem klassischen Logbuch in einer Word-Datei über Excel und SPSS (vgl. Person 1 Abs. 34 und Person 5 Abs. 32). Ein anderes Vorgehen ist die Dokumentation über die eigene Datenbank. In einem Projekt wurden Freitextfelder in nodegoat definiert, die für die Identifikation der Datenherkunft durch bspw. Archivsignaturen sowie Vergabe von IDs für die jeweiligen Datensätze verwendet wurden. (vgl. Person 6 Abs. 71) Dies erfordert zusätzlich eine Dokumentation der Datenfelder der projekteigenen Erschließungsdatenbank (siehe Kapitel 3.2 Anforderungen Datenqualität): „Da haben wir uns Standards gesetzt, da ist für, nicht für alle, aber für die meisten Zellen definiert, was da rein darf, was da drin ist.“ (Person 3 Abs. 20) Die Abspeicherung von Zwischenschritten in Form von Versionierungen wird auch als Form der Dokumentation praktiziert, um bei Fehlern auf ursprüngliches Datenmaterial zurückgreifen zu können. (vgl. Person 3 Abs. 28)

3.5.2 Dokumentation der Qualitätssicherung

Wichtig ist die Dokumentation auch für die Nachvollziehbarkeit von Entscheidungen, die besonders bei der Datenbereinigung oder der Verwendung von lückenhaften oder widersprüchlichen Quellen getroffen werden. So wird beispielsweise mittels Bemerkung – Fußnote oder ähnlichem – festgehalten, dass es Angaben gibt, die nach aktuellem Forschungsstand nicht lösbar sind. Ein Beispiel ist ein Widerspruch zwischen schriftlicher Quelle und Zeitzeugenaussage: „Ich habe das dann vorsichtig formuliert, dass - also laut schriftlichen Quellen - die Person dazugehört hat, Fußnote Zeitzeuge XY ist hier anderer Ansicht. In dem Knotenpunkt taucht die Person auf, weil ich für mich entschieden habe - ich kann ja nicht einen Halbpunkt einzeichnen. Für mich war

das Gewicht der drei schriftlichen Quellen größer wie die starke Aussage der einen Person, aber das war eine Entscheidung.“ (Person 1 Abs. 34) Diese Problematik der Sichtbarmachung von Unsicherheiten könnte durch Visualisierungen gelöst werden. (Mehr zum Umgang mit Datenunsicherheiten in Kapitel 3.3 Qualitätssicherung)

3.6 Datenaustausch und Datenpublikation

Die Publikation von Datensätzen und Analyseschritten ist ein wichtiges Thema für Forschende der historischen Netzwerkanalyse, da dadurch eine effektivere Nachnutzbarkeit der Daten gewährleistet werden kann. Als Beispiel für die Datenpublikation wird Zenodo genannt. (vgl. Person 2 Abs. 36)

Es wird eine einheitliche Datenpublikation vergleichsweise zu den Naturwissenschaften angestrebt: „Also in den Naturwissenschaften ist es ja gang und gebe, dass die Datensätze zu Artikeln die erscheinen eben parallel publiziert werden, das ist bei uns eben noch nicht üblich. Aber wäre lohnenswert, darüber nachzudenken.“ (Person 3 Abs. 20) Andere äußern Zweifel am Interesse der Geschichtswissenschaft an diesen Daten und damit auch an der Sinnhaftigkeit der Publikation dieser. „[...] man kann natürlich lange Seiten darüber schreiben, wie die Daten erhoben wurden, aber ich glaube also dem Betreuer und die Geschichtswissenschaft, die haben dafür wenig Interesse. Es gibt auch immer darüber Talk (lacht), dass man irgendwie alternative Publikationsplattformen schaffen sollte, auch dass Datenerhebung allein schon, sagen wir, ein Prozess ist, der ein Ergebnis ist, also eine Publikation wert ist. Also das die Datenerhebung überhaupt eine Wertigkeit hat ... und nicht nur Hilfsmittel ist ... Ja, ich hab meine privaten Notizen, aber inwiefern die auch veröffentlicht werden oder irgendwie auch jemanden interessieren, bin ich mir halt nicht so sicher.“ (Person 6 Abs. 69)

Zudem wird eine Diskrepanz zwischen individuellen Projekten und größeren Verbünden deutlich. Kolleg:innen seien offen und bereit die eigenen Daten bereitzustellen, bei größeren Institutionen hingegen – es wird konkret die ZDB genannt – gestaltet sich der Austausch schwieriger, besonders wenn es um die Korrektur von Fehlern geht. (vgl. Person 3 Abs. 22) Zwei Interviewte dachten zudem darüber nach, wie Daten aus individuellen Projekten an Verbünde zurückgespielt werden können. Dabei kommt aber die Frage nach der korrekten Abbildung von Provenienzen auf, die klare Rahmenbedingungen und Standards voraussetzt. (vgl. Person 1, Person 2)

4 Werkzeuge

Für die einzelnen Schritte des Forschungsprozesses werden verschiedene Werkzeuge unterstützend angewandt. Neben den spezifischen Werkzeugen werden erwähnte Funktionen aufgeführt. Zudem wird auf Funktionen eingegangen, welche durch die Interviewten als hilfreich und bisher nicht ausreichend abgedeckt beschrieben werden.

4.1 Verwendete Werkzeuge und Funktionen

Eine Zusammenfassung über verwendete Werkzeuge und entsprechende Funktionen ist Tabelle 1 zu entnehmen. Des Weiteren sind in den Interviews erwähnte positive sowie negative Aspekte der Werkzeuge mit aufgeführt.

Name Tool	Erwähnungen	genutzt für	positiv	negativ
Gephi	3	Netzwerkvisualisierung, Filterfunktion, Datentransformation von two-mode zu one-mode-network	Guter Support, große Community, Datenexport	Datenimport
NodeXL	2	mathematische Berechnungen, Netzwerkvisualisierung	Freeware, anwenderfreundlich	zweidimensional - keine Darstellung von Zeitlichkeit
nodegoat	2	Visualisierung, Filterfunktion, Datenspeicherung, Datenmodellierung, Dokumentation, mathematische Berechnungen, kollaborative Datensammlung	anwenderfreundlich, ermöglicht Entwicklung eigener Datenmodelle, kollaborative Datensammlung möglich	
neo4j	2	(nur testweise verwendet)		Informationsverlust
Visone	1	Kartenvisualisierung, Datentransformation	Datenimport/-export - unterstützt verschiedene Formate, "gut, einfach, stabil" Darstellung von Knotentypen, einfache und intuitive Bedienung	Performanz kritisch - besonders bei großen Datenmengen
OpenRefine	1	Datenanreicherung/-bereinigung, Datentransformation	Intuitive Bedienung, einfache Transformation und Anreicherung mit GND Informationen	
MySQL	1	Datenspeicherung		Technisch "herausfordernd", setzt SQL Kenntnisse voraus
QGIS	1	Georeferenzierung		
Stata	1	Statistische Analyse	Performance - schnell	
VennMaker	1	(nur erwähnt)		
(R)	1	Datenaufbereitung, Netzwerkvisualisierung	Gewährleistet Transparenz und Nachvollziehbarkeit	

Tabelle 3: Verwendete Werkzeuge

Genutzt werden Werkzeuge allgemein für Visualisierungen, Metadatenabgleich, multimodale Netzwerkanalyse, Datenmodellierung und -speicherung sowie statistische Berechnungen. Unterstützt werden sollte darüber hinaus Datenexport und -import, Such- und Filterfunktion sowie kollaborative Zusammenarbeit.

Usability, Kosten, Support und Performance wurden ebenfalls als wichtige Nutzungsaspekte erwähnt. Systeme sollten niedrigschwellig und anwenderfreundlich sein, kostenlose Nutzung wäre ein Pluspunkt. Unter Support wurden verschiedene Beispiele genannt: Tutorials auf Youtube, Filme zur Einführung in das Programm, Praxisberichte von Projekten, die das Programm genutzt haben und über Umgang und Resultate informieren sowie eine aktive Community um Fragen zu stellen und zu beantworten. Wichtig ist ein gesicherter Support über einen längeren Zeitraum hinweg: „Aber ich bin da sehr glücklich im Moment, das wird immer noch supportet - das ist ja auch so eine Sache, dass viele Sachen am Anfang sehr gehypt werden und dann nach zwei, drei oder fünf Jahren kein Support mehr da ist, keine Community, die das weiter pflegt [...]“. (Person 3 Abs. 16, Kommentar bezieht sich auf Gephi) Kommentare zur Performance beziehen sich auf Probleme mit großen Datenmengen und Umwandlungen/Berechnungen dieser. Diese führen teilweise soweit, dass Ergebnisse nicht produzierbar sind. (vgl. Person 2)

4.2 Anforderungen an Funktionalitäten

Im Folgenden sind Funktionen zusammengefasst, die von aktuellen Werkzeugen nicht oder nur ungenügend abgedeckt und als hilfreich eingeschätzt werden.

4.2.1 Datenexport und -import

Im Bereich Export- und Importfunktionen werden konkrete Formate genannt, die unterstützt werden sollten. Diese sind in Formate für Visualisierungen, Metadaten sowie Datenbankabfragen zu unterteilen. Für den Export von Visualisierungen wurde SVG als bevorzugtes Format erwähnt, weitere Programmiersprachen und Formate sind D3.js und Python sowie Java-Script (siehe Anhang Tabelle 2). Als Exportformate für Metadaten werden CSV und XML genannt. Für den Export einzelner Ausschnitte werden zudem APIs genannt. (siehe auch Kapitel 3.2.4 Erhebungsmethoden) Des Weiteren wird eine direkte Abfrage der Datenbank durch die Möglichkeit, vom System generierte Datenbankabfragen zu editieren oder selbst neue zu entwickeln.

4.2.2 Datenaufbereitung

Erwähnt wurde auch die Verfügbarkeit von Digitalisaten oder Volltexten sowie die Einbindung von Übersetzungen: „Dann den Text des Briefes wirklich original. In einer idealen Welt wäre auch noch eine Übersetzung dabei, weil frühes Neuhochdeutsch oder Latein ist jetzt nicht so meine Stärke, aber das ist wirklich schwer, das ist wirklich utopisch im Moment glaube ich, das automatisch zu übersetzen.“ (Person 4 Abs. 24)

4.2.3 Visualisierung und Netzwerkerkundungen

Im Bereich der Visualisierungen soll die Kombination bestimmter Knotentypen ohne die Abbildung direkter Zwischenknoten erleichtert werden (vgl. Person2 Abs. 32) sowie die Verbindung von zeitlichen und räumlichen Daten - beides hebt die Wichtigkeit von Filterfunktionen hervor. Das Netzwerk einer Person soll sich dem zeitlichen Verlauf anpassen, so dass sichtbar wird, wo sich eine Person zu einem bestimmten Zeitpunkt

befunden hat und wie sich dadurch ihre Beziehungen verändern - ein "Link zwischen Zeit, Raum und den Relationen". (Person 2) An anderer Stelle wird die Abbildung dieser verschiedenen (zeitlichen und räumlichen) Ebenen als "virtuell durch die Zeit gehen" (Person 1 Abs. 48) beschrieben.

Es werden sich Systeme gewünscht, die offen für explorative Vorgänge sind und so auch dabei helfen, neue Thematiken „spielerisch“ zu entdecken. Dies kann durch die Vorgabe von ausgewählten Kategorien oder die individuelle Erweiterung von Suchbegriffen erfolgen. Um anfängliche Schwellenängste vorwegzunehmen, sind Einführungsprogramme, die die Funktionen des System durch die Vorgabe von Kategorien erklären, hilfreich: „[...] eine Software, die zum einen, wenn ich [...] mich in das Thema neu rein bewege mir Kategorien vorgibt, die ich füllen kann, sodass ich auch von der Software lerne, wie ich sie anwenden kann [...]“ (Person 1 Abs. 36)

4.2.4 Dokumentation und Datenpublikation

Häufig erwähnt wurden Möglichkeiten der Kommunikation von Forschungsergebnissen und -prozessen (siehe auch Kapitel 3.6 Datenaustausch und Datenpublikation). Dies bezieht sich zum einen auf das sichere und standardisierte Teilen von Daten „[...] ich kann mir vorstellen, dass ich quasi Wissen teile, wenn ich quasi auch Wissen erhalte - also dieser Sharing-Gedanke, den wollte ich jetzt ausdrücken“ (Person 1 Abs. 36) sowie das Abbilden (Transparenz) des eigenen Arbeits- und Rechercheprozesses – zum Beispiel die Nachvollziehbarkeit des Suchwegs: wie oft und an welchen Stellen wurde das Netzwerk im Suchprozess gefiltert oder erweitert. (vgl. Person 2 und 3)

5 Problematik Datenqualität

Die Datenqualität ist ein wichtiges Problemfeld der historischen Netzwerkanalyse und betrifft alle Ebenen des Forschungsprozesses. In den Interviews wurden verschiedene Probleme, als auch Anforderungen, die beim Umgang mit historischen Daten berücksichtigt werden müssen, genannt. Als häufigste Probleme wurden Ambiguität und Ungenauigkeit sowie Informationslücken und Unvollständigkeit erwähnt. Ebenso sind die Historie der Metadaten selbst - sich über die Zeit verändernde Verständnisse von Begriffskategorien u. a. - sowie die Unsichtbarkeit von informellen Beziehungen Herausforderungen für die Datenqualität. Die Anforderungen umfassen: Transparenz, Dokumentation der Datengrundlage sowie die Verwendung von Standards.

5.1 Probleme mit der Datenqualität

5.1.1 Ungenauigkeit und Ambiguität

Am häufigsten wurden Ungenauigkeiten und Ambiguitäten als Problem im Datenmaterial benannt. Besonders kritisch ist die Namensgleichheit von Personen und Orten, doch auch ungenaue Zeiträume und widersprüchliche Klassifikationen zur Kategorisierung von Metadaten, ein Beispiel sind Berufsgruppen, wurden bemängelt. Als Beispiel für die Namensgleichheit von Personen wurde Martin Luther genannt: „[...] damals gab es ja noch keine einheitliche Rechtschreibung im 16. Jahrhundert, das heißt, Martin Luther hat sich halt nicht immer Luther genannt, sondern da, wo er gerade war "Martin auf der Wartburg" oder "Luder" oder dann mit "th" oder nicht, dann wusste man nicht ist ein Sohn gemeint oder jemand anderes - solche Sachen.“ (Person 4 Abs. 6) Die Problematik existiert auch für die Schreibweise von Orten. (vgl. Person 4 Abs. 6)

Die Angabe von Zeiträumen in den Originaldokumenten ist sehr unterschiedlich und somit nur schwer standardisierbar. Um einen Überblick darüber zu bekommen und so die Problematik greifbarer zu machen,

können Kategorisierungen dieser Ungenauigkeiten vorgenommen werden: „Eine ist zum Beispiel, ich habe es tagspezifisch genannt, da steht dann wirklich "15. Juli 1520", das ist das Beste, genau das will ich, die lass ich unberührt. Dann gibt es aber auch Kategorien, die bezeichnet Perioden oder mehrere Tage, irgendwie "13., 14., 15." oder "im Juli 1520" solche Sachen. Oder dann gibt es noch Kategorien, da nennen sie einen Ankerpoint praktisch, zum Beispiel "um den 15. Juli 1520" und sie wissen nicht, ob es der 15. oder irgendwas anderes [war]. Und dann gibt es natürlich auch noch Einträge, da steht gar nichts drin oder nur Fragezeichen oder sonst was.“ (Person 4 Abs. 6) Zum Teil wird die Konsequenz gezogen, Daten mit ungenauen Zeitangaben, wie zum Beispiel „um 1800“, von der Auswertung auszuschließen, da die Durchsicht und Korrektur bei großen Datenmengen mit einem zu hohen Aufwand verbunden ist. (vgl. Person 2)

5.1.2 Informationslücken und Unvollständigkeit

Für statistische Berechnungen werden zuverlässige und vollständige Daten benötigt, dennoch wird darauf hingewiesen, dass historische Datenbestände nie vollständig sind und immer nur einen Ausschnitt der Realität abbilden können, weshalb fast immer zum „Mut zur Lücke“ plädiert wurde: „[...] also prinzipiell [...] gab es keine Quelle, die ich wegen fehlenden Informationen ausgeschlossen habe. Es ist eher ein bisschen wie ein Gesamtpuzzle und zunächst nimmt man lieber mal alles auf oder zu viel - rausschmeißen kann man es immer noch.“ (Person 7 Abs. 18) An einer Stelle wird eine Vollständigkeit von 50% als Ideal sowie 33% als Untergrenze für die Weiterverarbeitung der Daten beschrieben – alles unter 33% wird ausgeschlossen: „[...] also so die Hälfte ist das, was ich immer so gerne hätte.. mindestens, manchmal sind es zwei Drittel, dann freu ich mich. Bei Sachen, wo ich dann wirklich nur noch ungefähr ein Drittel der Daten habe, da mache ich dann auch keine Netzwerkanalyse, [...] weil dann Leerstellen einfach zu groß sind.“ (Person 3 Abs. 12) Wichtig ist Transparenz, die durch genaue und zuverlässige Dokumentation auf beiden Seiten – Datenanbieter und Forschende – die Nachvollziehbarkeit von Entscheidungen gewährleistet. (siehe auch Kapitel 4.5 Dokumentation)

5.2 Anforderungen an die Datenqualität

Transparenz, Dokumentation der Datengrundlage sowie Standardisierung wurden als Kriterien für Datenqualität genannt.

5.2.1 Transparenz durch Dokumentation

Transparenz bezieht sich unter anderem auf die genaue Dokumentation von Entscheidungen beim Aufbau von Datenbanken oder anderen Systemen. Ein konkretes Beispiel hierfür ist die Kodierung von Zeiträumen, welche aufgrund ihrer vielfältigen und teilweise widersprüchlichen Erscheinungsweisen (s. o.) nicht einheitlich erfasst werden: : „[...] also manchmal hat man eine Gründungsurkunde, manchmal hat man konkurrierende Gründungsurkunden, manchmal hat man verschiedene erste Erwähnungen, manchmal weiß man nur "um 800" und so weiter.. und da muss man wissen, wie kamen die zu dieser Zahl und wie viele alternative Quellen oder wie viele alternative mögliche Jahreszahlen wurden nicht berücksichtigt oder auch wenn da so ein Zeitraum steht "800 bis 850", nach welcher Konversion wurde es jetzt dort eingegrenzt. Genau also wenn man weiß, die erste Erwähnung war "um 820", dann wird manchmal "800 bis 850" kodiert. Es muss alles genau dokumentiert sein oder es sollte bestenfalls alles genau dokumentiert sein.“ (Person 5 Abs. 20) Unterstützen können hierbei bspw. Versionierungsnummern, da diese eine genaue Nachvollziehbarkeit ermöglichen. (vgl. Person 2 Abs. 42)

5.2.2 Standards und Identifikatoren

In anderen Bereichen gibt es hingegen anerkannte und weit verbreitete Standards, welche datenbank-übergreifend eine einheitliche Erschließung und damit eine Vergleichbarkeit der Daten sowie Interoperabilität zu anderen Systemen gewährleisten. Schwieriger ist es bei interdisziplinären oder internationalen Projekten, da die Verwendung von Standards von disziplinären und regionalen Kontexten abhängig ist. Für einen dauerhaften Nachweis digitaler Objekte sind Persistent Identifiers (PIDs) eine Option. Die Verwendung von Normdaten als Mittel der Standardisierung ist hilfreich für eine eindeutige Identifikation und Disambiguierung von Entitäten und wird an einigen Stellen erwähnt. Doch auch diese sind nicht fehlerfrei und erfordern einen kritischen Umgang. Namentlich wurden Probleme mit der GND und Google Maps-APIs genannt. In einem Projekt wurden GND-IDs automatisch erhoben und im Anschluss nochmal händisch überprüft (vgl. Person 3 Abs. 33), da Personen z. B. doppelte Einträge haben. Bei der Georeferenzierung mittels Googlemaps-APIs war hingegen die fehlerhafte Datenbasis das Problem: „[...] da sind wir manuell wirklich noch mal rübergegangen und haben das gecheckt, vor allem die, die am Ende in den USA rauskamen, also wenn bei "Basel" das "L" weggelassen wird, dann kommt "Base" raus und Googlemaps denkt die ganze Zeit, es handelt sich um irgend eine Militärbasis in Ostflorida.“ (Person 4 Abs. 22)

5.3 Historische Daten als Herausforderung für Visualisierungen

Genau wie der Datenbestand, der der Visualisierung zugrunde liegt, ist auch die Visualisierung selbst nur eine Repräsentation und keine Replikation der Realität und geht daher immer auch mit einer Reduktion der Komplexität einher.¹⁰ Genau diese Reduktion wird kritisiert, da es Visualisierungen zu „höchst interpretierbedürftige[n] Quellen“ (Person 7 Abs. 30) macht und besonders in der öffentlichen Diskussion intensiver Erklärung bedarf. Deswegen sind einige der Interviewten dazu übergegangen, Visualisierungen lediglich für den eigenen Erkenntnisgewinn zu verwenden und nicht zur Darstellung von Forschungsergebnissen: „[...] immer, wenn man ein Bild also so eine Grafik zeigt, muss man die erstmal eine halbe Stunde erklären und sagen was sie jetzt nicht zeigt und was sie zeigt und warum der Algorithmus jetzt den einen Punkt in die Mitte macht und der andere am Rand, aber das eigentlich keine Wertung ist. Bilder zeigen ist sehr komplex und ich bin eher davon abgekommen.“ (Person 6 Abs. 27) Gleiches sagt Person 7 (Abs. 14): „[...] Netzwerkgrafiken bin ich vorsichtig, die habe ich auch in der aktiven Zeit, also als ich doch noch sehr stark in der Netzwerkanalyse involviert war, eigentlich nie gezeigt und wenn dann nur äußerst ungern, wenn es dann doch irgendwie verlangt war.“

Kritisiert wird darüber hinaus die unbegründete Verwendung von Visualisierungen ohne inhaltlichen Mehrwert: „Und das ist zum Beispiel halt auch mein Problem mit so einigen Sozialwissenschaftlern, die dann qualitative Netzwerkforschung machen, die meinen die Tatsache, dass Menschen irgendwie - Beziehungen bestehen zwischen Menschen, das wussten wir schon immer, dafür braucht man nicht Netzwerkanalyse - und nur, weil sie dann halt irgendwie so eine Visualisierung machen aus irgendwelchen qualitativen Daten, die sie haben und dann anfangen dieses Bild zu interpretieren ... also das ist ja keine Netzwerkanalyse, weil ... irgendwie ist es ja schon das formale Instrumentarium, das einem dann erlaubt Schlüsse zu ziehen, die eben nicht vielleicht offensichtlich sind nur indem man das irgendwie anschaut ... sondern dann ist das irgendwie so ein abstraktes Bild, dann kann man auch Kandinsky interpretieren.“ (Person 5 Abs. 48)

¹⁰ vgl. Linda Freyberg: Ikonizität als Erkenntnismittel - Vollständigkeit, Verständlichkeit und Kontextualisierung als Grundprinzipien der Visualisierung. In: E.Bisanz (ed.): Applied Interdisciplinary Peirce Studies, Peter Lang, 2020

Kritische Reaktionen sind auch aus der Community zu vernehmen: „[...] es kam dann immer so der Einwand: Ja, das ist ja spannend und sieht irgendwie cool aus, aber was hast du davon? Und ich glaube das ist eine berechnete Frage in vielen Bereichen.“ (Person 7 Abs. 16) Stärkere digitale Kompetenz, Transparenz des Forschungsprozesses, als auch Möglichkeiten einer breiteren Datenpublikation können helfen ein allgemeines Verständnis zu schaffen und produktive Diskurse zu erzeugen.

6 Zusammenfassung

Der Forschungsprozess von Expert:innen auf dem Gebiet der HNA lässt sich in sechs Prozessabschnitte zusammenfassen: Entwicklung von Forschungsfragen / Theoriebildung, Datenauswahl und -erhebung, Qualitätssicherung, Datenanalyse, Dokumentation und Datenaustausch und -publikation.

6.1 Anforderungen an die Infrastruktur

Die sich aus den Interviews ergebenden Anforderungen und Bedürfnisse konzentrieren sich auf drei Kategorien: Visualisierung und Netzwerkerkundungen, Dokumentation und Datenpublikation sowie Datenexport und -import.

Visualisierungen und Netzwerkerkundungen nehmen innerhalb des Prozesses der HNA eine Sonderstellung ein: Sie können am Anfang des Forschungsprozesses eingesetzt werden, um auf neue Forschungsfragen aufmerksam zu werden, sie können bei der Analyse hilfreich sein, auf Fehler aufmerksam machen oder Forschungsergebnisse veranschaulichen. Aufgrund sehr unterschiedlicher Vorkenntnisse werden Visualisierungen und Analysen im Allgemeinen zwar angewandt, in ihrem Potential aber nicht ausgeschöpft. (siehe Kapitel 4.2.3 Visualisierungen und Netzwerkerkundungen) Entsprechend fallen die Bedürfnisse an die “Visualisierungen und Netzwerkerkundungen” stark explorativ aus:

- Prozesshafte Entwicklung von Forschungsfragen mittels Visualisierung
- explorative Vorgänge, um neue Thematiken „spielerisch“ zu entdecken
- Vorgabe von Filtern zur Exploration eines Themas
- Kombination bestimmter Knotentypen ohne die Abbildung direkter Zwischenknoten
- Verbindung von zeitlichen und räumlichen Daten
- Abbildung multimodaler Netzwerke
- Abbildung von Ambiguitäten/Unsicherheiten im Datenmaterial
- Verfügbarkeit von Digitalisaten und Einbindung von Übersetzungen

Deutlich wird an vielen Stellen eine Vorsicht bei der Darstellung von Forschungsergebnissen mittels Visualisierungen. Wichtig ist daher die Kontextualisierung von Visualisierungen und den in ihnen enthaltenen Daten sowie die Transparenz durch Dokumentation der zugrundeliegenden Datenbankstrukturen, um eine verlässliche Grundlage für weitere Forschung zu gewährleisten. Wie deutlich wurde, ist darüber hinaus auch die Prozessdokumentation und Transparenz von Entscheidungen innerhalb des eigenen Forschungsprozesses sehr entscheidend. Entsprechend konzentrieren sich viele Bedürfnisse und Anforderungen besonders auf die “Dokumentation und Datenpublikation”, wobei folgende Bedürfnisse zu erkennen sind:

- Transparenz der Datenmodellierung und Datengrundlage (von Anbieterseite)
- Prozessdokumentation durch Abbildung des Arbeits- und Rechercheprozesses
- Abspeicherung von Zwischenschritten in Form von Versionierungen

- Kommunikation von Forschungsergebnissen und -prozessen
- Sicheres und standardisiertes Teilen von Daten
- Einheitliche Datenpublikation

Es sind zudem zwei Gruppen von Datennutzenden auszumachen: Zum einen die explorativ, prozesshaft und zum anderen die stark quantitativ, sozialwissenschaftlich Forschenden. Die erste Gruppe möchte sich Themen erschließen, auf neue Fragestellungen stoßen oder einfach unbekanntes entdecken und greift dabei gerne auf vorgefertigte Visualisierungen (mit transparenter Dokumentation von Modellierungsentscheidungen) zurück. Die zweite Gruppe hingegen konzentriert sich stark auf die Datengrundlage und sieht einen großen Nutzen in aggregierten, standardisierten Daten aus verschiedenen Datenquellen. Wobei der Vorteil weniger in der Aufbereitung dieser Daten mittels Visualisierung gesehen wird, als vielmehr in der Möglichkeit des Exports der Daten und der Weiterverarbeitung dieser für eigene Analysen. Visualisierungen werden bevorzugt selbst gebaut und erst im Anschluss an die Analyse verwendet, um eine Beeinflussung durch bildliche Verzerrungen zu verhindern. Es werden folgende Bedürfnisse an den Datenexport und -import gestellt:

- Exportformate für Visualisierungen: SVG, D3, PythonScript
- Exportformate für Metadaten: Excel, CSV, JSON, XML, SQL, BibTeX
- Suche mittels Abfragesprache über Expertensuche
- Export von Teilnetzwerken mittels APIs

SoNAR (IDH) kann den Forschungsprozess demnach in allen Prozessabschnitten unterstützen. Der Wunsch nach besseren Infrastrukturen für die HNA wurde geäußert.

6.2 Validierung des modellhaften Forschungsdesigns

Das durch die historisch arbeitenden Fachwissenschaftler:innen entwickelte modellhafte Forschungsdesign beschreibt Forschungsprozesse und -fragen der Historischen Netzwerkanalyse, wie diese methodisch bearbeitet werden und welche speziellen Anforderungen sich daraus ergeben. Schwerpunkte bilden:

- Datenerhebung, Datenauswahl und Datenqualität
- Anwendungsszenarien: Erkundung von Beziehungen, Gewichtung von Beziehungen und Quantität und Qualität von Beziehungen (entspricht Hypothesengenerierung und Datenanalyse)
- Transparenz und Versionierung
- Datensicherung und Datenexport

Alle Bereiche finden sich auch in dem durch die Interviewstudie beschriebenen Forschungsprozess. Konkrete Übereinstimmungen in Bezug auf den HNA-Forschungsprozess finden sich bei folgenden Aussagen:

Modellhaftes Forschungsdesign (AP 2)	Interviewauswertung (AP 4)
Die Datenerhebung und -aufbereitung aus häufig dezentral überlieferten Archiv- und Bibliotheksbeständen stellen besondere Herausforderungen/Anforderungen an die HNA.	3.2.1 Historisches Quellenmaterial

Die Auswahl der Netzwerkdaten ist stark von persönlichen Faktoren wie der Kenntnis über Datenbezug, -format und -auswahl sowie technische Verfahren abhängig.	2.2 Community > 2.2.1 HNA Infrastruktur 3.4 Datenauswahl > 3.4.1 Statistische Methoden
<i>Gewichtung von Beziehungen: Quantifizierend und hypothesentestend:</i> Die quantitative Netzwerkanalyse ist hypothesengeleitet und macht die Strukturen innerhalb des Netzwerks durch die Gewichtung der Beziehungen sichtbar – Zentralität, Prestige, Dichte, Identifizierung von Hubs.	3.4 Datenanalyse > 3.4.1 Statistische Methoden 3.1 Entwicklung von Forschungsfragen
<i>Erkundung von Beziehungen: Explorativ und hypothesenbildend</i> - Das Erkunden von Beziehungen kann als ein qualitativ-explorativer Ansatz verstanden werden, um sukzessive Zusammenhänge zu erkennen – zur Übersicht und Ermittlung von Schnittmengen.	3.4 Datenanalyse > 3.4.2 Visualisierung als Datenanalyse-Instrument
Die Identifikation auffälliger Strukturen in (Teil-) Datenbeständen kann wiederum Ausgangspunkt für weitergehende Forschungen sein.	4.2 Anforderungen an Funktionalitäten > 4.2.3 Visualisierung und Netzwerkerkundungen

Tab. 4: Übereinstimmungen Forschungsprozess modellhaftes Forschungsdesign

Daraus schlussfolgernd werden durch AP 2 konkrete Anforderungen an SoNAR (IDH) gestellt, welche grundlegend bestätigt werden können. Besonders folgende Anforderungen finden Übereinstimmung mit den in den Interviews erwähnten Bedürfnissen:

Modellhaftes Forschungsdesign (AP 2)	Interviewauswertung (AP 4)
Die Daten verbinden Ereignisse durch Kontextualisierung wie Entstehung (z.B. Autor, Ort, Zeit), Sache (z.B. Anlass, Thema) und Überlieferung (z.B. Bestandsbildner, Aufbewahrung, Eigentümer).	6.1 Erkenntnisse > Anforderungen an "Visualisierungen und Netzwerkerkundungen"
persistente adressierte Daten mit eindeutigen, systemunabhängigen Identnummern und URI	5.2 Anforderungen an die Datenqualität > 5.2.2 Standards und Identifikatoren
Datenmodifikationen werden sekundengenau gespeichert, sodass die für das Projekt wichtige Versionierung unterstützt wird.	5.2 Anforderungen an die Datenqualität > 5.2.1 Transparenz durch Dokumentation 3.5 Dokumentation > 3.5.1 Dokumentation des Forschungsprozesses

Eine größtmögliche Datentransparenz wird [...] bereits während des Forschungsprozesses angestrebt.	5.2 Anforderungen an die Datenqualität > 5.2.1 Transparenz durch Dokumentation 3.5 Dokumentation
Erschließungsregeln, z.B. die Ressource Description and Access (RDA) für ZDB und GND, sichern die Vergleichbarkeit der Daten; sie regeln die Art und Weise der Erfassung von Ausprägungen der Entitäten[S2]	5.2 Anforderungen an die Datenqualität > 5.2.2 Standards und Identifikatoren
Am Ende einer Untersuchung mit der Forschungstechnologie SoNAR (IDH) kann das Datenset und ggf. auch in Verbindung mit der Visualisierung und geeigneten Metadaten heruntergeladen werden.	4.2 Anforderungen an Funktionalitäten > 4.2.1 Datenexport und -import
Das Datenpaket kann in einem Forschungsdatenrepositorium einer wissenschaftlichen Einrichtung hinterlegt und zitiert werden.	3.2 Datenauswahl- und erhebung > 3.2.3 Erhebungsmethoden 4.2 Anforderungen an Funktionalitäten > 4.2.1 Datenexport und -import
Evolution von Entitäten und ihren Beziehungen > Gruppe von Entitäten und ihre Beziehungen in Form von Netzwerken oder hierarchischen Strukturen	6.1 Erkenntnisse > Anforderungen an "Visualisierungen und Netzwerkerkundungen"
Geographische Repräsentation > Personen eines Samples nach Orten quantitativ erfassen und zeitliche Verläufe darstellen.	6.1 Erkenntnisse > Anforderungen an "Visualisierungen und Netzwerkerkundungen"

Tab. 5: Übereinstimmungen Anforderungen modellhaftes Forschungsdesign

Leichte Abweichungen sind in dem Prozessabschnitt Datenerhebung/Datenauswahl festzustellen. Die Einschätzung, dass "[e]ine Reproduktion und Überprüfung der Ergebnisse quantitativer historischer Analysen [...] ebenso wie eine Sekundärnutzung der erhobenen Daten für neue Forschungsfragen kaum möglich [ist]", kann so nicht bestätigt werden. Die Verwendung von Sekundärdaten - als Ausgangspunkt für eigene Forschungen oder als Erweiterung des eigenen Datenbestandes - sowie die Dokumentation/Publikation von Erhebungs- und Analysedaten und die damit verbundene Ermächtigung einer Reproduktion und Überprüfung der Forschungsergebnisse, wurden als wichtige Schritte im Forschungsprozess genannt. (siehe ...) Es wurde aber deutlich, dass Verbesserungsbedarf in diesen Bereichen besteht. So konzentrieren sich die Bedürfnisse der Interviewten u. a. auf Schwerpunkte wie: die Kommunikation von Forschungsergebnissen und -prozessen, sicheres und standardisiertes Teilen von Daten sowie eine einheitliche Datenpublikation.

6.3 Fazit

Identifiziert wurden insgesamt sechs Prozessabschnitte welche sich zu angrenzenden Disziplinen besonders durch die starke Bedeutung von Qualitätssicherung und Dokumentation unterscheiden. Diese zwei

Prozessabschnitte weisen auf die Wichtigkeit von Transparenz sowie die Schwierigkeiten im Umgang mit historischen Daten bezogen auf die Datenqualität hin. Bedürfnisse zur Unterstützung zeigen sich besonders in den Bereichen Visualisierung und Netzwerkerkundungen, Dokumentation und Datenpublikation sowie Datenexport und -import.

Forscher:innen der HNA wünschen sich einen barrierearmen Zugriff auf für sie und ihre Forschungsfragen vorbereitete, verknüpfte Datenbestände und deren Export, Werkzeuge zur Exploration und Analyse dieser Datenbestände und eine Plattform, die Publikation und Wissenskommunikation ermöglicht und unterstützt. Visualisierungen können darüber hinaus neue Möglichkeiten in der Entdeckung von Forschungsfragen und in der Erarbeitung von Antworten eröffnen. Durch niedrigschwellige Angebote können außerdem Einstiegshürden gesenkt werden und einem großen Kreis von Forscher:innen neue Forschungsmethoden nahe bringen. Wichtig ist dabei die Kontextualisierung der Visualisierungen und den in ihnen enthaltenen Daten sowie die Transparenz durch Dokumentation der zugrundeliegenden Datenbankstrukturen, um eine verlässliche Grundlage für die Forschung zu gewährleisten.

Durch den vereinfachten Zugang zu komplexen Zusammenhängen ist es SoNAR (IDH) möglich eine größere Übersicht in der kritischen Datenanalyse zu schaffen und darüber hinaus alle Forschungsprozesse zu unterstützen. Um die digitale Forschungsinfrastruktur aufzubauen und zu verstetigen, muss die Transparenz weiter vorangetrieben werden. Insbesondere die Darstellung von Unsicherheiten und die Dokumentation des Systems sind wichtig. Das System muss Forscher:innen zudem bei der Dokumentation der eigenen Forschung unterstützen (Dokumentation von Forschungsprozess und -ergebnisse, Zitation, Archivierung). Stärkere digitale Kompetenz, Transparenz des Forschungsprozesses, als auch Möglichkeiten einer breiteren Datenpublikation können helfen ein allgemeines Verständnis zu schaffen und produktive Diskurse zu erzeugen.

Anhang

Interviewleitfaden

Leitfaden SoNAR Nutzerstudie – Interviewrunde I

Version 1.4

Zielsetzung:

Evaluation III dient der Überprüfung von AP2-1 (Modellhaftes Forschungsdesign). Dabei wird eine erste Fragerunde durchgeführt, die auf die Erhebung von Daten zur Arbeitsweise von Expert_innen auf dem Gebiet der HNA abzielt. Im Zuge des AP4-3 folgen die Interviews dabei dem verschachtelten Modell nach Munzner (2009). Dabei werden Daten zu folgenden Aspekten des *Nested Model* erhoben:

Ebene 1: Familiarisierung mit der Domäne;

Ebene 2: Verwendung adäquater Daten für die Zielgruppe;

Ebene 3: Konzeption der Visualisierungen;

Ebene 4: Erstellung der Algorithmen für den automatisierten Betrieb der Infrastruktur.

Es wird daher erwartet, über die Datenerhebung erste Schlüsse über Bedarf und Umfeld (z.B. Fachdisziplinen, für die HNA in der Praxis aktuell schon relevant ist) ableiten zu können.

Fragestellung:

Wie sieht der Forschungsprozess anderer ExpertInnen auf dem Gebiet der HNA aus, insbesondere in Bezug auf die Datenerhebung und -analyse? Welche Konsequenzen für SoNAR ergeben sich daraus in der Gegenüberstellung mit dem Modellhaften Forschungsdesign?

Literatur:

Bogner, Alexander; Littig, Beate; Menz, Wolfgang (2014): Interviews mit Experten. Eine praxisorientierte Einführung. Wiesbaden: Springer VS (Lehrbuch). Online verfügbar unter

<https://link.springer.com/content/pdf/10.1007%2F978-3-531-19416-5.pdf>.

Helfferich, Cornelia (2011): Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews. 4. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH Wiesbaden. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-531-92076-4>.

Helfferich, Cornelia (2014): Leitfaden- und Experteninterviews. In: Nina Baur und Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS, S. 559–574.

Vogt, Stefanie; Werner, Melanie (2014): Forschen mit Leitfadeninterviews und qualitativer Inhaltsanalyse. Skript. Fachhochschule Köln, Köln. Fakultät für angewandte Sozialwissenschaften. Online verfügbar unter https://www.f01.th-koeln.de/imperia/md/content/sozialarbeitplus/skript_interviewsqual_inhaltsanalyse.pdf

Grundsätzliches:

- Möglichst wenige geschlossene Fragen, wenn dann ergänzt durch vorgeplante Nachfragen.
- Keine Suggestivfragen!
- Es wird nicht die eigentliche Forschungsfrage direkt gestellt!
- Fragen aussortieren, die auch ohne Hilfe eines Experteninterviews beantwortet werden können (vgl. Bogner et al. 2014, S. 33).

- Grounded Theory-Ansatz der Sättigung (Saturation): „Sättigung heißt, dass keine zusätzlichen Daten mehr gefunden werden können, mit deren Hilfe der Soziologe weitere Eigenschaften der Kategorie entwickeln kann. Sobald er sieht, dass die Beispiele sich wiederholen, wird er davon ausgehen können, dass eine Kategorie gesättigt ist.“ (Glaser und Strauss 1998, S. 69).
- Offene Eingangsfragen und anschließende, antwortabhängige Vertiefungsfragen.
- Von allgemeinen hin zu spezifischen Themenbereichen.
- Angestrebte Maximaldauer jedes Fragenkomplexes festlegen.
- Der Ablauf der Fragen ist bis auf den Einstieg und Abschluss flexibel.
- Es sollten Aussagen zu jedem der Fragekomplexe (Hauptfragen) gemacht werden, das heißt aber nicht, dass alle gesammelten Fragen auch gestellt werden! Vielmehr dient der Leitfaden als Gedächtnisstütze, um das Gespräch auf die angedachten Aspekte (zurück) zu lenken.

Ablauf

Vor den Interviews:

Liegt unterzeichnetes *Informed Consent Dokument* vor?

Neben Zoom-Recording Zusatzaufnahme mit Handy-Aufnahmegerät für den Fall von Datenverlust bei Export oder Konvertierung!

EINLEITUNG

Zunächst einmal herzlichen Dank für Ihre Zeit und Ihre Bereitschaft zum heutigen Interview!

Im folgenden Gespräch möchte ich Sie einladen, uns von Ihrem Vorgehen bei Publikationen zur Netzwerkforschung zu berichten. Besonders interessiert mich dabei, wie Sie methodisch vorgehen.

Warum möchte ich das wissen? Ich bin Sina Menzel und führe im Arbeitspaket der Humboldt-Universität eine iStudie für das SoNAR-Projekt durch.

In diesem Projekt beschäftigen wir uns damit, wie wir vorhandene Daten aus Gedächtnisinstitutionen wie Archiven und Bibliotheken speziell angepasst auf die Bedürfnisse der Historischen Netzwerkanalyse aufbereiten und darstellen können.

Das Interview ist auf ca. 45 min. angelegt. Wie Sie bereits im Vorfeld gelesen und bestätigt haben, werde ich mit Ihrem Einverständnis das **Gespräch aufzeichnen**, sodass ich anschließend auf alle Informationen zugreifen kann. Selbstverständlich werden wir **sämtliche Daten anonymisieren**, sodass weder ein Rückschluss auf Sie persönlich, noch auf Ihre Institution möglich ist. Das heißt konkret, dass bereits die Audiodatei nicht mit Ihrem Namen benannt wird und das Transkript sofort anonymisiert wird.

Haben Sie noch Fragen zum Ablauf, bevor wir starten? Selbstverständlich können Sie auch während des Gesprächs jederzeit Rückfragen stellen.

ABSCHLUSS

Vielen Dank für das Gespräch! Wenn Sie möchten, halten wir Sie natürlich gerne über die Projektergebnisse auf dem Laufenden!

Fragen

Hauptfrage/ Erzählaufforderung	Ergänzungsfrage (Alternative Formulierung)	Aufrechterhaltungsfrage/ Steuerungsfrage	Zeit	Frageziel (Use Case)
ERÖFFNUNG				
1. Sie werden auf einer privaten Feier angesprochen, was Sie beruflich machen. Sie erzählen von Ihrer Tätigkeit und erwähnen Historische Netzwerkanalyse. Was antworten Sie, wenn Sie gefragt werden, warum Sie sich ausgerechnet Netzwerke anschauen?	Wie würden Sie die HNA in drei Sätzen für jemanden erklären, der sie nicht kennt? Wie ist es dazu gekommen, dass Sie die HNA zum Gegenstand Ihrer Forschungsarbeiten gemacht haben?	Was macht die Methode der HNA besonders relevant für Ihre Forschungsfragen?		Gesprächsstart Perspektive auf die Methodik Abgleich mit Einordnung der HNA aus Antrag und AP2-2
FORSCHUNGSINTERESSE				
2. Auf welcher Grundlage entwickeln Sie ihre Forschungsfragen? Bilden Sie Hypothesen?	Wie hat sich die Forschungsfrage/Hypothese zu Paper XY ergeben?			UC301: Erkundung und Hypothesenbildung UC303: Erkennen von Mustern (bird's perspective) Einstiegspunkte in der Datensammlung
DATENERHEBUNG UND DATENANALYSE (mit Aspekten von Bedarf und Umfeld)				

3. Beschreiben Sie bitte den Prozess der Datenauswahl für ihr letztes/aktuelles Projekt in dem Sie den Ansatz der HNA verwendet haben.	Wie beginnen Sie normalerweise, wenn Sie Daten suchen? Stand ein Quellenkorpus oder eine Forschungsfrage am Beginn?	<p>Gibt es eine Go-To Quelle, die Sie immer prüfen, unabhängig von der Fragestellung? Wenn ja, welche und warum?</p> <p>Welche übergreifenden Datenbanken nutzen Sie bevorzugt im Forschungsprozess (z.B. Bibliothekskataloge)? Nutzen Sie diese auch zur Datenerhebung?</p> <p>Wie war der Zugang zu Ihren Daten für Publikation XY? Waren die Quellen öffentlich oder nicht?</p>		<p>Datenquellen</p> <p>Bedarfsanalyse</p> <p>Kenntnis von Quellen</p>
4. Welche Kriterien müssen Daten erfüllen, bevor Sie Sie in eine Datenerhebung einbeziehen?	Was tun Sie, wenn Sie herausfinden möchten, ob eine Quelle Daten enthält, die Sie für Ihre Zwecke nutzen können?	<p>Was machen Sie mit widersprüchlichen oder unvollständigen Daten? Was machen Sie, wenn Daten lückenhaft oder ambig/uncindeutig sind?</p> <p>Was tun Sie, um Daten so zu verdichten, dass Sie für eine Analyse geeignet sind?</p> <p>Woher wissen Sie, dass Ihre Datenerhebung abgeschlossen ist?</p> <p>Was tun Sie, wenn Sie eine Verbindung sehr stark vermuten, aber keine Belege dafür finden?</p>		<p>UC302: Quantifizierung und Hypothesentest</p> <p>UC303: Erkennen von Mustern (bird's perspective)</p> <p>Notwendige Filtermöglichkeiten</p> <p>Kartenansicht</p> <p>Zeitleiste</p> <p>Umgang mit Datenunsicherheiten</p>

5. Welche Bedeutung hatte die HNA in ihrem methodischen Zugriff auf die erhobenen Daten? Haben Sie netzwerkanalytische Verfahren zur Auswertung verwendet?		Nutzen Sie mathematische Kennzahlen der Netzwerktheorie , wie Zentralität? Wenn ja, welchen Stellenwert nehmen diese Werte in Ihrer Datenanalyse und –interpretation ein?		UC304: Gewichtung von Beziehungen Mathematische Kennzahlen Netzwerktheorie Graphanalyse
--	--	--	--	--

BEDARF UND UMFELD

6. Gibt es Schritte in der Analyse von Netzwerken, die Sie besonders mühsam finden? Wenn ja, welche und warum?	Was sind ihrer Meinung nach die größten methodischen aber auch technischen Probleme bei der HNA?	Mit welchen Datenformaten arbeiten Sie für die Auswertung?		Technische Kenntnisse Unterstützungsbedarf in der Auswertung
7. Mit welcher Software haben Sie ihre Netzwerke visualisiert? Nach welchen Kriterien haben Sie diese Tools ausgesucht?	(Wenn vorhanden:) Womit haben Sie die Abbildungen in Ihrem Paper erstellt?	(z.B. VennMaker, Palladio, Gephi, NodeXL, Jupyter Notebooks) Wie importieren Sie die gesammelten Daten in Ihr Analyse-Tool?		Technische Kenntnisse Usability-Aspekte
8. Bibliothekskataloge enthalten eine Menge an Metadaten und Beziehungsinformationen. Haben Sie schon einmal versucht diese für Ihre Forschung zu nutzen?	Welche Such- und Exportfunktionen bräuchten Sie, damit sie Daten aus einem Bibliothekskatalog in eine Datenauswertung einbeziehen können?	Haben Sie spontan konkrete Ideen, wie Sie solche bereits vorhandenen und validierten Daten in Ihrem Forschungskontext nutzen könnten?		Konkrete Bedarfsabfrage

VERSIONIERUNG

9. Wie stellen Sie bei komplexen Netzwerkdarstellungen die Transparenz ihrer Datengrundlage sicher?	Wenn Sie in einer Folgetestreihe noch einmal die Daten ergänzen und aktualisieren würden, wie würden Sie vorgehen?	Haben Sie Erfahrungen mit dynamischen Datenquellen (etwa aus dem Internet)? Wie wissen Sie, welchen Ihrer Datenpunkte Sie aus welcher Quelle haben?		Versionierung Datenspeicherung PIM Zitation und Nachvollziehbarkeit
---	--	--	--	--

ABSCHLUSS

10. Wir haben jetzt über XY gesprochen. Gibt es noch etwas, das Sie gerne ergänzen würden? Haben Sie noch Fragen?				
---	--	--	--	--

Interviewteilnehmer:innen

Personenschlüssel	Position	Disziplin	Geschlecht
1	Wissenschaftliche Mitarbeiterin	Geschichtswissenschaft	w
2	Wissenschaftlicher Mitarbeiter, Postdoc	Kunstgeschichte	m
3	Wissenschaftliche Mitarbeiterin, Postdoc	Geschichtswissenschaft	w
4	Doktorandin	Sozialwissenschaft	w
5	Doktorand	Sozialwissenschaft	m
6	Wissenschaftliche Mitarbeiterin	Wirtschafts- und Sozialwissenschaft	w
7	Doktorandin	Geschichtswissenschaft	w

1 Repositorien/Normdaten/Formate	
1.1 Formate/Sprachen	1.2 Normdaten
1.1.1 BibTeX	1.2.1 Getty
1.1.2 RDF	1.2.2 GND
1.1.3 R	1.2.3 Wikidata
1.1.4 CSV	1.2.4 Geonames
1.1.5 SQL	1.2.5 Googlemaps
1.1.6 ProQuest	1.3 Datenrepositorien
1.1.7 JavaScript	1.3.1 english monastic archive
1.1.8 PythonScript	1.3.2 Germania Sacra
1.1.9 D3	1.3.3 Helvetia Sacra
1.1.10 SVG	1.3.4 CorrespSearch
1.1.11 XML	1.3.5 Wikipedia
1.1.12 Microsoft Excel	1.3.6 LONSEA
1.1.13 Microsoft Word	1.3.7 Kalliope
	1.3.8 ZDB - Zeitschriftendatenbank
2 Datenqualität	
2.1 Probleme	2.2 Anforderungen
2.1.1 Ambiguität und Uneindeutigkeit	2.2.1 Transparenz und Dokumentation

2.1.2 Informationslücken/Vollständigkeit	2.2.2 Standardisierung
2.1.3 Subjektivität des Datenmaterials	2.2.3 Provenienz
2.1.4 (Historie von Daten)	
3 Forschungsprozess	
3.1 Erste Schritte	3.5 Datenmodellierung
3.2 Forschungsfrage	3.6 Datenauswertung
3.2.1 prozesshaft	3.6.1 Mathematische Kennzahlen
3.2.2 theoriegeleitet	3.6.2 Berechnungen im two/one mode Network
3.2.3 datengeleitet	3.6.3 Aufteilung in Teilnetzwerke
3.3 Datenauswahl/-erhebung	3.7 Dokumentation
3.3.1 Recherche OPAC/DB	3.7.1 Versionierung
3.3.2 Arbeit mit analogen Quellen	3.7.2 Dokumentation der Datenfelder
3.3.3 Daten aus Zeitzeugeninterviews	3.7.3 Dokumentation der Datenerhebung
3.3.4 Datenexport, -import aus Repositorien	3.7.4 Auszeichnung von Datenunsicherheiten
3.3.5 Kollaborative Datenerhebung	3.7.5 Dokumentation über Identifikatoren
3.3.6 Daten aus Sekundärnutzung	3.7.6 Dokumentation Sekundär genutzter Datenbanken
3.4 Datenreinigung	3.8 Exploration
3.4.1 durch statistische Berechnung	
3.4.2 durch Intellektuelle Durchsicht	

4 Visualisierung	
4.1 Aktuelle Nutzung	4.2 Serendipität
	4.3 Kritik
5 Community	
5.1 Curriculum	5.4.3 Innsbruck
5.2 Infrastruktur	5.4.4 München
5.2.1 Veranstaltungen	5.4.5 Basel
5.3 Datenpublikation/Data Sharing	5.4.6 Berlin
5.4 Institutionen/Projekte	5.4.7 Heidelberg
5.4.1 Konstanz	5.4.8 Mainz
5.4.2 Trier	5.4.9 Palaio(?) – Projekt
6 Werkzeuge	
6.1 Verwendete Werkzeuge	6.2.2 Freeware
6.1.1 Visone	6.2.3 Support
6.1.2 OpenRefine	6.2.4 Performanz
6.1.3 MySQL	6.2.5 Visualisierung
6.1.4 nodegoat	6.2.6 Metadaten-Abgleich
6.1.5 QGIS	6.2.7 Multimodale Netzwerkanalyse
6.1.6 Stata	6.2.8 Datenmodellierung und -speicherung

6.1.7 VennMaker	6.2.9 Datenexport und -import
6.1.8 NodeXL	6.2.10 Such- und Filterfunktion
6.1.9 Gephi	6.2.11 Statistische Auswertung
6.1.10 neo4j	6.2.12 Kollaboration
6.2 Tools Funktionen +/-	6.3 Wunschfunktion
6.2.1 Usability	6.3.1 Wunsch-Metadaten
7 Netzwerke	
7.1 (Probleme)	7.3.3 Nicht-Information/Unsichtbares
7.2 Netzwerke (Verständnis)	7.3.4 Informelle Beziehungen
7.3 Netzwerkbeziehungen	7.4 Gründe für/gegen HNA
7.3.1 Netzwerkbeziehungen/Metadaten	7.5 Beschreibung Forschungsprojekt
7.3.2 Klassifizierung von Beziehungsarten	7.6 Netzwerkanalyse und historische Forschung