

SoNAR AP 2

Heiner Fangerau, Thorsten Halling, Eva Maria Holly, Michael Schneider

Inhalt:

A. Modellhaftes Forschungsdesign Historische Netzwerkanalyse	1
B. Anforderungskatalog SoNAR	11
C. Forschungsbericht AP2	18

A. Modellhaftes Forschungsdesign Historische Netzwerkanalyse

1. Vorbemerkung

Beziehungen zwischen Menschen bilden das Gewebe sozialer Ordnungen, sie konstituieren Möglichkeiten und Zwänge, sie beeinflussen den Zugang zu sozialem Kapital und damit Handlungs- und Wahloptionen¹. Die Analyse dieser Beziehungen und ihrer Entwicklung ist wesentlich für das Verstehen und Erklären von sozialen Prozessen in ihrer Wirkung auf kulturelle, politische, ökonomische und wissenschaftliche Phänomene. Unter dieser Annahme werden Methoden und Theorien aus der Sozialen Netzwerkanalyse (SNA) zunehmend auf historische Fragestellungen angewandt.² Visualisierungen von Netzwerken bieten im Idealfall eine Erkenntnisprogression durch eine (Re-) Strukturierung relevanter Informationen³.

Das Forschungsdesign der historischen Netzwerkanalyse (HNA) ist immer retrospektiv angelegt, d.h. historische Daten werden aus vorhandenen Quellen extrahiert⁴. Die historisch-hermeneutische Vorgehensweise impliziert in besonderem Maße einen „*zirkulären Arbeitsprozess*“, in dem die Arbeitsschritte in der Forschungspraxis ineinandergreifen, angepasst und wiederholt werden (müssen)[Hermeneutischer Zirkel]. Die Fragestellungen und Interpretationen der Forscher*in leiten dabei maßgeblich den laufenden Analyseprozess, der an die Quellenlage, den Zugang und den Inhalt angepasst wird. Darüber hinaus evozieren die quantifizierenden Elemente der HNA entsprechende Anpassungen von Fragestellungen oder veränderte Anforderungen an Datenquellen, Datenmodelle, Datenerhebung und -

¹ Lin, Nan: Social Capital. A Theory of Social Structure and Action. Cambridge, 2001 (2011)

² Bauerfeld, Daniel/Clemens, Lukas: Gesellschaftliche Umbrüche und religiöse Netzwerke. In: Bauerfeld, Daniel/Clemens, Lukas (Hg.): Gesellschaftliche Umbrüche und religiöse Netzwerke. Analysen von der Antike bis zur Gegenwart. Bielefeld, 2014; Düring, Marten et al. (Hg.): Handbuch Historische Netzwerkforschung. Grundlagen und Anwendungen. Münster, 2016

³ Fangerau H, Hentschel K (2018): Netzwerkanalysen in der Medizin- und Wissenschaftsgeschichte - Zur Einführung. Sudhoffs Archiv 102:133-145.

⁴ Krischel M, Fangerau H (2013): "Historical Network Analysis can be used to construct a social network of 19th-century evolutionists". In Fangerau H, Geisler H, Halling T, Martin W (Hrsg.): "Classification and Evolution in Biology, Linguistics and the History of Science. Concepts, Methods, Visualization", Steiner, Stuttgart 2013, S. 45-65

aufbereitung. Die skizzierte Vorgehensweise bestimmt dabei aber nicht nur den inneren Forschungsprozess der HNA, sondern auch die generelle Frage nach der Eignung des methodischen Zugriffs, die sich aus der genauen Kenntnis von Datenmenge und Datenqualität ergibt.

Die Aussagekraft von historischen Netzwerkanalysen (HNA) ist folglich abhängig vom jeweiligen Forschungsinteresse, der Datenqualität und der (Re-)kontextualisierung der Befunde. Das hier vorgestellte modelhafte Forschungsdesign adressiert typisierte Strategien und zentrale Aspekte der historischen Netzwerkanalyse entlang von konkreten Arbeitsschritten, wie das

- Entwickeln von Fragestellungen und Hypothesenbildung
- Ermitteln von Datenquellen
- Erstellen von Datenmodellen
- Erheben, Aufbereiten und Selektieren von Daten
- Analysieren und Visualisieren von Daten
- Sichern und Publizieren von Daten
- Dokumentieren von Prozessschritten

2. Entwickeln von Fragestellungen und Hypothesenbildung

Die Entwicklung von konkreten Fragestellungen und Hypothesenbildung können in der Historischen Netzwerkanalyse (HNA) in zwei grundsätzliche Vorgehensweisen unterteilt werden: Explorativ-datengeleitet und theorie-/hypothesengeleitet.

2.1 Exploratives bzw. datengeleitetes Vorgehen

Das explorative bzw. datengeleitete Vorgehen zeichnet sich dadurch aus, dass ausgehend von bereits erhobenen Datenbeständen geeignete Fragestellungen entwickelt werden und die HNA als ein möglicher methodischer Zugriff getestet wird.

Charakterisierend für die explorative Vorgehensweise ist die Bereitschaft und Offenheit den Prozess von einer konkreteren Ideenentwicklung bis hin zur definierten Forschungsfrage - „durch die Daten geleitet“ - zu vollziehen. An das Vorgehen ist zugleich die Erwartungshaltung (Hoffnung) geknüpft durch die Daten auf überraschende/unterwartete Phänomene aufmerksam zu werden, die zur Entwicklung und Konkretisierung des Forschungsvorhabens beitragen.

In der Praxis bedeutet dies, dass der/die Forscher*in auf Grundlage der bisherigen Forschung zunächst erste noch vage Ideen entwickelt und grobe Anhaltspunkte festlegt. Durch die explorative Vorgehensweise werden diese Ideen entweder gefestigt und weiter ausgebaut, oder verworfen und nochmals neue Ideen anhand der Eindrücke und Erkenntnisse aus den ersten Abfragen entwickelt. Als Anhaltspunkte beim explorativen Vorgehen können beispielsweise Orte, Personen oder Institutionen dienen, die dem/der Forscher*in im untersuchten Kontext bereits bekannt sind.

Bei den Abfragen überprüft der/die Forscher*in unterschiedliche Kriterien, die folgenden Anforderungen entsprechen:

- Ist eine ausreichend große Menge an validen Beziehungsinformationen enthalten? R051, R052
- Welche Beziehungsinformationen sind zu finden? R053
- Inwiefern sind die Daten aufbereitet (standardisiert, (ein-)eindeutig) und (digital) verfügbar?
- Sind die Provenienzen ausreichend transparent? R052
- Wofür stehen Knoten, wofür Kanten: Was passiert beim Wechsel von Knoten zu Kanten und umgekehrt?
- Lassen sich Knoten für mehrere Entitäten ermitteln, die über gleiche Kanten verbunden werden können und lassen sich Knoten, die für unterschiedliche Entitäten stehen gemeinsam abbilden, d.h. lassen sich mehrere Netzwerkmodi (multimodale Abbildungen) in einem Netzwerk abbilden?

Durch deskriptive Maßzahlen (Centrality, Betweenness etc.) können Verteilungen für eine Menge von Entitäten numerisch ausgedrückt werden, Strukturen werden so erkennbar und können Ausgangspunkt für konkrete Forschungsfragen bzw. die Hypothesenentwicklung sein. Je nach Visualisierungsformat können neben den im Rahmen der SNA erprobten Maßzahlen auch räumliche Anordnungen der Hypothesenbildung dienen. R 036, R047

Zentral für die HNA ist jeweils die Positionierung der erhobenen Maßzahlen und Strukturen in einer chronologischen Ordnung, so dass sich Informationen über Entwicklungen, Entwicklungshemmnisse, Dynamiken in Netzwerkstrukturen etc. erheben lassen.

Anforderungen:

- Welche Algorithmen liegen zugrunde? R038, R039

2.2 Theorie-/Hypothesengeleitetes Vorgehen

Das Theorie-/ bzw. hypothesengeleitete Vorgehen zeichnet sich dadurch aus, dass anhand der Datenbestände Fragestellungen bearbeitet und Hypothesen überprüft werden.

Charakterisierend für dieses Vorgehen ist, dass der/die Forscher*in bereits ein konkretes Erkenntnisinteresse bzw. eine Fragestellung im Vorfeld formuliert hat und die Datenabfragen daran orientiert. An das Vorgehen ist die Erwartungshaltung geknüpft, durch eine tiefgehende Datenanalyse Antworten bzw. konkrete Erkenntnisse mit Blick auf die jeweilige Forschungsfrage zu erzielen.

Der/die Forscher*in hat anhand des Forschungsstandes bereits eine konkrete Fragestellung entwickelt und eine Reihe von Subfragen bzw. Hypothesen aufgestellt, die zu einer systematisierten Datenabfrage führen und den gesamten Forschungsprozess leiten. Bei den Abfragen werden beispielsweise Fragestellungen und Hypothesen zu Ausprägungen von Strukturen im zeitlichen Verlauf, der Identifikation von Rollen und Positionen (Annäherung, Abgrenzung, Polarisierung) oder der Interdependenzen von Netzwerken und Handlungsebenen zugrunde gelegt.

Durch das theorie-/hypothesengeleitete Vorgehen werden die Subfragen beantwortet oder problematisiert und die Hypothesen überprüft.

- Ist eine ausreichend große Menge an validen Beziehungsinformationen enthalten? R051, R052
- Welche Beziehungsinformationen sind zu finden? R053
- Inwiefern sind die Daten aufbereitet (standardisiert, (ein-)eindeutig) und (digital) verfügbar? R031, R032, R046
- Sind die Provenienzen ausreichend transparent? R052

Hieraus ergeben sich weitere Anforderungen an die Analyse und Visualisierung der Daten (vgl.6. Analysieren und Visualisieren von Daten)

3 Ermitteln von Datenquellen

Datenquellen können alle Quellen sein, aus denen sich potentiell netzwerkfähige, d.h. relationale oder relationierbare Daten erheben lassen. Die Suche und Auswahl der Datenquellen orientiert sich am jeweiligen Forschungsinteresse sowie an thematischen, geographischen, zeitlichen und sprachlichen Kriterien. Darüber hinaus stellt die möglichst unkomplizierte Verwertbarkeit der Daten für die computergestützte Analyse ein weiteres Auswahlkriterium dar.

Datenquellen können u.a. über Bibliothekskataloge, Archivportale oder Bibliographische-, Personen- Objekt-, Projekt-, und Volltextdatenbanken ermittelt werden, die je nach Bestand Text-, Sach- audiovisuelle und/oder multimediale Quellen bereithalten, wie z.B.: Urkunden, diverse Aktenbestände, archäologische Fundstücke, digitalisierte Zeitschriften, Tonbänder und Filme etc. Die in den Quellen enthaltenen historischen Daten müssen für die HNA exzerpiert, abstrahiert und in Informationseinheiten überführt werden. Dadurch wird der Informationsgehalt der Quelle reduziert und auf relevante Merkmale für die HNA fokussiert. Innerhalb des Netzwerks können dadurch komplexe Strukturen und Relationsinhalte sichtbar und greifbar werden.

4 Erstellen von Datenmodellen

Für die SNA hat Jansen⁵ vor einigen Jahren festgehalten, dass sich Netzwerkrelationen hinsichtlich des Inhaltes, der Intensität und der Form (gerichtet, ungerichtet, unidirektional, reziprok) unterscheiden lassen. Diese Unterscheidung ist grundsätzlich auch für die HNA sinnvoll. Allerdings müssen aus geschichtswissenschaftlicher Perspektive vor allem bei den

⁵ Vgl. Jansen, Dorothea. Einführung in die Netzwerkanalyse: Grundlagen, Methoden, Forschungsbeispiele. 3, überarb. Aufl. Wiesbaden 2006. S 58 ff.

Relationsinhalten und der Relationsintensität einige zentrale Aspekte zusätzlich berücksichtigt werden.

Grundsätzlich spielt der Faktor Zeit in historischen/geschichtswissenschaftlichen Untersuchung von Netzwerken eine entscheidende Rolle hinsichtlich der Veränderung vermeintlich fester Konstanten oder der Bewertungen von Relationen. Dies wird vor allem durch Untersuchungen deutlich, die Entwicklungen über längere Zeitabschnitte betrachten oder diachrone Vergleiche über große zeitliche Distanzen vornehmen.

Mit Blick auf die Relationsinhalte können - je nach Forschungsfrage - grundlegende Veränderungen gesellschaftlicher, (geo-)politischer und wirtschaftlicher Zusammenhänge erheblichen Einfluss auf die Struktur von Netzwerken haben. Konkret lassen sich diese Veränderungen z.B. an sprachlichen Besonderheiten, Definitionen oder dem Gebrauch und Verständnis von spezifischen Worten, Begriffskonzepten und Denkweisen ablesen. Des Weiteren demonstrieren territoriale Umstrukturierungen wie neue Grenzziehungen oder der Wechsel von politischen Systemen und wirtschaftlichen Grundordnungen Einschnitte, die auf die Netzwerkbildung deutliche Auswirkungen haben können.

Diese kontextuellen Veränderungen vermeintlich fester Konstanten und ihr Einfluss auf die untersuchten Netzwerke müssen daher, ebenso wie die Veränderungen in der (normativen) Bewertung von Relationen (positiv – negativ, eng – distanziert, freudig oder traurig, etc.), gesondert reflektiert werden.

Die Relationsintensität umfasst nach Jansen qualitative Parameter, z.B. die Häufigkeit eines Austausches, sowie quantitative Parameter wie die Wichtigkeit/Bedeutung einer Relation. Letztere sind jedoch an (normative) Bewertungen gebunden, die sich entweder innerhalb der Akteure im Netzwerk zeigen können oder von den Forschenden selbst vorgenommen werden. Über die Zeit hinweg können sich Veränderungen in der Bewertung von Relationen, Motivationen und Handlungen ergeben, beispielsweise wenn sich gesellschaftliche Moralvorstellungen, Werte und Normen geändert haben (z.B. Ehrenrettung durch Duellieren).

Darüber hinaus bilden die zur Verfügung stehenden Quellen eine gesetzte Grundlage für die Untersuchungen historischer Netzwerke, die nicht ohne Weiteres, wie z.B. durch die Generierung neuer Information mittels Befragungen von Akteuren, ergänzt werden kann. Neben Bestandslücken sehen sich Historiker*innen oftmals aufgrund des Untersuchungsgegenstandes und des Untersuchungszeitraums mit unterschiedlichen Quellenarten konfrontiert, deren Spezifika und Informationsgehalt gesondert reflektiert werden müssen. Zudem können sich innerhalb einer bestimmten Untersuchungszeit Quellen verändern, indem beispielsweise durch technische Entwicklungen eine Quellenart als Informationsträger abgelöst und dafür neue etabliert werden (Brief, Telegraphie, Rohrpost, etc.).

Relationsinhalte und Datenquellen der HNA

Eine Grundannahme in der Historischen Netzwerkanalyse besteht darin, dass Netzwerke neben einer zeitlichen auch eine räumliche Dimension haben, in denen die unterschiedlichen Beziehungen abgebildet werden können. Ob die synchrone oder diachrone Darstellung

relevant ist oder, ob eine regionale, transterritoriale (-nationale) oder globale Betrachtungsweise von Interesse ist, hängt zu allererst von der konkreten Forschungsfrage ab. Aber auch die zur Verfügung stehenden Datenquellen können die jeweils angesetzte Perspektive bahnen.

Der Versuch einer systematischen Kategorisierung der Erkenntnisinteressen und der mit ihnen verbundenen Relationsinhalte („Was wird untersucht?“) und Datenquellen („Woran wird untersucht?“) verdeutlicht die Probleme einer trennscharfen Klassifikation. Denn zentrale Relationsinhalte und Datenquellen werden oft mehrdeutig genutzt, oder die Datenquellen bedienen mehrere Relationsinhalte.

Wenn sich das Erkenntnisinteresse beispielsweise auf Machtverhältnisse richtet, die ausgehend von Netzwerken rekonstruiert werden sollen, dann könnten diese Netzwerke soziale Beziehungen, Finanzströme oder Ideen in den Blick nehmen und mit Hilfe von verschiedenen Datenquellen wie Briefen, Schuldscheinen oder Zitaten untersucht werden. Je nach Erkenntnisinteresse können zudem unterschiedliche Relationsinhalte mit stellvertretender Funktion relevant werden. Diese Inhalte treten häufig -aber nicht ausschließlich- dann auf, wenn nicht-physische Relationsinhalte wie Ideen oder (politische und wirtschaftliche) Prozesse und Entwicklungen betrachtet werden, die sich mit Methoden der Netzwerkanalyse nicht direkt untersuchen lassen. Ein naheliegendes Beispiel sind Orte oder Institutionen, die stellvertretend für bestimmte Denkrichtungen („Schulen“) stehen.

Erkenntnisziele und Relationsinhalte besitzen überdies zumeist eine Doppelfunktion. Ob ein Untersuchungsmerkmal in der Funktion Relationsinhalt oder Erkenntnisinteresse erscheint, steht immer in Anhängigkeit zur jeweiligen Forschungsfrage.

Es können beispielsweise Marktentwicklungen (Erkenntnisinteresse) anhand von Innovationen (Haupt-Relationsinhalt) wie der Dampfmaschine oder des e-Motors untersucht werden. Dafür werden weitere Inhalte wie bestimmte Orte, Produktionsstätten und Einsatzgebiete, soziale Beziehungen von Käufern und Verkäufern, etc. in den Blick genommen. Umgekehrt kann die Marktentwicklung aber der Haupt-Relationsinhalt eines anderen Erkenntnisinteresses sein, z.B. können Marktprozesse anhand der Agrarwirtschaft in der Frühen Neuzeit untersucht werden, um herauszustellen wie stark bäuerliche Landwirtschaften in den Geldverkehr eingebunden waren.

Daher erscheint es sinnvoll, von einer Ordnung anhand fester Kategorien abzusehen und stattdessen den Möglichkeitsraum mit Hilfe von „tags“ bzw. einem „tag-Prinzip“ zu strukturieren.⁶

Im Folgenden werden auf dieser Grundlage exemplarisch Relationsinhalte/Erkenntnisinteressen der HNA aufgelistet. Die meisten Relationsinhalte/Erkenntnisinteressen lassen sich prinzipiell noch weiter ausdifferenzieren wie beispielsweise soziale Beziehungen, die sich zunächst in private und nicht-private

⁶ [Das „tag“ ist ein Etikett bzw. eine Auszeichnung und hilft u.a. E-Mails zu strukturieren. Die Rechnung eines Freundes kann nur schwer zugeordnet werden, wenn es ein festes Ordnersystem gibt, in dem z.B. der Ordner „Freunde“ und der Ordner „Rechnung“ existieren. Wird die E-Mail stattdessen mit den „tags“ *Freund* und *Rechnung* versehen ist sie schnell wiederzufinden und abrufbar.]

Kontakte untergliedern lassen. Bei den privaten Kontakten können verwandtschaftliche oder freundschaftliche Beziehungen differenziert werden. Zu den nicht-privaten Kontakten zählen u.a. berufliche Kontakte oder auch flüchtige bzw. funktionale Bekanntschaften an der Bushaltestelle oder bei Behörden. Um die Tabelle möglichst übersichtlich zu halten, wird jedoch von einer weiteren Untergliederung der Kriterien abgesehen.

Tab. 1 Relationsinhalte/Erkenntnisinteressen der HNA (in Auswahl)

Relationsinhalte /Erkenntnisinteresse	Datenquellen
Soziale Beziehungen	<p>Schriftliche Dokumente (öffentlich, offiziell, kirchlich, privat)</p> <ul style="list-style-type: none"> • Akten (politisch, wirtschaftlich, rechtlich etc.) • Korrespondenzen • Schuldscheine • Protokolle • Flugblätter • Listen jegl. Art (von Buchhaltung bis hin zu Feierlichkeiten Einladungen und Tischordnungen) • Amtsblätter • Literatur, Zitate • Landkarten
Objekte	
Institutionen	
Ideen und Wissen	
(Technische) Infrastruktur (Verkehrsnetze)	
Religion	
Stile und „Richtungen“	
Sprache	
Marktentwicklungen	
Finanzströme	
Kapital	
Handel	
Entscheidungen; Entscheidungsprozesse	<p>Medien (hier: Massenmedien)</p> <ul style="list-style-type: none"> • Zeitungen TV, Radio Film, Kino, Internet --> Social Media etc. <p>aber auch Theater, Lesungen, Vorträge</p>
Kunst und Kultur	
Orte	
Macht und Einfluss	<p>Verkehrswege</p> <ul style="list-style-type: none"> • Straßen, Kanäle, Flug- und Schifffahrtswege, Schienennetz • Funkfrequenzen; Kabelnetze, Serverstandorte
Gefühle	
Medien! (können zusätzlich zu Kriterium und Erkenntnisinteresse auch noch Indikator sein)	<p>Sachquellen</p> <ul style="list-style-type: none"> • archäologische Gegenstände (Vasen, Schmuck, Münzen, Werkzeuge) • Photographien • Kunst (z.B. Bilder, Gemälde, Statuen etc.) • Einrichtungsgegenstände • Kleidung • Instrumente (musikalisch und medizinisch) • Maschinen und Werkzeuge • Substanzen

Aus diesem Ansatz ergeben sich unterschiedliche Anforderung an die Analyse und Visualisierung der Daten (siehe 6. Analysieren und Visualisieren von Daten)

5 Erheben, Aufbereiten und Selektieren von Daten

Anders als in der SNA, die ihre Daten etwa durch standardisierte Befragungen generiert, ist die Datenerhebung in der HNA nicht in gleichem Maße vorstrukturiert. Gleichzeitig können verschiedenste Informationen (Relationsinhalte s.o.) als Stellvertreterinformationen z.B. für die Vernetzung von Institutionen, den Wissenstransfer oder für soziale Beziehungen herangezogen werden. Eine gängige Hypothese zu letztgenannten, die eine solche Bündelung von Beziehungsinformationen verlangt, geht davon aus, dass mit der Zahl der möglichen unterschiedlichen Kontaktpunkte zweier Personen die Wahrscheinlichkeit ansteigt, dass beide in Beziehung zueinander standen.⁷

Für die HNA sind demnach sowohl Daten relevant, die a) direkte Beziehungsinformationen enthalten, als auch solche die b) indirekte bzw. stellvertretende Informationen über Beziehungen liefern. Dabei kann es sich um Normdaten, aber auch andere Daten, Metadaten oder Volltexte handeln.

Zu Veranschaulichung wird dies am Beispiel der sozialen Beziehungen detaillierter ausgeführt:

a) direkte Beziehungsinformationen

Zum einen können soziale Beziehungen unmittelbar angenommen werden, wenn familiäre oder berufliche Verbindungen in den Daten explizit ausgewiesen sind. Solche Informationen werden u.a. in der Gemeinsamen Normdatei (GND) bereitgestellt.

Die Verwendung von Normdaten für die HNA bietet im Hinblick auf die Datenqualität diverse Vorteile für die Forschenden: Diese betreffen zum einen die Disambiguierung von Personennamen, die in der Datenaufbereitung eine zentrale Herausforderung darstellt. Zusätzlich zur ein-eindeutigen Zuweisungen, kann davon ausgegangen werden, dass die Daten aus standardisierten Pflichtfeldern immer abrufbar und zusätzlich Informationen in weiteren Feldern und unterschiedlichen Schreibweisen enthalten sind. Die Herkunft der Daten ist nachvollziehbar und nachprüfbar und gerade bei öffentlich zugänglichen Daten ist der unbeschränkte Zugang im Gegensatz zu privaten Anbietern gewährleistet.

Zum anderen liefern auch nicht normierte Daten aus Aufsatzdatenbanken, Archivdatenbanken etc., Informationen zu direkten sozialen Beziehungen, die beispielsweise anhand von Briefen (Adressat – Adressant/Verfasser, Herausgeberschaften oder Co-Autorenschaft) extrahiert werden können. Hier kann davon ausgegangen werden, dass der Verfasser und der Adressat eines Briefes oder Autoren, die gemeinsam ein Buch verfassten, in einem sozialen Verhältnis zueinander standen. R016; R022

b) Indirekte bzw. stellvertretende Beziehungsinformation

Anhand von Berufsbezeichnungen, Orten (Geburts-, Sterbe-, Wirkungsorte), Veranstaltungen (Feiern, Kongresse), Zugehörigkeiten zu bestimmten Institutionen (Vereine, Parteien

⁷ Alexander Rausch: Bimodale Netzwerke. In: Stegbauer, C./Häußling, R. (Hrsg.) Handbuch Netzwerkforschung. VS Verlag für Sozialwissenschaften S. 421-432.

Universitäten, Verbänden) oder Schlagworten/Sachbegriffen können unterschiedlichen Kontaktpunkte von Akteuren identifiziert werden, anhand derer auf soziale Beziehungen geschlossen werden kann. Die Daten enthalten also stellvertretende Informationen über mögliche soziale Interaktionen, die ebenso wie die direkten Beziehungsinformationen aus Norm-, Metadaten und Volltexten generiert werden.

Der grundlegende Indikator für eine mögliche soziale Beziehung sind übereinstimmende Lebensdaten der Akteure.⁸ Ausgehend davon lässt sich anhand von Stellvertreterinformationen auf die Plausibilität bzw. den Grad an Wahrscheinlichkeit einer sozialen Beziehung schließen. Je mehr Überschneidung bzw. mögliche Kontaktpunkte zwischen den Akteuren festzustellen sind, umso höher ist auch die Wahrscheinlichkeit, dass eine direkte soziale Beziehung zwischen diesen existiert(e). Die Wahrscheinlichkeit bzw. welche Kontaktpunkte erfüllt sind, sollte den Forschenden transparent angezeigt werden; Zumal die Möglichkeit einer sozialen Beziehung bei einem oder mehreren gemeinsamen Wirkungsorten im gleichen Zeitraum plausibler erscheint als lediglich ein gemeinsamer Sterbeort (R006). An dieser Stelle drängt sich die Frage nach einer Gewichtung der Stellvertreterinformationen geradezu auf. Eine valide Gewichtung erfordert zunächst eine umfassende mathematische Prüfung der stellvertretenden Indikatoren. R017-R022

Intellektuelle Beziehungen

Eine Sonderform stellen intellektuelle Beziehungen dar. Strenggenommen sind diese keine reinen sozialen Beziehungen, da die grundlegende Anforderung übereinstimmender Lebensdaten nicht zwingend erfüllt sein muss. Gleichzeitig besteht eine (intellektuelle) Verbindung zwischen Personen, die zumeist durch eine dritte Person konstruiert wurde.

Vorstellbar sind beispielsweise Sammlungen wie Fotosammlungen und Karteikäsen, in denen verschiedene Akteure von dem /der Bestandsbildner*in in eine (thematische) Relation gesetzt wurden, ohne dass zwischen diesen jemals eine soziale Beziehung bestand; gleiches gilt u.a. für schriftliche Konversation z.B. Briefe oder audiovisuellen Aufzeichnungen z.B. Interviews mit Zeitzeugen. Das aus wissenschaftlicher Perspektive wohl nahe liegende Beispiel sind Zitationen in Texten. Hier werden vom Autor*in durch die Zitationen bestimmte Personen mit ähnlichen oder völlig gegensätzlichen Ansichten explizit in Relation gesetzt, unabhängig davon, ob diese sich gekannt haben oder nicht. R005

Anforderungen:

- Wofür stehen Knoten, wofür Kanten: Was passiert beim Wechsel von Knoten zu Kanten und umgekehrt? R029; R030
- Sind multimodale Abbildungen möglich? R037
- Anhand welcher Quellen werden die Beziehungsinformationen angezeigt? R029, R030, R046

⁸ Da familiäre Beziehungen eine besondere Form sozialer Beziehungen darstellen, sollten diese ggf. auch unabhängig von sich überschneidenden Lebensdaten angezeigt werden.

- Wie werden die Zentralitätsmaße berechnet? R037
- Wie werden die Algorithmen transparent gemacht? R038
- Wie können Algorithmen ausgewählt werden? R039
- Sind alle Datenfelder ausgefüllt, um auf die relevanten Informationen zugreifen zu können? R052
- Wie werden direkte und wahrscheinliche soziale Beziehungen gekennzeichnet, um sie bei Bedarf separat anzeigen zu lassen. R053
- Wie wird die Wahrscheinlichkeit einer sozialen Beziehung den Forschenden transparent aufgezeigt? R006
- Werden soziale und intellektuelle Beziehungen separat angezeigt? R004; R005
- Wie wird der/die Bestandsbildner*in innerhalb der intellektuellen Beziehungen gekennzeichnet?

6 Analysieren und Visualisieren von Daten

Aus dem oben geschilderten Kapiteln ergeben sich konkrete Anforderungen an die Darstellung und Visualisierung, die für die eine solide Analyse historischer Netzwerke unabdingbar sind:

Anforderungen:

1. Kann die Darstellungsart (Datenquellen oder Relationsinhalte) gewählt werden. R023
2. Kann die Darstellungsart (Datenquellen oder Relationsinhalte) für unterschiedliche Zeiträume (im Sinne von Zeitabschnitten) gewählt werden.
3. Können Bewertungen von Relationen unterschiedlich dargestellt werden?
4. Können in der Darstellung verschiedene Relationsinhalte kombiniert und hervorgehoben werden? R034
5. Welche Daten können als Stellvertreter für welche Untersuchungseinheiten genutzt werden?
6. Wie können welche Daten als Knoten und Kanten für Netzwerkanalysen eingesetzt werden?
7. Wie kann eine zeitliche Dimension dargestellt werden? R048
8. Lassen sich Knoten für mehrere Entitäten ermitteln, die über gleiche Kanten verbunden werden können und lassen sich Knoten, die für unterschiedliche Entitäten stehen gemeinsam abbilden, d.h. lassen sich mehrere Netzwerkmodi in einem Netzwerk abbilden? (Multimodale Abbildungen) R037

B. Anforderungskatalog SoNAR

Heiner Fangerau, Thorsten Halling, Eva Maria Holly, Michael Schneider

Tab. 2 Anforderungen an die HNA im Allgemeinen und an SoNAR im Speziellen

ID	Titel	Beschreibung
R001	Soziale Relationen in den Normdaten suchen	Forscher*in möchte, dass alle in den Normdaten enthaltenen direkten Beziehungsinformationen (familiäre, berufliche Relationen, Affiliationen zu Körperschaften) im Datenmodell abgebildet werden, um direkte soziale Beziehungen abfragen zu können.
R003	Soziale Relationen in den Metadaten suchen	Forscher*in möchte, dass alle in den Metadaten der Ressourcen enthaltenen Beziehungsinformationen als abgeleitete soziale Beziehungen im Datenmodell abgebildet werden, um direkte soziale Beziehungen abfragen zu können.
R004	Nicht-soziale ("intellektuelle") Relationen von sozialen unterscheiden	Forscher*in möchte, dass die in den Ressourcen enthaltenen Informationen zu nicht-sozialen bzw. durch Dritte vermittelten ("intellektuellen") Relationen im Datenmodell von direkten sozialen Relationen unterschieden werden, um soziale und intellektuelle Netzwerke abbilden zu können
R005	Intellektuelle Relationen ausdifferenzieren	Forscher*in möchte nicht-soziale ("intellektuellen Relationen") (z.B. Co-Nennungen in "Zettelkästen") und nicht explizite soziale Relationen (z.B. in Briefen "erwähnt und behandelt") mit weiteren Bedingungen (überlappende Lebensspanne, gleicher Beruf, gleicher Geburts-, Wirkungsort, gleiche Affiliation (z.B. Universität) verknüpfen, um potentielle soziale Beziehungen identifizieren zu können
R006	potentielle soziale Relationen identifizieren	Forscher*in möchte Personen, denen vergleichbare Attribute zugeordnet sind (überlappende Lebensspanne, gleicher Beruf, gleicher Geburts-, Wirkungsort, gleiche Affiliation (z.B. Universität) gruppieren können, um potentielle soziale Relationen identifizieren können ("umso mehr Übereinstimmungen umso wahrscheinlicher ist eine soziale Relation")
R007	Nach Personen suchen	Forscher*in möchte nach Personen suchen können, um Personennetzwerke zu identifizieren.
R008	Nach Schlagworten suchen	Forscher*in möchte nach Begriffen suchen können, um themenbasierte soziale Netzwerke zu identifizieren.
R009	Nach Körperschaften suchen	Forscher*in möchte nach Körperschaften suchen können, um Institutionennetzwerke zu identifizieren.
R010	Nach Orten suchen	Forscher*in möchte nach Orten suchen können, um geographische Schwerpunkt von Personennetzwerken zu identifizieren.

R011	Mit Hilfe von Trunkierte suchen	Forscher*in möchte trunkiert suchen können (mind. 3 Buchstaben + *), um größere Themenfelder mit einer Abfrage zu erfassen
R012	Anwendungsszenarien anzeigen lassen	Forscher*in möchte wissen, welche HNA-Anwendungsszenarien in SoNAR zu realisieren sind, um die eigenen Forschungsfragen damit abzugleichen oder ggf. anzupassen
R013	Provenienz der Ausgangsdaten anzeigen lassen	Forscher*in möchte wissen, welche Datenrepositorien aktuell dem Datenset zu Grunde liegen, um die Relevanz für die eigene Forschungsfragen beurteilen zu können
R014	Evolution der Ausgangsdaten darstellen lassen	Forscher*in möchte wissen, welche Veränderungen im Datenbestand erfolgt sind, um erstens die Notwendigkeit einer erneuten Abfrage einschätzen bzw. um zweitens Veränderungen der Ergebnisse erklären zu können
R015	Aktualität der Ausgangsdaten anzeigen lassen	Forscher*in möchte wissen, wann die Ausgangsdaten zuletzt aktualisiert wurden, um ggf. Widersprüche zu Direktabfragen in den Ausgangsrepositorien aufzuklären
R016	Relationen Personen zu Personen anzeigen lassen	Forscher*in möchte eine Liste mit den Personen, die mit der gesuchten Person verbunden sind, um unabhängig von einer graphischen Netzwerkdarstellung, die Relationen übersichtlich/tabellarisch zur Verfügung zu haben, auch um unabhängig von Sonar mit diesen Daten weiter arbeiten zu können
R017	Relationen Personen zu Körperschaften anzeigen lassen	Forscher*in möchte eine Liste mit den Körperschaften, die mit der gesuchten Person verbunden sind, um unabhängig von einer graphischen Netzwerkdarstellung, die Relationen übersichtlich/tabellarisch zur Verfügung zu haben, auch um unabhängig von Sonar mit diesen Daten weiter arbeiten zu können
R018	Relationen Personen zu Körperschaften zu Personen anzeigen lassen	Forscher*in möchte eine Liste mit den Personen, mit der die gesuchte Person über eine Körperschaft verbunden ist, , um unabhängig von einer graphischen Netzwerkdarstellung, die Relationen übersichtlich/tabellarisch zur Verfügung zu haben, auch um unabhängig von Sonar mit diesen Daten weiter arbeiten zu können
R019	Relationen Personen zu Sachschlagworten anzeigen lassen	Forscher*in möchte eine Liste mit den Sachbegriffen, die mit der gesuchten Person verbunden sind, um unabhängig von einer graphischen Netzwerkdarstellung, die Relationen übersichtlich/tabellarisch zur Verfügung zu haben, auch um unabhängig von Sonar mit diesen Daten weiter arbeiten zu können

R020	Relationen Personen zu Konferenzen anzeigen lassen	Forscher*in möchte eine Liste mit den Konferenzen, die mit der gesuchten Person verbunden sind, um unabhängig von einer graphischen Netzwerkdarstellung, die Relationen übersichtlich/tabellarisch zur Verfügung zu haben, auch um unabhängig von Sonar mit diesen Daten weiter arbeiten zu können
R021	Relationen Personen zu Geographika anzeigen lassen	Forscher*in möchte eine Liste mit den geographischen Angaben, die mit der gesuchten Person verbunden sind, um unabhängig von einer graphischen Netzwerkdarstellung, die Relationen übersichtlich/tabellarisch zur Verfügung zu haben, auch um unabhängig von Sonar mit diesen Daten weiter arbeiten zu können
R022	Relationen Personen zu Publikationen anzeigen lassen	Forscher*in möchte eine Liste mit den Publikationen, die mit der gesuchten Person verbunden sind, um unabhängig von einer graphischen Netzwerkdarstellung, die Relationen übersichtlich/tabellarisch zur Verfügung zu haben, auch um unabhängig von Sonar mit diesen Daten weiter arbeiten zu können
R023	Beziehungstypen anzeigen lassen	Forscher*in möchte eine Liste mit den unterschiedlichen Beziehungstypen, um die Art der Vernetzung zu verstehen
R024	Beziehungszusatzangaben anzeigen lassen	Forscher*in möchte eine Liste mit den Beziehungszusatzangaben
R025	Fortschritt bei den Abfrageverarbeitung anzeigen	Forscher*in möchte einen Hinweis (ggf. mit Begründung) bekommen, sollte das Laden der Daten länger als 5/X Sekunden dauern (evtl. mit ungefähre Zeitangabe, die das Laden in Anspruch nimmt), um ggf. die Abfragenkomplexität zu reduzieren
R026	Egozentrierte Netzwerke (einfach) darstellen	Forscher*in möchte eine egozentrierte Darstellung von Personen
R027	Egozentrierte Netzwerke (komplex) darstellen	Forscher*in möchte eine egozentrierte Darstellung von Personen, deren Relationen untereinander ebenfalls angezeigt werden, um die Einbettung einer Person besser zu verstehen und um ggf. Unterstützernetzwerke zu identifizieren
R027	Netzwerke zu mehreren Körperschaften darstellen	Forscher*in möchte eine Darstellung von Personen die über mehrere Körperschaft miteinander verbunden sind, um die Verbindungen der gesuchten Person und anderer Personen in mehreren Institutionen zu identifizieren

R028	Netzwerke darstellen	Körperschaften	Forscher*in möchte eine Darstellung von Personen die über eine Körperschaft miteinander verbunden sind, um die Rolle von bestimmten Institutionen für Relationen untersuchen zu können
R029	Transparenz sicherstellen	der Knoten	Forscher*in möchte zu jedem Knoten eine Beschreibung der potentiell vorhandenen Merkmale und die jeweilige Datenprovenienz, um einen größtmöglichen Überblick zu den Daten zu erhalten
R030	Transparenz sicherstellen	der Kanten	Forscher*in möchte zu jeder Kante eine Beschreibung der potentiell vorhandenen Merkmale und die jeweilige Datenprovenienz, um einen größtmöglichen Überblick zu den Daten zu erhalten
R031	Knoten verlinken		Forscher*in möchte von jedem Knoten aus eine direkte Verlinkung auf den Datensatz in den Ausgangsdaten, um die Quelle prüfen zu können
R032	Kanten verlinken		Forscher*in möchte von jeder Kante aus eine direkte Verlinkung auf den Datensatz in den Ausgangsdaten, um die Quelle prüfen zu können
R033	Kantenstärke anzeigen lassen		Forscher*in möchte anhand der Kantenstärke auf die "Qualität" der Beziehung schließen können z.B.: über Ressourcen (viele Briefe) suggeriert stärkeren Kontakt (ob dieser dann positiv oder negativ zu bewerten ist vermutlich nicht abbildbar)
R034	Kanten differenziert anzeigen		Forscher*in möchte die unterschiedlichen Relationsarten einer Kante (z.B. familiären/berufliche Relationen) auch auf separaten Netzwerken anzeigen können, um durch Anzahl und Art der Relation auf die Qualität/Intensität der einzelnen Beziehung schließen zu können (vgl. NDB)
R035	Egozentrierte Netzwerke erweitern können		Forscher*in möchte ein egozentriertes Netzwerk sukzessive um die vermittelten Beziehungen erweitern können (vgl. NDB)
R036	Beziehungen gewichten		Forscher*in möchte mit Hilfe von Zentralitätsmaßen das Netzwerk gewichten (Hypothese: Akteure mit vielen Kontakten verfügen über mehr Ressourcen, Informationen, Macht/Prestige)
R037	Zentralitätsmaße berechnen		Forscher*in möchte Zentralitätsmaße berechnen lassen können, um Netzwerkbeziehungen gewichten und damit qualitativ interpretieren zu können.

R038	Verwendete Algorithmen transparent darstellen	Forscher*in möchte wissen, welcher Algorithmus der Gewichtung zu Grunde liegt, um dessen Arbeitsweise prüfen zu können und dessen Eignung vor dem Hintergrund der eigenen Forschungsfrage beurteilen zu können
R039	Algorithmus auswählen	Forscher*in möchte unter gängigen Algorithmen auswählen können, um die Gewichtung vor dem Hintergrund der eigenen Forschungsfrage zu variieren
R040	„Vermittler/Hubs“ identifizieren	Forscher*in möchte so genannte "weak ties" in Netzwerken angezeigt bekommen (Hypothese: weak ties wirken als Brücken zwischen Wissensfeldern/Personen)
R041	statischen Zeitfilter einbauen	Forscher*in möchte Netzwerke in Zeitschnitten visualisieren können, um Veränderungen im Zeitverlauf, speziell vor dem Hintergrund von spezifischen Zeitpunkten analysieren zu können
R042	dynamischen Zeitfilter einbauen	Forscher*in möchte Netzwerke dynamisch visualisieren können, um Veränderungen im Zeitverlauf analysieren zu können
R043	Lebensdatenfilter der Personen im Netzwerk einbauen	Ein Filter um anhand der Lebensdaten der Personen im Netzwerk nach bestimmten Abschnitten um z.B. Verbindungen auf das Erwachsenenalter eingrenzen zu können.
R044	Raumfilter einbauen	Forscher*in möchte über einen Raumfilter verfügen, um geographische Schwerpunkte von Beziehungsnetzwerken abbildbar machen können.
R045	Merkmalsausprägungen als Filter für Graphen	Forscher*in möchte Merkmalsausprägungen als Filter für Graphen: z.B. Geschlecht, Alterskohorte, Herausgeberschaft, Beruf, Affiliationen, etc.
R046	Dokumente verlinken	Forscher*in möchte per Mausklick auf Originaldokumente/Volltexte zugreifen können, um die angezeigten Relationen validieren und kontextualisieren zu können
R047	räumliche Beziehungen darstellen	Forscher*in möchte verschiedene Optionen der räumlicher Darstellung (z.B. als Landkarte)
R048	zeitliche Dimensionen darstellen	Forscher*in möchte verschiedene Optionen der zeitlichen Dimensionen (Z.B. Knotengröße für Zahlen, z.B. Baumringe für Jahre?)
R049	Workspace zur Verfügung stellen	Forscher*in möchte Zwischenergebnisse in einem Workspace ablegen können, um in mehreren Sitzungen in SoNAR abfragen zu können

R050	Zwischenergebnisse versionieren	Die Zwischenergebnisse sollen versioniert/datiert sein ebenso wie die verwendete SoNAR Version für Visualisierung
R051	Ergebnisse versionieren	Forscher*in möchte eine Versionierung der Ergebnisse, um .
R052	Anzahl der Datensätze für jede Abfrage anzeigen	Forscher*in möchte statistische Angaben zu den zu Grunde liegenden Datensätzen, gegliedert nach Datenrepositorien, um die jeweilige Quellengrundlage einschätzen zu können
R053	Anzahl der belegten beziehungsrelevanten Datenfelder anzeigen	Forscher*in möchte statistische Angaben zu den tatsächlich belegten beziehungsrelevanten Datenfeldern, um die Frage der fehlenden Daten besser einschätzen zu können
R055	Reproduzierbarkeit sicherstellen	Die für die Visualisierung verwendeten Daten sollen als Datendownload zur Verfügung stehen können, die Reproduzierbarkeit der Visualisierungsergebnisse sicherzustellen.
R056	Versionierung der zur Verfügung gestellten Daten	Die verwendete Daten sollen eindeutig versioniert sein
R057	Druckfähiges Bild ermöglichen	Forscher*in möchte die Ergebnisse der HNA als druckfähiges Bild exportieren können, um sie in Publikationen integrieren zu können
R058	druckfähiges Bild versionieren	Forscher*in möchte eine automatische Versionierung des Bildes und ggf. den Namen/Version von SoNAR
R059	Exportfähige Datenformate bereitstellen	Forscher*in möchte die Ergebnisse der Abfrage als editierbaren Datensatz exportieren können, um sie mit eigenen Daten zu kombinieren
R060	exportierten Datensatz versionieren	Forscher*in möchte eine Versionierung/Datierung des exportierten Datensatz unter Angabe der SoNAR Version
R061	Datensatz beschreiben	Forscher*in möchte wissen, welche Arten von Beziehungsinformationen im Datensatz enthalten sind, um die Aussagefähigkeit und ggf. die fehlenden Relationstypen für die Interpretation der Daten berücksichtigen zu können.
R062	Multimodale Netzwerke anzeigen	Forscher*in möchte Knoten für mehrere Entitäten ermitteln, die über gleiche Kanten verbunden sind, bzw. Knoten, die für unterschiedliche Entitäten stehen gemeinsam abbilden, um mehrere Netzwerkmodi in einem Netzwerk abzubilden?

C. Forschungsbericht AP 2 Heinrich-Heine-Universität

Heiner Fangerau, Thorsten Halling, Eva Maria Holly, Michael Schneider

1. Vorbemerkung

Mit mehreren Forschungstests wurde die Eignung des modellhaften Forschungsdesigns überprüft. Sie wurden nacheinander durchgeführt, damit Resultate und Erfahrungen aus dem ersten Test durch Optimierung der Datenbereitstellung (AP1) und der -visualisierung (AP3) bereits in den zweiten Test einfließen konnten. Wie zu erwarten, stießen die beiden Forschungstests auf vergleichbare Probleme: fehlende oder fehlerhafte Daten verzerrten die Ergebnisse und hätte zu Fehlschlüssen führen können.

Die Forschungstests waren entscheidend dafür, das modellhafte Forschungsdesign methodenzentriert und offen für möglichst breite Forschungsfragen und -themen zu erarbeiten. Hierfür untersuchten die Tests verschiedene fachwissenschaftliche Forschungsthemen mit sehr ähnlichen methodischen Ansätzen: Entlang des aktuellen (historiografischen) Forschungsstands wurden für die Forschungstests Hypothesen zur Wirkung von sozialen Beziehungen auf beobachtete Phänomene gebildet. Diese wurden mit Ergebnissen der Forschungstechnologie SoNAR (IDH) verglichen. Die Resultate wurden eingeordnet und bewertet: Als Bewertungskategorien dienen Gleichheit, Ungleichheit, Unbrauchbarkeit oder, im Idealfall, eine neue Erkenntnis.

Der hier vorgestellte Forschungsbericht adressiert Herausforderungen und Lösungsstrategien für zentrale Aspekte der historischen Netzwerkanalyse entlang der konkreten, im modellhaften Forschungsdesign skizzierten Arbeitsschritte, wie das

- Entwickeln von Fragestellungen und Hypothesenbildung
- Anwendungsbezogen Analyse der verwendeten Datenrepositorien
- Erstellen, Prüfung und Anpassung von Datenmodellen
- Erheben, Aufbereiten und Selektieren von Daten
- Analysieren und Visualisieren von Daten
- Sichern und Publizieren von Daten
- Dokumentieren von Prozessschritten

Die historisch-hermeneutische Vorgehensweise implizierte einen *zirkulären Arbeitsprozess*, in Hinblick auf die Zusammenarbeit mit AP 3 und AP 1. Die Interpretationen der Ergebnisse der Fallstudien durch die/den Forscher*in evozierten entsprechende Anpassungen bzw. veränderte Anforderungen an Datenquellen, Datenmodelle, Datenerhebung und -aufbereitung.

2. Entwickeln von Fragestellungen und Hypothesenbildung

Die Entwicklung von Fragestellungen und die Hypothesenbildung erfolgte vor dem Hintergrund langjähriger wissenschafts- und medizinhistorischer Forschungsschwerpunkte der Projektgruppe der Heinrich-Heine-Universität (AP2). Sie basierte daher auf spezifischen Kenntnissen einerseits auf der Forschungsliteratur als auch der Datenquellen zum jeweiligen Thema, andererseits auf ersten strukturellen Kenntnissen der Datenrepositorien von SoNAR.

Im Mittelpunkt des ersten Forschungstests stand die Entwicklung eines medizinisch-physiologischen Forscherkollektivs im 19. und frühen 20. Jahrhunderts. Basis waren Personen- und Sachindizes einschlägiger Forschungsarbeiten zum Untersuchungsgegenstand. Gefragt werden sollte nach sozialen, institutionellen und intellektuellen Relationen (z.B. verbindende Themen, Co-Nennungen). Darüber hinaus sollte auf Grundlage digitalisierter Zeitungsvolltexte (ZEFYS) die öffentliche Wahrnehmung von Forschern und ihrer Themen („Public Engagement in Science“) in den genannten Epochen analysiert werden.

Im zweiten Forschungstest wurden Hypothesen über die Wirkung deutscher und österreichischer Nationalökonominnen in der zweiten Hälfte des 19. und der ersten Hälfte des 20. Jahrhunderts gebildet, u.a. zu den Unterstützernetzwerken der verschiedenen Berufenen Max Webers. Der Untersuchungsgegenstand grenzte sich inhaltlich deutlich vom ersten Forschungstest ab, verfolgte aber ein vergleichbares historisch-hermeneutisches Erkenntnisinteresse: die Identifikation des Zusammenwirkens von Akteuren im Hinblick auf die Mechanismen disziplinärer Evolutionen, ihre öffentliche Wahrnehmung und ihre Wirkung auf politische Entscheidungsprozesse.

Um die im Projektantrag geforderten iterativen Prozesse zur Entwicklung von Anforderungen an das Datenmodell (AP1) und die Visualisierung (AP3) erforderte kleinere Pretests, die wiederum zur Entwicklung neuer Fragestellungen führten. Diese Vorgehensweise diente somit nicht nur der Überprüfung der Resultate, sondern zugleich der Anpassung von Anforderungen an das Datenmodell, die Visualisierung sowie die Datenrepositorien.

Sie ist damit durchaus vergleichbar mit einem modellhaften Vorgehen in der HNA, das ausgehend von einem Forschungsinteresse die zur Verfügung stehenden Daten auf ihre Eignung für eine netzwerkanalytische Auswertung prüfen muss. Im historisch-hermeneutischen Forschungsprozess unterliegen Fragestellungen und Hypothesen einer zyklischen Bewertung und Anpassung.

Die von AP2 durchgeführten Fallstudien umfassten sowohl explorativ-datengeleitet als auch Theorie-/Hypothesengeleitete Vorgehensweisen.

2.1 explorativ-datengeleitet

Deskriptive Maßzahlen drücken Verteilungen für eine Menge von Entitäten numerisch aus, so dass Strukturen so sichtbar werden und Ausgangspunkt für konkrete Forschungsfragen bzw.

die Hypothesenentwicklung sein können. Am Fallbeispiel „Neurologen-Netzwerke im Nationalsozialismus“ wurde nach institutionellen und inhaltlichen Clustern unter Neurologen/ vor, während und nach der Zeit des NS gefragt.

Das Theorie- bzw. Hypothesengeleitete Vorgehen beinhaltet einen Abgleich eigener fachwissenschaftlichen Forschungsergebnisse zu konkreten Fragestellungen mit Abfragen in SoNAR.

2.2 Theorie- bzw. Hypothesengeleitet

Folgenden Fragestellungen standen im Fokus:

2.2.1 Konkrete Ausprägung einer Struktur mit der Hypothese, dass der Aufbau stabiler Ordnungen mit Normen stellvertretend anhand von Personennetzwerken im Zeitverlauf untersucht werden können.

Anhand der Fallstudie „Disziplinenbildung in der Urologie“ wurde der konkreten Frage nachgegangen, wann sich die Urologie in Deutschland als eigenständige medizinische Fachdisziplin etabliert hat?

2.2.2 Identifizieren von Rollen und Positionen mit der Hypothese, dass ein Wandel durch Selektionsprozesse stellvertretend anhand von Personennetzwerken im Zeitverlauf untersucht werden, d.h. beispielsweise: „Wer bleibt, wer verschwindet?“ oder „Wer ist zentral, wer ist Broker (graue Eminenz)“.

Anhand der Fallstudie „Cholera Bekämpfung: Experimentelle Hygiene vs. Bakteriologie“ wurde untersucht, welche Bedeutung den wissenschaftlichen Netzwerken in der Auseinandersetzung dieser zwei Konzepte zukam?

2.2.3 Verteilung und Austausch von Informationen und Ressourcen mit der Hypothese, dass Ressourcen (hier wissenschaftliche Anerkennung) eine zentrale Rolle für Machtzuwachs und Machtverlust spielen.

Anhand der Fallstudie „Nobelpreis im wissenschaftlichen Diskurs“ wurde die Bedeutung des Nobelpreises in und außerhalb der scientific community analysiert.

Anhand der Fallstudie zu Robert Liefmann wurden intellektuelle Netzwerke entlang von Rezensionen und Forschungsdiskussionen nachgezeichnet.

2.2.4 Egozentrierte Netzwerke (familiär/ beruflich) mit der Hypothese, dass in egozentrierten Netzwerke berufliche und private Bündnisse in zeitlicher und geographischer Dimensionierung abgebildet werden können.

Die Fallstudie „Wilhelm Roux – wechselnde Bündnisse in den Wissenschaften“ untersuchte, ausgehend von den von Roux in seiner Autobiographie hervorgehoben Kollegen, die ihn schon immer seine wissenschaftlichen Positionen unterstützt hätten, welche wechselnden Bündnisse in dem egozentrierten Netzwerk von Wilhelm Roux identifiziert werden können?

Die Fallstudie zu „Max Weber“ analysierte relevante Netzwerke und Akteure im Kontext seiner ersten Berufung und damit, ob quantifizierenden Netzwerkanalysen die Frage nach der Bedeutung von akademischen und familiären Verbindungen im Karriereverlauf eines Wissenschaftlers unterstützen können.

Eine geplante Fokussierung auf Themencluster zur öffentlichen Wahrnehmung von Forschern und ihren Themen („Public Engagement in Science“), die auf Inhalten von Zeitungsvolltexten beruhen, konnte nicht erfolgen, da das entsprechende Repositorium (ZEFYS) in der Projektlaufzeit nicht zur Verfügung stand.

3 Anwendungsbezogene Analyse der verwendeten Datenrepositorien

Dem Forschungsprojekt SoNAR lag die durch eine umfassende Potentialanalyse (Vgl. Antrag, Anhang 2) gestützte Annahme zu Grunde, dass sich die verwendeten Datenrepositorien in Qualität, Umfang, Aufbereitung der Daten für die Bedürfnisse historische Netzwerkanalysen im Kontext der Wissenschaftsforschung im 19. und 20. Jahrhundert eignen. Überprüft werden sollten mit Hilfe von Beispielen, inwieweit die Varianz historischer Fragestellungen abgebildet werden können und fehlende oder fehlerhafte Daten trotz großer Datenmengen Ergebnisse substantiell verfälschen.

3.1 Nutzung von Normdaten in der HNA

Die Historische Netzwerkanalyse ist auf ein einheitliches Datenset angewiesen, um Strukturen rekonstruieren zu können. Dem stehen zumeist heterogene historische Quellenbestände gegenüber, die formal uneinheitliche und nicht eindeutige Beziehungsinformationen enthalten. Zentrale Probleme der HNA sind daher eine zeitintensive Datenauswahl bzw. Datenerhebung sowie die Datenaufbereitung. Die in der Fallstudie zum Physiologen-Netzwerks im 19. und frühen 20. Jahrhundert vorgenommene manuelle Erhebung mit über 200 Knoten (Physiologen) und mehreren hundert Kanten (berufliche Beziehungen) verdeutlicht die enormen personellen Ressourcen, die allein für die Vorbereitung weitergehender quantifizierender und qualitativen Analysen notwendig sind. Auch in der Nutzerstudie wurde deutlich, dass der Grad der Digitalisierung nicht nur die Datenerhebung erheblich beeinflusst, sondern generellen Einfluss auf die Entscheidung für eine HNA hat (Vgl. HU Nutzerstudie,) und die Disambiguierung von Personennamen als zentrale Herausforderung in der Datenaufbereitung gesehen wird.

SoNAR bietet eine automatisierte Disambiguierung durch die Nutzung von Normdaten. Die Herkunft der Daten ist nachvollziehbar und nachprüfbar, die Existenz bestimmter Daten wird

durch die Pflichtfelder gewährleistet (die Felder sind „immer“ ausgefüllt), der Zugang zu Datenbeständen ist unbeschränkt, die Datenqualität ist gesichert, Zuweisungen sind eindeutig. Die Nutzung von Normdaten ist somit gut geeignet für historische Forschung und die damit verbunden Ansprüche an Daten/Quellen (→ Kriterien von Wissenschaftlichkeit) erfüllt.

Die Forschungstests haben diese Annahme für die, dem Projekt zugrunde gelegte GND („Gemeinsame -**Norm**-Datei“) grundsätzlich bestätigt, so lassen sich Personennetzwerke nachprüfbar (re-)konstruieren, Personen müssen auch bei Namensgleichheiten oder Ähnlichkeiten nicht mehr verifiziert werden, vor allem die Berufszuordnung ermöglicht eine gezielte Eingrenzung relevanter Personen (Beispiele: Netzwerke zu Max Weber, Physiologen, Nobelpreisträger, Urologen, Bakteriologen, etc.). Auch Körperschaften und Orte können eindeutig zugeordnet werden, was insbesondere zuverlässige geographische Verortungen ermöglicht (Beispiel: Die Netzwerke von Max Weber konnten auch nach Wirkungsorten differenziert werden. Zugleich konnten indirekte Verbindungen über die Indikatoren Ort/Zeit (re-) konstruiert werden)

Die Forschungstests haben zugleich einige Aspekte mit einschränkendem Einfluss auf die Nutzarmachung der Normdaten für die HNA ergeben. Sie erklären sich vor allem, aber nicht ausschließlich aus dem primären Zweck der GND (eindeutige Identifizierung von Autoren): u.a. die Diskrepanz zwischen Datenfeldern mit Beziehungsinformationen und den tatsächlichen Daten (Beispiel: die Befüllung „Berufliche und familiäre Beziehungen“), die Vorgaben für die Datenzuordnung zu bestimmten Datenfeldern (Beispiel: Berufskategorien („Primärberuf“), d.h. Personen aus dem in den Forschungstest zu erwartenden Personenkreis, können nur als Arzt, nicht aber als Neurologe identifiziert werden. Da „Physiologe“ als Primär gelistet ist, trat dieses Problem in Forschungstest 1 nicht auf), die Praxis der Zuordnung von Informationen zu bestimmten Datenfeldern (Beispiel: Beruf „Nobelpreisträger“, obwohl es ein Feld „Ehrungen“ gibt)

Die Nutzung von GND Normdaten in SoNAR ist ein Alleinstellungsmerkmal der Forschungstechnologie und hat sich in den Forschungstests grundsätzlich bewährt, die GND ist ein normierter, aber kein einheitlicher Datensatz, hier ist das Problem „fehlender Daten“ immanent. Es ist anzunehmen, dass sich für andere Fragestellungen weitere logische Inkonsistenzen ergeben. Das Problem der „fehlenden“ Daten in der GND konnte bisher nicht systematisiert gelöst werden. Eine angedachte statistische Auswertung von tatsächlich „befüllten“ Datenfeldern für die jeweilige Suchabfrage konnte aus Gründen der „Performance“ nicht realisiert werden.

3.2 Anreicherung der Normdaten und Auswahl weiterer Datenrepositorien

Wie bereits im Antrag angenommen, grenzt der Normdatensatz GND, die HNA – Fragestellungen allerdings auch ein. Diese Einschränkungen ergeben sich durch den Fokus auf nationale Bibliotheksbestände (Deutschland) und damit einen überproportionalen Anteil deutschsprachiger Autoren. Gleichzeitig werden damit ausschließlich Menschen die auch publiziert haben. Die Forschungstests ergaben, dass die Einschränkungen erheblichen Einfluss

auf historische Fragestellungen nehmen. Dazu gehören ein strukturell bedingter Fokus auf Elitenforschung sowie ein „Gender-Bias“ (geringere Anzahl von Autorinnen insbes. vor zweiter Hälfte 20. Jahrhunderts). Transnationale Beziehungen können nur selektiv rekonstruiert werden und ökonomische Fragestellungen sind kaum möglich, da weder Personen noch Körperschaften aus diesem Feld substantiell in der GND abgebildet werden. Die GND ist strukturell daher nur für bestimmte Fragestellungen nutzbar. Die Anreicherung der GND-Daten durch andere Normdaten (Wikidata), wie im Antrag vorgesehen, ist noch nicht abgeschlossen und konnte daher in den Forschungstests nicht überprüft werden.

Die in SoNAR hinzugezogenen Metadaten (ZDB, DNB KPE) erwiesen sich als grundsätzlich geeignet für die konzipierten Forschungstests. Digitalisierte Zeitschriften aus dem ZEFYS (Volltexte) waren innerhalb der Projektlaufzeit noch nicht verfügbar. Die Forschungstests ergaben, dass die in SoNAR hinzugezogenen Metadaten zu substantiellen Ergebnissen für Fragestellungen im Kontext der Medizin-, Wissenschafts- und Wirtschaftsgeschichte führen. Besonders breit gefasste Fragestellungen ergaben durch Schlagwortsuchen (u.a. mit Trunkierung) gute Ergebnisse (Beispiel: *physiolog*). Personensuchen, besonders auch in Kombinationen (mehrere Personen gleichzeitig) erbrachten auch bei spezifischen Fragestellungen zufriedenstellende Ergebnisse (Beispiel: „Streit um die Großhirnrinde“). Aus der Forschung bekannte zentrale Akteure waren in den jeweiligen Netzwerken wiederzufinden (Beispiel: Die in SoNAR erstellten Physiologen-Netzwerke enthalten nahezu alle wichtigen Fachvertreter bis in die Mitte des 20. Jahrhunderts).

Egozentrierte Netzwerke führten in der zweiten Fallstudie zu unübersichtlichen Ergebnissen, wobei Fehlerquellen in der regelbasierten Ableitung von sozialen Beziehungen inzwischen abgestellt werden konnten. (Beispiel: Das egozentrierte Netzwerk zu Max Weber).

Auch Zentralitätswerte, die auf Grundlage der bisherigen Forsch zu erwarten waren, konnten in SoNAR abgebildet werden, zugleich konnten auch im positiven Sinne überraschende Verknüpfungen „entdeckt“ werden. (Beispiel: Der Pretest zu Bakteriologie und Experimenteller Hygiene (Pettenkofer-Koch) ergab einerseits eine klare (zu erwartende) Abgrenzung von Unterstützernetzwerken, aber auch (überraschende) Verbindungslinien.

Herauszustellen ist die zentrale Bedeutung von Kalliope, sowohl qualitativ als auch quantitativ innerhalb des SoNAR Datensets. Erhebliche Probleme ergaben sich in den Forschungstests aus der heterogenen Erschließung in Kalliope. Dazu gehört u.a., dass die Daten zu Nachlässen unterschiedlich (Konvolute vs. Einzelblatt) eingepflegt sind, dass große Nachlässe wie die Max-Weber Gesamtausgabe (insbesondere die Briefbände) nicht in Kalliope erschlossen sind, dass das Datenfeld Schlagwort aktuell kaum genutzt wird und die Bestände daher im Gegensatz zu den Bibliotheksbeständen wesentlich schlechter bei Themensuchen genutzt werden können, dass nicht-normierte Datenfelder z.T. zentrale Informationen enthalten, die von SoNAR aktuell nicht angesprochen werden können (Beispiel Schlagwort „Nobelpreis“ häufig nur im Textfeld als Zusatzinformation erwähnt, daher ist ein Teil der Korrespondenz mit Nobelpreisträgern mit einer Schlagwortsuche nicht abbildbar).

Datenumfang und Datenqualität der DNB erfüllten nicht die Anforderungen (Beispiel: Die DNB enthält primär Literatur nach 1912, somit ungeeignet für die HNA zur Physiologie im 19. Jahrhundert.) Bei mehreren Autoren, Herausgeber sind nicht alle GND-hinterlegt, somit für

SoNAR nicht abrufbar. Die Integration eines weiteren Bibliothekskatalogs (SBB) konnte diese Probleme weitgehend lösen

Die ZDB-Daten enthalten weniger Beziehungsinformationen als erwartet. Erwartete Informationen zum Wandel von personellen Strukturen in Herausgebergremien sind nur selten enthalten.

Auch vor dem Hintergrund aktueller Forschungsschwerpunkte zu Korrespondenznetzwerken (vgl. u.a. Programm HNR+ResHist 2021) ist Kalliope als Kernbestand (Metadaten) von SoNAR anzusehen, mit einem erheblichen Entwicklungspotential durch eine wachsende Anzahl von beteiligten Archiven). Ähnlich der klassischen Quellenerhebung sind Repositorien wie die KPE durch eine große Heterogenität der Datenerhebung geprägt. Diese verringert die Qualität der SoNAR-Netzwerkergebnisse teilweise massiv. Dieser Qualitätsverlust ist derzeit durch andere Datenquellen nicht zu kompensieren.

Die Integration der unterschiedlichen Datenbestände in ein einheitliches Datenmodell hat sich als komplex erwiesen, u.a. ist die Ableitung sozialer Beziehungen aus Ressourcen (vgl. Regelset) fehleranfällig (Beispiel KPE: uneinheitliche Syntax in den Ausgangsdaten wie Unterschiede in Groß- und Kleinschreibung oder synonyme Verwendung von „behandelt/erwähnt“ führte zu gravierenden Datenlücken). Besondere Probleme liegen in der Komplexität der Datenstruktur („jeder Datensatz ist anders“). **Diese wird auch bei der Integration weiterer Datenbestände ist die Qualität der Ausgangsdaten ein bestimmender Faktor sein**, denn die Struktur der Ausgangsdaten muss zunächst erschlossen werden, um eine sinnvolle Integration zu gewährleisten. Eine technische, aber auch eine fachwissenschaftliche Prüfung für jedes neue Datenrepository ist unabdingbar. Trotz der genannten Herausforderungen ist die Ausweitung auf z.B. internationale (Norm-)datenbanken (u.a. VIAF) notwendig, um auch für die internationale HNA Forschung interessant zu werden.

4 Erstellen, Prüfen und Anpassen des Datenmodells

Die unter Punkt 2 skizzierten Fallstudien dienen zu allererst der iterativ gestalteten Entwicklung, Prüfung und Anpassung des SoNAR Datenmodells (Vgl. auch AP 1). Ihrerseits liegt Ihnen ein jeweils eigenes Datenmodell zugrunde, dass sowohl auf allgemeingültigen quantitativen und qualitative Beziehungsmerkmalen und Quellengattungen (Vgl. hierzu auch das Modellhafte Forschungsdesign), als auf spezifischen Indikatoren, die sich in den in SoNAR verwendeten Datenrepositorien enthalten sind, beruht.

4.1. Erstellen des Datenmodells

Die Fragestellung einer historischen Netzwerkanalyse und damit auch das jeweilige Datenmodell von den überlieferten Quellen beeinflusst, nicht selten präformiert. In SoNAR waren somit die Potenziale der Verbundrepositorien, Daten (Merkmalsausprägungen) zu Entitäten (Merkmalsträger) in Beschreibungselementen (Merkmale) für quantitative Analysen und Visualisierungen zu erfassen, zu berücksichtigen. In der Gegenüberstellung Tabelle 3 wird

deutlich, dass die SoNAR Repositorien ein breites Spektrum an quantifizierenden Merkmalsträgern und Merkmalen umfassen, auf qualifizierende Merkmalsausprägungen hingegen weitgehend verzichtet werden muss.

Tab. 3 Allgemeine Quantitative und Qualitative Beziehungsmerkmale

Merkmalsträger	Merkmal	Quellen (allgemein)	Quellen (SoNAR)
Personen	Selbstzuschreibungen („Ich bin ein Schüler/Freund/Kollege/Weggefährte von“)	Autobiographien/ Selbstzeugnisse, Interviews	-
Personen	Fremdzuschreibungen („Er war ein Schüler/Freund/Kollege/Weggefährte von“)	Nachrufe, Biographische Schriften, Berichte, Presse	-
Personen	Verwandtschaft	Biographische Nachschlagewerke, Selbstzeugnisse	GND
Personen	Korrespondenz	Nachlässe	Kalliope
Personen	Affiliation (dynamisch): Labore/Institute/Fakultäten /Universitäten	Biographische Nachschlagewerke/ Selbstzeugnisse, GND	GND
Personen	Kongressteilnahme (Sektion/Panel)	Tagungsberichte	GND
Personen; Werke	Öffentlicher Austausch (Rezensionen; „Bemerkungen zu“)	Fachzeitschriften	ZEFYS
Personen	Promotionsverhältnis (Betreuer, Co-Betreuer, Prüfer)	Dissertationen	GND
Personen; Werke	Zitationen	Publikationen	-
Personen; Körperschaften	Mitgliedschaften (Fachgesellschaften, Vereine, etc.)	Mitgliederverzeichnisse, Selbstzeugnisse	-
Personen; Körperschaften	Herausgeberschaften (Fachzeitschriften, Handbüchern)	Bibliothekskataloge	ZDB; DNB; StabiKat *
Personen; Körperschaften; Orte	Co-Nennungen	Nachrufe, Zeitungsberichten, hist. Forschung	[ZEFYS]
Personen; Körperschaften	Ehrungen (Preise, Nominierungen, z.B. für den Nobelpreis)	Biographische Nachschlagewerke/ Selbstzeugnisse, Nobelpreisarchiv	GND

Thema	Sachschlagwort	Thesaurus; Bibliothekskataloge; Indizes	GND; DNB; <i>Stabikat</i> *, ZDB
Orte	Ereignisse; Tagungen		GND
Tagungen	Themen; Personen, Orte	Kongress- /Tagungsprogramme und berichte,	GND

*Im Projektverlauf ergänzter Bibliothekskatalog

4.2. Prüfen und Anpassen des Datenmodells

Für die Pretests wurden aus dem Merkmalset jeweils bestimmende Indikatoren identifiziert. Für die „Neurologen-Netzwerke im Nationalsozialismus“ standen die beruflichen Verbindungen über Affiliationen (GND), verwandtschaftliche Beziehungen (GND) gemeinsame Herausgeberschaften (ZDB, DNB) sowie die vorhandene Korrespondenz (Kalliope) in Fokus. Zu diesem Zeitpunkt war die erste Version des Datenmodells noch eine reine Verknüpfung der Einzelressourcen. Soziale Beziehungen im eigentlichen Sinne wurden nur durch GND-Daten hergestellt. Die Ableitungen sozialer Beziehungen aus den Ressourcen erfolgte nach einem regelbasierten Verfahren (Vgl. REGELSET sowie R003).

Für den Pretest „Disziplinenbildung in der Urologie“ wurden als wichtige Indikatoren die Facharztbezeichnung (GND-Berufssachschlagworte), Fachzeitschriften (ZDB), Lehrstühle (evtl. Kalliope) sowie Hand- und Lehrbücher (DNB) herangezogen. Diskrepanzen ergaben sich zwischen den relevanten Datensätzen und Informationen, die in den Ausgangsdaten enthalten sind und denen, die zu damaligen Zeitpunkt mit SoNAR anzusprechen und zu verarbeiten waren. Zentrale Urologen (u.a. Präsidenten der Deutschen Gesellschaft für Urologie) konnten in SoNAR zunächst nicht über das Schlagwort Urologe angesprochen werden, obwohl sie in der GND verzeichnet sind. Schlagworte der DNB fehlten in der ersten Version des SoNAR Datenmodells. Weitere Diskrepanzen, die sich aus der Qualität der Ausgangsdaten ergeben, blieben bestehen. Beispiel Berufsangaben: Für einen GND Eintrag ist nur das Feld 550 „berc“: Charakteristische Berufe verpflichtend. Hierunter sind vorwiegend ausgeübte Berufe zu verstehen. Bei Akademischen Berufen wird vor allem nach Hauptrichtungen (Ausdruck?) unterschieden: Arzt, Physiker, Jurist, Theologe, etc. Urologe wäre in diesem Falle im Zusatzfeld: 550 „beru“: weitere Berufe einzutragen. Dieses Feld ist keine GND Pflichtfeld und ist daher der Akribie des jeweiligen Datenproduzenten überlassen. Daraus ergab sich eine Anpassung des Datenmodells sowie Anforderungen an die Datengrundlage für eine zeitliche Dimension (R048) und an die transparente Darstellung für den Nutzer, welche Datenfelder (z.B. in der GND) verpflichtend sind und welche Datenfelder tendenziell zufällig tatsächlich Daten enthalten.

Ein prägnantes Beispiel wie sich fehlende Daten, insbesondere in Hinblick auf die Nachlassdatenbank Kalliope massiv auf die Ergebnisse auswirkt ergab der Pretest zur „Cholera Bekämpfung: Experimentelle Hygiene vs. Bakteriologie“. Als Indikatoren dienten hier

universitäre Affiliationen (GND), Verbindungen in Politik und Wirtschaft (GND), Fachzeitschriften (ZDB) und Korrespondenz (Kalliope). Die umfangreichen Korrespondenzbestände in der Bayrischen Staatsbibliothek zu Max von Pettenkofer suggerierten eine viel größere Anzahl an Unterstützern für Pettenkofer als für Koch. Die weitgehende fehlende Koch-Korrespondenz vermittelt hier ein Bild vom Grad der jeweiligen Vernetzung in der community, das einer vertieften Analyse bedarf. Eine Anforderung daraus war die Transparenz der Daten für jede einzelne Abfrage, damit es aufgrund von fehlenden Daten nicht zu Fehlschlüssen kommt.

Ähnlich gelagerte Probleme ergaben sich in dem Pretest „Nobelpreis im wissenschaftlichen Diskurs“ mit den Indikatoren Sachschlagwort zu einer Person (GND, DNB, ZDB), Nominierungen (Kalliope) und Erwähnung in Korrespondenz, u.a. Gratulation, Kritik (Kalliope). In der *GND* finden sich lt. SoNAR Abfrage 457 Nobelpreisträger, ca. 50 % der Personen, die einen Nobelpreis erhalten haben. Mehrere Gründe sind für diese Diskrepanz denkbar: Nicht alle NP haben eine selbständige Publikation: Problem der selbständigen Publikation in den Naturwissenschaften? Nicht alle Kulturkreise sind gleich stark in der *GND* vertreten. Vor allem für die Bewertung des Nobelpreises interessant ist dessen normierte Zuordnung. Vorwiegend als Beruf, aber auch als Adelstitel oder akademischer Grad. Dies entspricht der wissenschaftshistorischen Einschätzung, dass die meisten Nobelpreisträger nicht mehr Physiker, Chemiker oder Mediziner seien, sondern von ihren ursprünglichen Professionen entkoppelte übergeordnete Instanzen wissenschaftlicher Expertise. Nicht normiert sind 95% der in *Kalliope* enthaltenen Informationen/Daten zum Nobelpreis. Sie befinden sich in Textfeldern und können von SoNAR daher nicht verarbeitet werden. Normierte Schlagworte werden hier nur selten verwendet. Auch hier betrafen die resultierenden Anpassungen und Anforderungen, die Frage der Transparenz, welche Datenfelder (z.B. in der *GND*) verpflichtend sind und welche Datenfelder tendenziell zufällig tatsächlich Daten enthalten sowie Angabe von Diskrepanzen zwischen Volltext-Abfragen in den Ausgangsdaten und den SoNAR-Abfragen, die ausschließlich normierte Datenfelder ansprechen.

Im Pretest „Wilhelm Roux – wechselnde Bündnisse in den Wissenschaften“ (Indikatoren: berufliche Verbindungen über Affiliationen (GND), Verwandtschaftliche Beziehungen (GND), gemeinsame Herausgeberschaften (ZDB, DNB), Korrespondenz (Kalliope)) zeigte sich, dass von den in der Autobiographie genannten Personen in der SoNAR Grafik nur wenige wiederfinden. Hieraus ergab sich die spannende Frage, wie er mit den Personen verbunden war, die offenbar nicht zu seinen Unterstützern zählte. Zu den Anpassungen und Anforderungen gehörten Überlegungen zur zeitlichen und räumlichen Dimensionierung von Netzwerken in SoNAR. Insbesondere Merkmale, die eine zeitliche Differenzierung von Netzwerken über längere Zeiträume hinweg ermöglichen fehlen weitgehend. Die zeitliche Einordnung erfolgt primär über die Lebensdaten der Akteure und umschreibt damit große Zeitspannen, die der Dynamik von wissenschaftlichen Beziehungen nicht gerecht werden und eher traditionellen Vorstellungen von dauerhaften Lehrer-Schüler-Verbindungen entsprechen.

5 Erheben, Aufbereiten und Selektieren von Daten

Die Datenerhebung, die Aufbereitung und die Selektierung von Daten erfolgte für die Forschungstests auf Grundlage der Forschungsliteratur. Die Datengrundlage für die Pretests waren jeweils einzelne Referenzstudien⁹ zumeist von Autoren aus dem Forschungsteam.

In der zweiten Hälfte des Projektes wurden das angepasste Datenmodell und vor allem die formulierten Anforderungen an die Visualisierungskonzepte durch zwei größere Fallstudien getestet.

5.1 "Das medizinisch-physiologische Forscherkollektiv des 19. und frühen 20. Jahrhunderts"

Diese Fallstudie ist ein Beispiel für eine themenzentrierte Suche sowie für die gleichzeitige Suche nach mehreren Personen. Auf der Grundlage bisheriger Forschung erfolgte die Bildung von Hypothesen zu den sozialen Bedingungen der Entstehung von wissenschaftlichen Disziplinen, die von allgemeinen zu speziellen Annahmen reichen. Basis sind Personen- und Sachindizes einschlägiger Forschungsarbeiten zum Untersuchungsgegenstand. Die so gewonnenen Personen- und Themencluster wurden mit netzwerkanalytischen Methoden visualisiert, beschrieben und mit Ergebnissen aus dem Datenbestand von SoNAR (IDH) verglichen.

Ein guten ersten Anhaltspunkt für die Quantität des medizinisch-physiologischen Forscherkollektivs im Untersuchungszeitraum bietet die Studie des Medizinhistorikers und Physiologen Karl Rothschuh. Sein Personenindex umfasst ca. 1200 Namen. Soziale und intellektuelle Beziehungen lassen sich aus den Texten und insbesondere auch aus den Stammtafeln ableiten. Aus diesen und anderen Quellen kennen wir wichtige Protagonisten, ihre Schüler, zentrale Orte und vor allem wissenschaftliche Schnittpunkte und Differenzen.

5.2.1 „Max Weber“

Den inhaltlichen Rahmen der zweiten Fallstudie bildeten Netzwerke deutscher und österreichischer Nationalökonomien des 19. und 20. Jahrhunderts. Inhaltlich lag das Erkenntnisinteresse primär auf egozentrierten Netzwerken, den ausgehend von hypothesengeleiteten Abfragen in SoNAR nachgegangen wurde. Um potentiell viele bzw. unterschiedliche Ergebnisse in SoNAR erzielen zu können, sollte die untersuchte Person im ersten Beispiel einen hohen Bekanntheitsgrad innerhalb der Nationalökonomie im entspr. Untersuchungszeitraum aufweisen. Als Beispiel wurde die Person „Max Weber“ ausgewählt

⁹ Karenberg A, Fangerau H, Martin M (2020): Neurologen und Neurowissenschaftler in der NS-Zeit: Versuch einer Bewertung. *Nervenarzt* 91: 128-45; Fangerau H, Imhof C (2015): "Medizinische Spezialisierung: Wege der Urologie in beiden deutschen Staaten und die Gründung der Deutschen Gesellschaft für Urologie der DDR". In Halling T, Moll F, Fangerau H (Hrsg.): "Urologie 1945-1990. Entwicklung und Vernetzung der Medizin in beiden deutschen Staaten", Springer, Heidelberg, S. 21-34; Hansson N, Halling T, Fangerau H (eds): *Attributing Excellence in Medicine: The History of the Nobel Prize (Clio Medica 98)*. Brill & Rodopi, Leiden 2019; Halling T, Hansson N, Fangerau H (2018): „Prisvärdig“ Forschung? Wilhelm Roux und sein Programm der Entwicklungsmechanik. *Berichte zur Wissenschaftsgeschichte* 41:73-97

und konkret nach Netzwerken und Akteuren gefragt, die im Zuge der ersten Berufung Webers an die Albert-Ludwigs-Universität Freiburg besondere Relevanz hatten. Dabei sollten berufliche und auch private Netzwerke genauer untersucht werden.

Insgesamt betrachtet ist die Person Max Weber (1864-1920) sehr gut erschlossen. Neben zahlreichen Publikationen liefert die *Max Weber Gesamtausgabe* eine hervorragende Quellengrundlage: mit insges. 47. editierten Bänden zu Schriften und Reden, Briefen sowie Vorlesungen und Vorlesungsnachschriften. Um die Qualität der Abfrageergebnisse in SoNAR zu bewerten, stellten insbesondere die Briefbände der MWG eine gute Referenz da, zumal hier Briefe zum Bewerbungsverfahren Webers enthalten sind, denen wertvolle Hinweise auf Personen, Akteure und Netzwerke entnommen werden können. Den Quellen und der Literatur konnte entnommen werden, dass im Zuge der Berufungen nach Freiburg nicht nur Webers fachliche Expertise, sondern auch unterschiedliche Netzwerke von Bedeutung waren, in denen einzelne Akteure den Berufungsprozess - mehr oder minder - stark beeinflussten.

5.2.2 Robert Liefmann

Im nächsten Beispiel der zweiten Fallstudie sollten die für SoNAR gewonnenen Ergebnisse und Erkenntnisse aus der „Max Weber“ Studie überprüft und weitere Funktionalitäten SoNARs untersucht werden.

Dafür wurde der Fokus nochmals bewusst auf ein egozentriertes Netzwerk eines deutschen Nationalökonomen gelegt. Die Person sollte einige Parallelen zu Weber aufweisen (z.B. gleiche Wirkungsorte oder Generationszugehörigkeit), jedoch einen deutlich geringen Bekanntheitsgrad besitzen. Unter Berücksichtigung dessen, fiel die Auswahl auf Robert Liefmann (1874 – 1941).

Dieser studierte Rechtswissenschaften und Nationalökonomie in München, Brüssel und Berlin, verbrachte aber den Großteil seiner Studienzeit an der Universität Freiburg. Hier wurde er bei Max Weber mit der Doktorarbeit „Die Unternehmerverbände (Konventionen und Kartelle), ihr Wesen und ihre Bedeutung“ promoviert und erlangte später eine außerplanmäßige Professur in Freiburg. Liefmann beschäftigte sich in seiner Forschung primär mit Monopolen und Kartellen. Innerhalb der Nationalökonomie wurden seine wissenschaftlichen Ausführungen vergleichsweise wenig Bedeutung zugemessen. Von einem Großteil der Community wurde er eher kritisch beäugt und galt als Sonderling, was vor allem seinem hohen Maß an Selbstüberzeugung und seinen polemischen Äußerungen und Reaktionen geschuldet war. Dieses Verhalten lässt sich auch in einer ganzen Reihe von Rezensionen und Gegenrezensionen erkennen.¹⁰

Um anhand des zweiten Beispiels die Funktionsweisen SoNARs und die Möglichkeiten für die HNA weiter zu erproben, wurde der Fokus auf die Untersuchung intellektueller Netzwerke

¹⁰ Roman Köster, Die Wissenschaft der Außenseiter. Die Krise der Nationalökonomie in der Weimarer Republik, Göttingen 2011, S. 112f.

gelegt. Dabei stand die Frage nach intellektuellen Beziehungen um die Person Robert Liefmann im Mittelpunkt. Als Quellenmaterial wurden die oben erwähnten Rezensionen, Gegenrezensionen und Abhandlungen herangezogen. Ausgehend von der zugrunde gelegten Fragestellung ergaben sich weitere Fragen, mit denen die hierarchischen Stellungen der Akteure (Personen und Institutionen) innerhalb und ggf. außerhalb der Community sowie ihr Einfluss im Diskurs identifiziert werden sollte.

6 Analysieren und Visualisieren von Daten

In SoNAR wurde grundsätzlich von einer explorativen als auch einer thesengeleiteten Aussagekraft von Netzwerkvisualisierungen ausgegangen. Die in SoNAR gefundenen Lösungsansätze (vgl. Ergebnisse AP 3) zur Datentransparenz ermöglichen eine vollständige Kontextualisierung aller Knoten und Kanten (z.B. Verlinkungen zu den Ressourcen, Auffächern von Kanten), die zur zeitlichen und räumlichen Dimensionierung ermöglichen wiederum dynamische Netzwerkdarstellungen (u.a. Entwicklung von Netzwerken über die Zeit, Aufscheinen und Verschwinden von Akteuren und Beziehungen (Beziehungstypen)).

6.1 „Das medizinisch-physiologische Forscherkollektiv des 19. und frühen 20. Jahrhunderts“

Anhand der ersten Fallstudie konnte demonstriert werden, wie wir uns mit Hilfe der SoNAR Forschungstechnologie sozialen und intellektuellen Netzwerken genähert haben, beginnend mit einer trunkierten Sachschlagwortsuche nach „physiolog“. Verschiedene Anwendungsszenarien der HNA wie Erkundung, Gewichtung und zeitliche oder räumliche Dimensionierung von sozialen Beziehungen sowie die Erkennung von Mustern, gingen dabei in der Forschungspraxis ineinander übergehen oder wechselten einander ab.

Bei der Exploration der Datenbestände zum Thema Physiologie dienen im SoNAR Prototyp eine Zeitleiste, eine Personenliste sowie eine Statistik zu Knoten und Kanten der Darstellung der Orientierung des Nutzers. Für die genannte Schlagwortsuche umfasst der Zeitraum fast 2000 Jahre. Die Mengenverteilung auf der Skala verdeutlicht aber, dass sich der Untersuchungszeitraum, 19. und frühes 20. Jahrhundert, als Blütezeit der Physiologie als Forschungsgebiet in SoNAR widerspiegelt.

Die absolute Personenanzahl ist mit den Angaben in der Forschungsliteratur vergleichbar. Die Personenliste verdeutlicht welche Akteure der Nutzer im Netzwerk wiederfinden kann und wie viele Personen mit ihnen verknüpft sind. Wir finden hier viele der Namen, die wir erwartet haben, aber auch einige, die auf den ersten Blick verwundern und erklärungsbedürftig sind. Durch die direkte Verknüpfung mit den jeweiligen Ressourcen können die Daten zielgenau erkundet werden.

Die Darstellungen in SoNAR sind zugleich gewichtend, wobei der zu Grunde liegende Algorithmus immer kenntlich gemacht wird. In dieser birds-eye Perspektive erkennen wir allerdings noch keine Details, aber schon mindestens drei Personencluster. Sie bilden sich um Nathan Zuntz, um Wilhelm Wundt und etwas abseits um den erwähnten John Eccles. Alle drei

Akteure sind tatsächlich zentral für verschiedene Aspekte der Physiologie, Wundt verbindet die Physiologie zudem mit dem Forscherkollektiv der Psychologen. Man kann ihn somit also als klassischen Hub zwischen zwei Denkkollektiven bezeichnen. Die exponierte Stellung der drei in unserem Netzwerk ist aber auch auf die bereits in den Pretests festgestellte desperate Quellenlage in SoNAR. Der ab den 1930er Jahren in Großbritannien, Australien und den USA tätige Physiologe, Neurowissenschaftler und Nobelpreisträger John C. Eccles (1903-1997) verfügt in den Visualisierungen auch deshalb ein so großes und im Vergleich zu anderen in ihrer wissenschaftlichen Bedeutung durchaus gleichwertigen Forschern seiner Generationen durchaus dominantes Netzwerk, da sein Korrespondenznachlass nahezu vollständig in Kalliope erfasst ist. Für Wundt und Zuntz gibt es ebenfalls Briefnachlässe, die in Kalliope erfasst sind. Die durchgängige Datentransparenz erschließen sich dem Nutzer solche Auffälligkeiten aber umgehend.

Ein Beispiel für das weniger explorative, sondern mehr hypothesentestende Anwendungsszenario ist dieses Netzwerk zu Akteuren, die wir aus unserer eigenen Forschung als Protagonisten eines wissenschaftlichen Streits um die sogenannte Großhirnrinde identifiziert haben. Hier bietet SoNAR die Möglichkeit ein bereits definiertes Netzwerk in einen größeren Zusammenhang einzubetten. In einer ganz konkreten Gegenüberstellung von klassischer Analyse und SoNAR-Befund zur Person Friedrich Goltz zeigte sich wiederum, dass besonders bei sehr engen Fragestellungen fehlende Daten zu extremen Verzerrungen der Ergebnisse führen können: Im genannten Beispiel fanden sich keine Überschneidungen, das bedeutet, nachweislich wichtige Kontaktpersonen von Goltz sind in den SoNAR Daten nicht enthalten bzw. nicht mit Goltz verknüpft, dafür zusätzliche andere Personen und damit neue Ansätze, denen nachgegangen werden kann.

6.2.1 „Max Weber“

In der zweiten Case Study stellten sehr große Datenmengen eine besondere Herausforderung dar. Die insgesamt fast 2000 Knoten und über 10.000 Kanten bzw. 110.000 Kanten (mit Querbeziehungen¹¹ erforderten für die weitere Analyse modifizierte Visualisierungskonzepte.

Völlig anders gelagerte Fragen wurden z.B. durch Verbindungen zwischen Max Weber und Plato aufgeworfen, zumal eine soziale Beziehung zwischen diesen auszuschließen ist. Es stellte sich heraus, dass beide Namen Bestandteil eines Karteikastens sind, der in Kalliope verzeichnet ist. In diesem Karteikasten befinden sich unterschiedliche Namen, die vom Bestandsbildner - aus nicht ersichtlichem Grund - in Verbindung gebracht wurden. Neben sozial formierten Netzwerken können in SoNAR also auch (in diesem Fall zunächst unbeabsichtigt) „intellektuelle Netzwerke“ identifiziert werden (Vgl. hierzu im modellhaften Forschungsdesign 5.)

Aufgrund der Komplexität des Weber-Netzwerkes und insbesondere mit Blick auf die dargelegte Forschungsfrage, mussten Möglichkeiten für die Selektion und Reduktion der Daten gefunden werden. Dafür wurden Filter entwickelt, mittels derer gezielt nach Personen und Beziehungen in ausgewählten Zeiträumen gesucht werden konnte. Die Filter

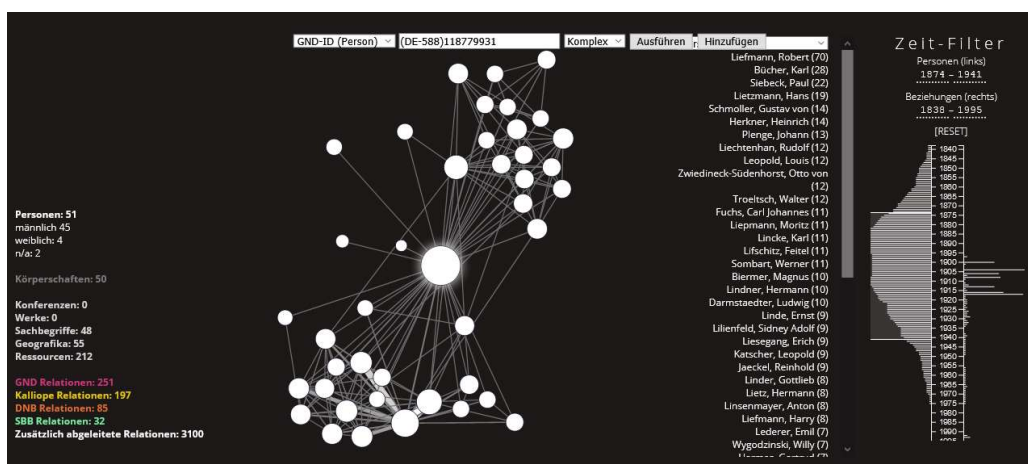
¹¹ d.h. zusätzlich zur den Verbindungen zu Weber werden auch die zu anderen Knoten angezeigt

ermöglichten somit die Daten im Weber-Netzwerk für die angestrebte Analyse handhabbar zu machen und den Fokus auf jene Netzwerke zu legen, die im Zuge der Berufung Webers 1894 besonders relevant waren. Am Beispiel von Friedrich Althoff, eines zentralen Protagonisten im Berufungsverfahren, lässt sich ein weiterer Vorteil von SoNAR herausstellen. Mit dem Klick auf die Kante werden den Forschenden jene Ressourcen angezeigt, die Althoff und Weber miteinander verbinden. Durch die Verlinkung zur Originalquelle im Ausgangsrepositorium konnten diese verifiziert und weitere Informationen eingesehen werden. Für die gesamte Fallstudie enthielten die KPE-Ressourcen allerdings kaum relevante Daten bereit, die zur Beantwortung der Fragestellung beigetragen hätten. Ein grundsätzliche Schwierigkeit bestand darin, dass die vielen Briefe, die im Kontext von Webers Berufung verfasst wurden, zwar in der MWG allerdings nicht in KPE zu finden sind. Hier bestätigten sich die Befunde aus den Pretests und der ersten Fallstudie: Die Ergebnisse in SoNAR sind stark vom jeweiligen Forschungsinteresse und den erfassten Datenbeständen abhängig.

6.2.2. „Robert Liefmann“

Vor den Abfragen in SoNAR wurden für das zweite Beispiel zunächst Rezensionen und Gegenrezensionen recherchiert und aus diesen die für die Fragestellung relevanten Namen händisch extrahiert. Dabei wurden primär der/die Rezensent*in und der/die Autor*in des rezensierten Werkes festgehalten ebenso wie Namen, die mit Bezug auf die Rezension bzw. Gegenrezension prominent im jeweiligen Text erwähnt wurden. Die Namen sollten für die Abfragen in SoNAR als Referenz herangezogen werden.

Die ersten Ergebnisse zur Abfrage „Robert Liefmann“ zeigten ein Netzwerk mit Relationen (1841-1975), die über den Lebenszeitraum Liefmanns deutlich hinausreichten. Da das Forschungsinteresse primär auf die intellektuellen Beziehungen zu Liefmanns Lebzeiten fokussierte, wurde der entwickelte Zeitfilter auf 1874-1941 gesetzt (vgl. Bild 1). Unabhängig von einer zu großen Datenmenge, die bei Max Weber den Ausschlag für die Entwicklung des Filters gegeben hatte, zeigte sich auch bei den deutlich geringeren Datenmengen im Liefmann-Beispiel der Vorteil von Zeitfilterfunktionen, die den Forschenden eine gezielte Analyse ermöglichen.



Die angezeigten Knoten wurden einzeln überprüft, zumal die Darstellung keine Differenzierung von sozialen und intellektuellen Beziehungen zulässt. Die Namen im Netzwerk wurden mit denen aus der händisch erstellten Liste abgeglichen. Anschließend wurde anhand der Kanten überprüft, ob durch die Quellenverweise zur entsprechenden Relation weitere verwertbare Informationen angezeigt werden. Es stellte sich heraus, dass *Zwiedineck*¹² die einzige (lediglich) namentliche Übereinstimmung der SoNAR-Ergebnisse mit der händisch erstellten Liste darstellt. Die angezeigte Verbindung zu Liefmann beruhte auf einer Korrespondenz zwischen Zwiedineck und Karl Bücher, in der Liefmann namentlich erwähnt worden war. Unabhängig davon, dass das gewünschte Ergebnis mit Blick auf die inhaltliche Fragestellung eher ernüchternd ausfiel, zeigte sich an dieser Stelle nochmals der direkte Vorteil, der aus einer direkten Verknüpfung der Kante mit der Ressource (hier KPE) hervorgeht.

Mit Blick auf die Fragestellung bleibt festzuhalten: Es konnten keine Relationen identifiziert werden, die über Rezensionen, Gegenrezensionen oder Abhandlungen zustande gekommen sind. Bei der Prüfung aller Kanten im Netzwerk konnten ebenfalls keine weiteren Verweise auf Rezensionen o.Ä. gefunden werden. SoNAR kann also die erwartbaren intellektuellen Beziehungen derzeit nicht abbilden. Demnach ist es lohnenswert eine Integration weiterer Datenbanken insbesondere von Rezensionsdatenbanken für zukünftige Untersuchungen intellektueller Beziehungen anzustreben.

Anforderung:

- Können die am rechten Rand angezeigten Personen nur dem gefilterten Zeitraum zugeordnet werden? Wie können die relevanten Konten mit Hilfe der legende am rechten Bildrand schnell identifiziert werden? Können die Personennamen alphabetisch angeordnet werden und mit einem Klick auf den Namen, farblich in Netzwerk gekennzeichnet werden?
- Wie können die Forschenden soziale und intellektuelle Beziehungen unmittelbar im Netzwerk unterscheiden?

¹² Vermutlich handelt es sich um *Otto von Südenhorst -Zwiedineck*

7 Fazit

Für das Vorhaben, eine Forschungstechnologie für die Historische Netzwerkanalyse aufzubauen resultiert daraus: Die Nutzung von GND Normdaten in SoNAR ist ein Alleinstellungsmerkmal der Forschungstechnologie und hat sich in den Forschungstests grundsätzlich bewährt. Die GND ist strukturell aber nur für bestimmte Fragestellungen nutzbar.

Die Anreicherung der GND-Daten durch andere Normdaten (Wikidata), wie im Antrag vorgesehen, ist noch nicht abgeschlossen und konnte daher noch nicht in den Forschungstests überprüft werden.

Die in SoNAR hinzugezogenen Metadaten (ZDB, KPE) erwiesen sich als grundsätzlich geeignet für die konzipierten Forschungstests. Auch vor dem Hintergrund aktueller Forschungsschwerpunkte zu Korrespondenznetzwerken (vgl. u.a. Programm HNR+ResHist 2021) ist Kalliope als Kernbestand (Metadaten) von SoNAR anzusehen, mit einem erheblichen Entwicklungspotential durch eine wachsende Anzahl von beteiligten Archiven). Ähnlich der klassischen Quellenerhebung sind Repositorien wie die KPE durch eine große Heterogenität der Datenerhebung geprägt. Diese verringert die Qualität der SoNAR-Netzwerkergebnisse teilweise massiv. Die Integration der unterschiedlichen Datenbestände in ein einheitliches Datenmodell hat sich als komplex erwiesen, u.a. ist die Ableitung sozialer Beziehungen aus Ressourcen (Regelset) fehleranfällig.

Besondere Probleme liegen in der Komplexität der Datenstruktur („jeder Datensatz ist anders“). Diese wird auch bei der Integration weiterer Datenbestände ist die Qualität der Ausgangsdaten ein bestimmender Faktor sein, denn die Struktur der Ausgangsdaten muss zunächst erschlossen werden, um eine sinnvolle Integration zu gewährleisten. Eine technische, aber auch eine fachwissenschaftliche Prüfung für jedes neue Datenrepository ist unabdingbar. Trotz der genannten Herausforderungen ist die Ausweitung auf z.B. internationale (Norm-)datenbanken (u.a. Vial) notwendig, um auch für die internationale HNA Forschung interessant zu werden.

Die gefundenen Lösungen in der automatisierten Datenauswertung und der transparenten Visualisierung, erfüllen bisher in der HNA in Ansätzen realisierte Anforderungen an dynamische, zeitlich und räumlich dimensionierte Darstellungsformen. Die Anforderungen an ein Datenmodell, dass die Integration weiterer Datenbestände ermöglicht, stellen sich aus fachwissenschaftlicher Sicht als ressourcenintensives, aber letztendlich lohnenswertes Ziel für die zweite Projektphase von SoNAR dar