



Auswertungsbericht

Visualisierungsworkshop SoNAR (IDH)

Explorative Perspektiven auf soziosemantische Netzwerke

Sina Menzel & Vivien Petras

Humboldt-Universität zu Berlin

Oktober 2019

Abstract

Dieses Dokument fasst die Auswertung des Workshops zur Visualisierungsherangehensweise vom September 2019 im Projekt SoNAR (IDH) nach Zielsetzung des Arbeitspakets 4 zusammen.

Inhalt

1.	Setting.....	3
2.	Ziel der Auswertung.....	3
3.	Datengrundlage	4
4.	Methodik	4
5.	Ergebnisse.....	5
	5.1 Aussagen der HNA/SNA-ExpertInnen.....	5
	5.2 Aspekte für die Leitfäden.....	7
	5.3 Diskussionspunkte	10
6.	Limitierungen.....	10
7.	Zusammenfassung	10
8.	Referenzen.....	11
	Anhang A: Codebuch	12
	Anhang B: Datenmaterial.....	15

1. Setting

Organisation durch AP3: Mark-Jan Bludau und Marian Dörk

Datum: 16.09.19

Zeit: 14:00-18:10 Uhr

Ort: Raum-D223, Designhaus, FH Potsdam, Kiepenheuerallee 5, 14469 Potsdam.

10 Teilnehmende:

Name	Institution	Fachlicher Hintergrund	Projektbezug
Bludau, Mark-Jan	FHP, UCLAB	Informationsvisualisierung	Intern
Dörk, Marian	FHP	Informationsvisualisierung	Intern
Halling, Thorsten	HHU	HNA, Medizingeschichte	Intern
Helm, Susanne	FU, FHP	SNA, Soziologie	Extern
Leitner, Elena	DFKI	Datenaufbereitung	Intern
Menzel, Sina	HU	Qualitätssicherung	Intern
Petras, Vivien	HU	Qualitätssicherung	Intern
Schuster, Paula	FHP	Informationsvisualisierung	Extern
Topp, Sascha	MPIWG	HNA, Wissenschaftsgeschichte	Extern
Wintergrün, Dirk	MPIWG	HNA, Wissenschaftsgeschichte	Extern

2. Ziel der Auswertung

Die vorliegende Auswertung basiert auf dem durch AP3 konzipierten und ausgerichteten Workshop zum Thema „Explorative Perspektiven auf soziosemantische Netzwerke“ vom 16.09.2019. Die systematische Auswertung der Workshop-Materialien durch AP4 erfolgt als Vorstudie zu den Arbeitsschritten AP4-4 und AP4-5 (vgl. Menzel/Petras 2019, S. 7-8) und ist damit Teilschritt der Qualitätssicherung im Projekt SoNAR (IDH).

Ziel ist es, durch einheitliche Codierung von Beobachtungsnotizen und Audiomitschnitten fundierte Bedürfnisäußerungen aus dem Workshop sowie Vorstellungen zu Funktionen, die als hilfreich eingestuft werden, zu erfassen. Diese bilden gemeinsam mit der Auswertung des HNR-Workshops (Historical Network Research) aus Arbeitspaket AP2-2 die Grundlage für die Leitfadenerstellung zu den späteren ExpertInnen-Interviews, Case Studies und Cognitive Walkthroughs (vgl.ebd.). Auf diese Weise wird sichergestellt, dass zentrale Aspekte aus den beiden Workshops systematisch einbezogen und überprüft werden, indem die ProbandInnen der Evaluierungen III und IV sie gezielt adressieren.

Die vorliegende Zielsetzung ist losgelöst von der Zielsetzung von AP3 zu betrachten, auch wenn die Ergebnisse anschließend gemeinsam verglichen und diskutiert werden. Als Zielsetzung wurde von Seiten AP3s festgehalten:

1. Identifizierung von Ansätzen zur Visualisierung von Netzwerken
(Mail_Auswertungsansätze_AP3_Doerk, A. 40).
2. Herstellung von Beziehungen zur Forschungsweise der FachwissenschaftlerInnen und Identifizierung von Zielkonflikten zwischen Designern und FachwissenschaftlerInnen
(Visualisierungsworkshop_Notizen_Petras, A. 14).
3. Definition und Prioritisierung von (semantischen) Einstiegspunkten für die Visualisierung
(Visualisierungsworkshop_Notizen_Petras, A. 42).

3. Datengrundlage

Die vorliegende Auswertung basiert auf den Audioaufnahmen, die mit Einverständnis aller Teilnehmenden durch AP3 von Teilen des Workshopverlaufs angefertigt wurden. Das Sampling wurde ebenfalls durch AP3 durchgeführt, dazu wurde auf eigene Kontakte der FHP, sowie Kontakte von AP2 zurückgegriffen und besonderer Wert auf diverse fachliche Hintergründe (Geschichts- und Sozialwissenschaft, Informationsvisualisierung, andere Fachbereiche) der Teilnehmenden gelegt. Zusätzlich bezieht sich die Auswertung auf die durch AP4 (Menzel/ Petras) angefertigten Notizen und eine von AP3 (Dörk) versandte E-Mail, die im Workshop-Nachgang bereits einige Beobachtungen stichpunktartig für die Projektpartner zusammenfasste.

Für die Reihenfolge des Kodierungsprozesses (Kapitel 3) wurde den vorliegenden Dateien verschiedene Wichtigkeitsgrade zugemessen, je nach inhaltlichem Anteil der ExpertInnen. Dabei wurde der Abschlussdiskussion am meisten Gewicht beigemessen, da hier die bisherigen Workshopinhalte systematisch von allen Teilnehmenden adressiert und bewertet wurden. Das entsprechende Audio wurde daher vollständig transkribiert und enthält ergo die meisten Redebeiträge. In abfallender Reihenfolge wurden anschließend die Präsentation der HNA-Arbeitsweise mit Diskussion, die einzelnen Spekulationen über die im Workshop erstellten Kollagen und schließlich die Sekundärdokumente ausgewertet (siehe Tabelle 1).

Nr.	Quellenart	Titel	Anzahl Redebeiträge inkl. Moderation
1	Primär	5_Audio_Kollagen_Spekulationen_IV_Abschlussstatements	80
2	Primär	1_Audio_Praesentation_Halling_Diskussion_Einstieg	54
3	Primär	4_Audio_Kollagen_Spekulationen_III	17
4	Primär	3_Audio_Kollagen_Spekulationen_II	7
5	Primär	2_Audio_Kollagen_Spekulationen_I	5
6	Sekundär	Visualisierungsworkshop_Notizen_Petas	X
7	Sekundär	Visualisierungsworkshop_Notizen_Menzel	X
8	Sekundär	Mail_Auswertungsansätze_AP3_Doerk	X

Tabelle 1: Hierarchie des Datenmaterials.

4. Methodik

Die Evaluierung der Arbeitsschritte AP2-1 (Modellhaftes Forschungsdesign) und AP3-2 (Prototyp(en) der Visualisierung) orientiert sich an der *Grounded Theory* (vgl. Hunger/Müller 2016). Diese setzt die Einbeziehung konkreter Nutzungsszenarien der Forschungsinfrastruktur von Anbeginn voraus und ist gerade in der Visualisierung durch die Herangehensweise nach Isenberg et. al (2008) in der Evaluierung eng an die Einbeziehung der potenziellen Nutzenden gebunden.

Die vorliegende Auswertung fokussiert daher besonders auf die Einschätzung der ExpertInnen auf dem Gebiet der Sozialen bzw. Historischen Netzwerkanalyse (SNA und HNA). Demnach ergibt sich eine verdoppelte Aussagengewichtung für die Teilnehmenden mit entsprechendem fachlichen Hintergrund (siehe Tabelle 2) sowie eine gesonderte Auswertung ihrer Äußerungen (siehe Kapitel 4.2).

Projektbezug	Fachlicher Hintergrund	Score
Extern	HNA	2
Intern	HNA	2
Extern	SNA	2
Extern	Informationsvisualisierung	1
Intern	Informationsvisualisierung	1
Intern	Datenaufbereitung	1
Intern	Qualitätssicherung	1

Tabelle 2: Aussagengewichtung.

Die in Kapitel 2 beschriebenen Daten wurden zunächst in die Software MAXQDA 2018 (Release 18.2.0) importiert. Sämtliche Audiodateien wurden gemäß des in Kapitel 1 beschriebenen Ziels teiltranskribiert und in durch die MAXQDA-Fokusgruppen-Funktion nach SprecherInnen ausdifferenziert. Der o.g. Score wurde durch die Codegewichtungs-Funktion in MAXQDA erreicht.

Anschließend wurde in einem ersten Zirkel *in vivo* kodiert, d.h. es wurden Codes vergeben, die das durch die Teilnehmenden verwendete Vokabular aufgriffen (vgl. Saldaña 2016, S. 102). Daraufhin fand eine erste Code-Gruppierung statt, in der unter anderem inhaltlich verwandte Aussagen zusammengefasst und einheitlich benannt wurden. Die Codes wurden hier zur eventuellen Weiterverwendung im Projekt ins Englische übertragen. Nach dem zweiten Codierungszirkel war damit ein Codesystem geschaffen, dass nun hierarchisch geordnet, klassifiziert und beschrieben werden konnte. Die Codes sind in Anhang A definiert.

5. Ergebnisse

Das folgende Kapitel stellt zunächst die Ergebnisse der Codierung der Workshop-Beiträge durch die HNA/SNA-ExpertInnen vor und begründet auf der Basis der in Kapitel 4 vorgestellten Aussagegewichtung anschließend die Aspekte, die in die Leitfäden der Evaluierung in SoNAR (IDH) einfließen werden.

5.1 Aussagen der HNA/SNA-ExpertInnen

Als zentrales Bedürfnis wurde von drei der vier FachwissenschaftlerInnen die **Exploration** der Daten in der Forschungsinfrastruktur genannt. Konkret ging es um das Verständnis und Kennenlernen des Datenkorpus' über die Visualisierung:

Thorsten Halling: „*Unsere Idee jetzt - oder wie wir HNA verstehen, ist dieses Explorative. Also weniger das Beschreibende. Das ist auch gut, nur das Explorative, also wenn ich einfach Daten habe, wo ich im Idealfalle - erst eine Frage durch die Visualisierung SEHE.*“
 (1_Audio_Praesentation_Halling_Diskussion_Einstieg, A. 2)

Dirk Wintergrün: „*Um was es geht, ist ja erstmal die innere Konsistenz dieser Datensätze verstehen.*“
 (1_Audio_Praesentation_Halling_Diskussion_Einstieg, A. 36)

Das fasst die übergeordnete Erwartung auf, dass durch algorithmisch automatisierte Datenvisualisierung neue **Forschungsfragen sichtbar werden könnten**, die durch manuelle Möglichkeiten der Netzwerkanalyse aufgrund der Datenmenge nicht erfassbar wären. Als Voraussetzung zur Generierung dieser neuen Perspektiven wurde der **dynamische Umgang mit der Darstellungsweise** der Daten genannt. Hier wurde unter anderem der **Wechsel von Datenperspektiven** angesprochen, also die Makroperspektive, die einen Überblick über den Gesamtteil oder einen Großteil der vorhandenen Daten verschafft, sowie die Mikroperspektive, in der man einen Datenpunkt

detailliert betrachten kann. Dies setzt **unterschiedliche Einstiegspunkte** in die Forschungsinfrastruktur voraus.

Sascha Topp: „Also die Flexibilität begrüße ich auch sehr, die die Programme jetzt bieten. Weil ich nämlich dann anhalten kann, Screenshots machen kann und mir das näher anschauen kann. Also ohne das macht wirklich die Netzwerkanalyse kaum Sinn.“
(5_Audio_Kollagen_Spekulationen_IV_Abschlussstatements, A. 40)

Anhand des Zitats wird bereits deutlich, dass die übergeordneten Bedürfnisse der Datenexploration und -flexibilität in einem Großteil der Aussagen durch konkrete Feature-Vorstellungen begleitet wurden. Dazu gehört u.a. das **Festhalten und Einfrieren** von Visualisierungen, die **Farbgebung** sowie **Quellenverlinkungen** der dargestellten Datenpunkte. Der Aspekt der Flexibilität wurde begleitet vom sehr präsenten Wunsch der drei HNA-Experten nach **Filtern**, z.B. nach **geografischem Ort**. Als besonders zentrale Filterfunktion wurde die **Zeitleiste** betont:

Sascha Topp: „Aber ohne Timelines nützen mir die Visualisierungen nichts! Weder für die Analyse noch für die Vorstellung von Ergebnissen.“
(5_Audio_Kollagen_Spekulationen_IV_Abschlussstatements, A. 36)

Als weiteres Bedürfnis stellte sich vor allen bei den drei HNA-Experten die **offene und vollständige Darstellung der Daten** heraus. Am wichtigsten wurde hier von drei der vier HNA/SNA-ExpertInnen die **visuelle Auszeichnung von Datenunsicherheiten** genannt, aber auch die **Aktualität der Daten** wurde mehrfach angesprochen. Darüber hinaus wurde die **Darstellung der Datendichte** wurde als zentral empfunden:

Dirk Wintergrün: „Das macht auch Sinn, das vorher [...] zu wissen [...], ob man eine Frage überhaupt stellen kann.“
(5_Audio_Kollagen_Spekulationen_IV_Abschlussstatements, A.22)

Zum einen wurde in diesem Zusammenhang der Wunsch nach algorithmisch aufbereiteter Darstellung zum Datenpotenzial geäußert, konkret z.B. zur **Wahrscheinlichkeit, zum Typ und zur Gewichtung von Relationen** zwischen Entitäten. Hierzu gehören auch die sogenannten **schwachen Verbindungen (weak ties)**, also in der Regel einfache Verbindungen zwischen zwei erkennbaren Anballungen vielfach miteinander verbundener Entitäten (*hubs*). Zum anderen gab es von zwei der HNA-Experten Äußerungen zur **heuristischen Herangehensweise** an die Daten:

Dirk Wintergrün: „Aber manchmal ist das mehr so ein intuitives Problem. Das man sich eigentlich nicht denkt, also aus anderen Quellen ergänzende Quellen. [...] [D]as ist mehr als Statistik, das ist eher auch so ein [...] heuristisches Experimentieren mit den Daten.“
(1_Audio_Praesentation_Halling_Diskussion_Einstieg, A. 26)

5.2 Aspekte für die Leitfäden

Während des Codierungsprozesses wurde deutlich, dass sich die Äußerungen der Teilnehmenden in zwei zentrale Gruppen aufteilen lassen:

1. Codegruppe I: Bedürfnisäußerungen (Needs, 17 Codes)

Hierbei handelt es sich um Erläuterungen der eigenen Arbeitsweise durch die HNA/SNA-ExpertInnen sowie allgemeine Einschätzungen und Spekulationen zum Potenzial einer Forschungsinfrastruktur für historische Netzwerke. In dieser Gruppe sind Voraussetzungen für die Arbeit der Zielgruppe mit der Infrastruktur zusammengefasst, die eher in abstrakter Form geäußert wurden.

2. Codegruppe II: Äußerungen zu konkreten Funktionen (Features, 12 Codes)

Bei der zweiten Gruppe dagegen handelt es sich um konkrete Aussagen zu Anforderungen an eine Forschungsinfrastruktur für historische Netzwerke. Hier sind Äußerungen zusammengefasst, die sich auf unmittelbare Vorstellungen zum Umgang mit algorithmisch unterstützter wissenschaftlicher Arbeit mit historischen Daten beziehen.

Beide zentrale Codegruppen beinhalten diverse Über- und Untercodes (siehe Tabelle 3).

Codegruppe I: Needs	Score Verg.	Anz. Pers.	\bar{X}	Codegruppe II: Features	Score Verg.	Anz. Pers.	\bar{X}	
Data Density	26	4	15	Filters	16	3	9,5	
Heuristic Approach	9	4	6,5	Timeline	22	4	13	
Data Potential	27	4	15,5	Location	8	3	5,5	
Data Explanations	16	4	10	Freeze Option and Citation	13	4	8,5	
Uncertainty	25	7	16	Source Linking	6	4	5	
Entry Points	17	4	10,5	Type of Relation	5	5	5	
New Perspectives	17	2	9,5	Weak Ties	8	6	7	
Data Perspectives: Micro vs. Macro	7	6	6,5	Weight of Relation	11	5	8	
Exploration	13	5	9	Network Statistics	5	2	3,5	
Flexibility	7	2	4,5	Probabilities for automatic data enrichments	18	4	11	
Innovation vs. Digital Habits	7	3	5	Semantic Search	4	1	2,5	
Data Recency	6	3	4,5	Colors	4	1	2,5	
User Centered Design	3	1	2					
Form follows Function	10	1	5,5					
Task Approach	3	1	2					
Target Group	9	4	6,5					
Use Cases	14	4	9					

Tabelle 3: Codes mit Vergabehäufigkeit (=Score Verg.; doppelte Zählung der Teilnehmenden mit dem Score 2) und Anzahl der Personen, die sich zu diesem Aspekt geäußert haben. Die Codes sind geordnet nach Vergabehäufigkeit der Übercodes (**fett**). Zentrale Codes für die Leitfäden der Evaluierungen durch AP4 sind farblich hinterlegt. Die Codes sind in Anhang A (Codebuch) definiert.

Die Anzahl der Code-Vergabe im Datenmaterial lässt Schlüsse auf die Präsenz des Codes in den aufgezeichneten Workshop-Sequenzen zu. Kombiniert mit der Anzahl der Personen, die sich zum Code-Thema äußerten, ergibt sich eine Gewichtung der Codes anhand des Mittelwerts. Da es sich um eine qualitative Auswertung handelt, deren zugrundeliegender Datensatz lediglich ein kleines, nicht repräsentatives Sample einfasst, dient dieser Mittelwert lediglich als Orientierung zur Auswahl der zentralen Codes für die Leitfadenerstellung im weiteren Evaluierungsverlauf. Differenziert wird außerdem in die Einbeziehung zu den Interview- und/oder Beobachtungsleitfäden.

Dementsprechend wurden z.T. Codes ausgeschlossen, die für die Evaluierung der Visualisierungen nicht zentral sind, wie z.B. die Codes zum nutzerzentrierten Design. Andere Codes mit vergleichsweise geringem Mittelwert wurden dagegen einbezogen, z.B. die Verlinkung von Datenquellen, da diese als zentral für die wissenschaftlichen Anforderungen der HNA eingestuft wurden.

Sowohl für die Interview- als auch für die Beobachtungs-Leitfäden (IBL) gelten folgende Aspekte als zentral:

IBL1: Freeze Option and Citation, insb. Source Linking

Befragung ProbandInnen der Evaluierungen III zur Bedeutung von Speicher- und Zitationsfunktionen sowie Quellenangaben. Im Fall, dass diese in die Prototypen eingebunden sind: Beobachtung der ProbandInnen der Evaluierungen IV zu deren Nutzung.

IBL2: Filters, insbes. Timeline und Location

Beobachtung der ProbandInnen der Evaluierungen III und IV zur Nutzung von Filtern allgemein, insbesondere Karten, Zeitleisten oder anderen geografischen/temporären Visualisierungen. Erfassung von Äußerungen zur Bedeutung von Karten, Zeitleisten oder anderen geografischen/temporären Filtermöglichkeiten in den Befragungen.

IBL3: Probabilities for automatic data enrichments

Befragung ProbandInnen der Evaluierungen III zur Bedeutung zur automatisierten Ergänzung der Daten durch probabilistische Methoden. Im Fall, dass diese in die Prototypen eingebunden ist: Beobachtung der ProbandInnen der Evaluierungen IV zu deren Nutzung.

Für die Interview-Leitfäden (IL) gelten folgende Aspekte als zentral:

IL1: Data Density und Data Potential

Befragung der ProbandInnen der Evaluierungen III und IV zur Datendichte. Erfassung von Äußerungen zum Bedürfnis, in der Forschungsinfrastruktur enthaltene Daten und deren Lücken (*missing information*) zu erfassen.

IL2: New Perspectives

Befragung der ProbandInnen der Evaluierungen III und IV zur Motivation, eine Forschungsinfrastruktur im Bereich der Sozialen- und Historischen Netzwerkanalyse zu nutzen, mit besonderem Fokus auf den Aspekt der Generierung neuer Forschungsfragen durch automatisierte Verfahren (etwa im Kontext der *serendipity*).

Für die Beobachtungs-Leitfäden (BL) gelten folgende Aspekte als zentral:

BL1: Entry Points

Beobachtung der ProbandInnen der Evaluierungen III und IV auf deren Einstiegspunkte in die Forschungsinfrastruktur.

BL2: Exploration

Beobachtung der ProbandInnen der Evaluierungen III und IV zum explorativen Vorgehen in den Prototypen der Forschungsinfrastruktur.

BL3: Data Explanations, insb. Uncertainty

Beobachtung der ProbandInnen der Evaluierungen III und IV auf die Wahrnehmung oder Bedürfnisäußerungen zu erklärenden Elementen und Texteinsteigen in die Forschungsinfrastruktur, insbesondere der Auszeichnung von vorhandenen Datenunsicherheiten.

5.3 Diskussionspunkte

Die Workshopkonzeption folgte der Herangehensweise des nutzerzentrierten Designs. Als besonders wichtig wurde dabei von den Teilnehmenden die Einbeziehung konkreter Anwendungsszenarien (*use cases*) in die Entwicklung der Visualisierungen eingestuft (4 von 10 erwähnten dies explizit).

Eine klare Zielgruppe wurde im Zuge des Workshops nicht definiert (vgl. Tagstellen *Target Group*). Konsens bestand allerdings darüber, dass die Visualisierungen im Sinne des o.g. nutzerzentrierten Designs an der Forschungsweise von HNA-FachwissenschaftlerInnen orientiert sein sollte. Hier traten sowohl Überschneidungen als auch leichte Differenzen zur sozialwissenschaftlichen SNA auf (vgl. 5_Audio_Kollagen_Spekulationen_IV_Abschlussstatements, A. 54).

In diesem Zusammenhang entstand auch ein Dialog über Vermittlung der Daten an einsteigende NutzerInnen, also z.B. textuelle Erklärungen zu den angezeigten Daten (vgl. Tagstellen *Data Explanations*). Ebenfalls angesprochen wurde das Spannungsfeld zwischen Innovationen in der Informationsvisualisierung und dem Aufgreifen von Anwendungsgewohnheiten der Nutzenden aus bereits etablierten Forschungstools.

Darüber hinaus wurden an zwei Textstellen die Vorteile einer semantischen Suche nach Themenkomplexen erwähnt. Hier entstand die Diskussion, ob deren Auslesung innerhalb des Projekts auch aus den Volltexten möglich sei.

6. Limitierungen

Die vorliegende Auswertung bezieht sich lediglich auf ein keines Sample von zehn Personen, von denen vier im Sinne des Evaluierungskonzepts (vgl. Menzel/Petras 2019, S. 8) als ExpertInnen eingestuft wurden. Da es sich bei der Auswertung lediglich um eine von zwei Vorstudien zur Leitfadenentwicklung handelt, werden deren Aussagen im Laufe des Evaluierungsprozesses durch weitere Einschätzungen von ExpertInnen ergänzt und dadurch validiert oder falsifiziert.

7. Zusammenfassung

Anhand der Auswertungsergebnisse sollten gemäß des in Kapitel 1 beschriebenen Ziels durch AP4 die anhand der Codierungshäufigkeiten als zentral eingestuften Bedürfnisse und Funktionen in die Leitfäden der Evaluierungen III und IV einfließen.

Zentrale Bedürfnisse:

- Datenpotenzial und Datendichte erkunden
- Neue Forschungsfragen generieren
- Möglichkeiten zum explorativen Vorgehen mit den Daten.

Es wurde darüber hinaus deutlich, dass die Forschungsinfrastruktur durch konkrete Features den Ansprüchen wissenschaftlicher Arbeitsweise genügen muss.

Zentrale Funktionen:

- Möglichst aktuelle Datenlage (Live-Synchronisation mit Datenpools)
- Diverse Einstiegspunkte in die Daten unter anderem durch Filter (z.B. Zeitleiste und Karten)
- Zitationsfähigkeit der individuellen Visualisierungsansichten
- Klare Quellenangaben zum Datenstand (Pool, Einheiten) und zu den einzelnen Datenpunkten
- Klare Auszeichnung von Datenunsicherheiten

- Zusätzliche Verknüpfungen der Entitäten durch mathematisch errechnete Verbindungen (Wahrscheinlichkeiten).

Diese Punkte werden in Form von Aufgaben oder Fragen in die vorgesehene Nutzerstudie einbezogen, indem sie in die Beobachtungs- und Interviewleitfäden einfließen. Eine Rücksprache zu den Ergebnissen mit AP3 findet statt.

8. Referenzen

Hunger, Ina; Müller, Johannes (2016): Barney G. Glaser/Anselm L. Strauss: The Discovery of Grounded Theory. Strategies for Qualitative Research, Aldine Publishing Company: Chicago 1967, 271 S. (dt. Grounded Theory. Strategien qualitativer Forschung, Bern: Huber 1998, 270 S.). In: Samuel Salzborn (Hg.): Klassiker der Sozialwissenschaften. Wiesbaden: Springer Fachmedien Wiesbaden, S. 259–262. DOI: 10.1007/978-3-658-13213-2_59

Isenberg, Petra; Zuk, Torre; Collins, Christopher; Carpendale, Sheelagh (2008): Grounded evaluation of information visualizations. In: Proceedings of the 2008 conference on BEyond time and errors novel evaLuation methods for Information Visualization - BELIV '08. Florence, Italy: ACM Press, S. 1. Online verfügbar unter <http://portal.acm.org/citation.cfm?doid=1377966.1377974>

Menzel, Sina; Petras, Vivien (2019): Evaluierungskonzept für SoNAR (IDH). Version I vom 27.09.19. Projektintern verfügbar unter:

<https://onedrive.live.com/?authkey=%21AH5z6Zyly9DgH80&cid=3B9129E4B3C7C3C9&id=3B9129E4B3C7C3C9%21362&parId=3B9129E4B3C7C3C9%21357&o=OneUp>

Saldaña, Johnny (2016): The coding manual for qualitative researchers. 3. edition. Los Angeles, Calif., London, New Delhi, Singapore, Washington DC: SAGE.

Anhang A: Codebuch

Codegruppe I: Needs

Die Codegruppe *Needs* beinhaltet alle Codes, die sich auf Bedürfnisäußerungen im Workshop beziehen. Hier geht es insbesondere um die Beschreibung der Arbeitspraxis sowie Hoffnungen und Erwartungen an eine als hilfreich empfundene Forschungsinfrastruktur, ohne notwendigerweise Lösungsvorschläge oder konkrete Vorstellungen zu äußern.

Data Recency

Dieser Code bezieht sich auf Äußerungen zur Aktualität der Daten, auf deren Basis die Visualisierungen generiert werden.

Entry Points

Dieser Code bündelt Aussagen zu denkbaren Einstiegspunkten in die Forschungsinfrastruktur.

New Perspectives

Unter diesem Code sind Aussagen von Teilnehmenden zusammengefasst, in denen die Erwartung oder Hoffnung ausgedrückt wird, dass die Visualisierungen neue Forschungsfragen generiert, neue Perspektiven eröffnet und damit ein Zugang zu den Daten erzeugt wird, der durch nicht-maschinell unterstützte kognitive Arbeit kaum möglich ist.

Data Perspectives: Micro vs. Macro

Untercode zu *New Perspectives*. Dieser Code beinhaltet Äußerungen zur Menge der Daten, die in einer Visualisierung überblickt werden (können), sowie die Potenziale von Mikro- und Makroperspektiven und Zoomfunktionen.

Data Density

Codiert Stellen, die sich auf die Datendichte beziehen, also die Übersicht über in der Forschungsinfrastruktur enthaltene Daten und deren Lücken (*missing information*).

Heuristic Approach

Untercode zu *Data Density*. Dieser Code beschreibt innerhalb der Problematik der Datendichte Aussagen, in denen der Wunsch geäußert wird, auch Darstellungsformen nicht statistisch darstellbarer "gefühlter" Datenlücken zu haben.

Data Potential

Untercode zu *Data Density*. Dieser Code bezieht sich auf Aussagen, die sich innerhalb der Visualisierung der Datendichte die Darstellung des Datenpotenzials wünschen, d.h. Schlüsse auf den Möglichkeitsraum der Daten zulässt.

Exploration

Dieser Code fasst Äußerungen zur explorativen Herangehensweise an die Datenmenge innerhalb der Forschungsinfrastruktur zusammen.

Flexibility

Dieser Code fasst Aussagen zusammen, die den dynamischen bzw. flexiblen Umgang mit den Daten

adressieren.

User Centered Design

Dieser Code bündelt zum einen allgemeine Aussagen zur Herangehensweise des nutzerzentrierten Designs und beinhaltet zum anderen als Übercode spezifischere Codes zu diesem Thema.

Form follows Function

Untercode zu *User Centered Design*. Dieser Code bezieht sich auf Aussagen, die sich auf das Zusammenwirken bzw. Spannungsfeld von Ästhetik und Funktionalität von Design beziehen.

Task Approach

Untercode zu *User Centered Design*. Dieser Code beinhaltet Textstellen aus den Beobachtungsprotokollen, die Herangehensweise der Fachwissenschaftler an die Gestaltungsaufgabe im Workshop thematisieren.

Target Group

Untercode zu *User Centered Design*. Der Code fasst Aussagen zusammen, die sich mit der angedachten Zielgruppe für die Forschungsinfrastruktur befassen.

Use Cases

Untercode zu *User Centered Design*. Dieser Code beinhaltet Aussagen zur Erstellung konkreter Anwendungsszenarien der Forschungsinfrastruktur.

Data Explanations

Dieser Code bündelt zum einen allgemeine Aussagen zu erklärenden Elementen und Texteinstiegen in die Forschungsinfrastruktur und beinhaltet zum anderen als Übercode spezifischere Codes zu diesem Thema.

Uncertainty

Untercode zu *Data Explanations*. Dieser Code bezieht sich auf Aussagen zum Umgang mit Datenunsicherheiten, inklusive *predicted errors* und Inkonsistenzen.

Innovation vs. Digital Habits

Dieser Code beschreibt Textstellen, in denen das Spannungsfeld zwischen digitalen Gewohnheiten bzw. der durch bekannte digitale Strukturen erzeugten Erwartungshaltung und digitaler Innovation angesprochen wird.

Codegruppe II: Features

Die Codegruppe *Features* beinhaltet alle Codes, die sich auf Anwendungsäußerungen im Workshop beziehen. Hier geht es insbesondere um die Beschreibung konkreter gewünschter Features, sowie Hoffnungen und Erwartungen an eine als hilfreich empfundene Forschungsinfrastruktur, verknüpft mit konkreten Lösungsvorstellungen.

Network Statistics

Dieser Code umfasst Aussagen zu statistischen Visualisierungen innerhalb von Netzwerkdaten. Dazu gehören u.a. der Clustering-Koeffizient, die Closeness-Zentralität, die Degree-Distribution, die Degree-Zentralität und die Betweenness-Zentralität. *Network Statistics* beinhaltet darüber hinaus als

Übercode die Aussagen zu Verbindungswahrscheinlichkeiten.

Probabilities for automatic data enrichments

Untercode zu *Network Statistics*. Dieser Code bezieht sich auf Äußerungen zu potenziellen Links zwischen Entitäten, die automatisiert durch die Errechnung von Wahrscheinlichkeiten dargestellt werden könnten.

Type of Relation

Dieser Code bündelt Aussagen zur Darstellung von Beziehungsarten und beinhaltet gleichzeitig als Übercode Aussagen zu deren Gewichtung.

Weak Ties

Untercode zu *Type of Relation*. Dieser Code umfasst Aussagen zur Visualisierung von schwachen (in der Regel nur durch eine Verbindung manifestierten) Verbindungen in Netzwerken, die zwischen dichteren Einzelnetzwerken ("Bällen"; "Hubs") bestehen.

Weight of Relation

Untercode von *Type of Relation*. Dieser Code bezieht sich auf Äußerungen, die mit der Gewichtung von Verbindungen zwischen Entitäten zusammenhängen.

Timeline

Unter diesem Code sind Aussagen zur Notwendigkeit einer Zeitleiste zusammengefasst.

Freeze Option and Citation

Dieser Code beinhaltet Textstellen zur Zitationsfähigkeit und zum Festhalten von Informationen des Datenstandes zum Abrufzeitpunkt.

Source Linking

Untercode zu *Freeze Option and Citation*. Dieser Code bezieht sich auf Textstellen, an denen die Verlinkung von Datenquellen in der Visualisierung angesprochen werden.

Filters

Dieser Code fasst Aussagen über Filteroptionen zusammen.

Semantic Search

Dieser Code bezieht sich auf Textstellen, in denen von einem semantischen Sucheinstieg (z.B. über Sachbegriffe) die Rede ist.

Colors

Dieser Code bündelt Aussagen zur Farbgebung innerhalb der Forschungsinfrastruktur.

Location

Dieser Code bezieht sich auf Äußerungen zur Bedeutung der geografischen Zusatzinformationen zu einzelnen Verbindungen, z.B. anhand wissenschaftlicher Anwendungsfälle wie der *Nationalen Geschichtssteuerung* oder dem *Stille-Post-Effekt*.

Anhang B: 1_Audio_Praesentation_Halling_Diskussion_Einstieg

1 [Vortrag Thorsten Halling]

2 **Thorsten Halling:** ...Mehrwert der Beschreibung. Und ich glaube, das ist auch das Hauptproblem, das, was uns so antreibt: Du (Anm. gemeint ist Mark-Jan Bludau) hast eben wunderbare Visualisierungen gehabt, die gibt es eben in der HNR oder HNA auch, häufig jetzt schon. Und ganz oft sage ich: "Schön!" - Aber es ist eben einfach nur ästhetisch, aber ich habe keinen Erkenntnisgewinn, wenn ich da draufgucke. Da kann ich eine Stunde draufgucken und ich verstehe es erst, wenn ich die Erklärung dazu lese. Und dann ist die Frage: Warum brauche ich die Grafik, wenn ich dann doch den Text lese? Und ist da überhaupt - und auch dann erst mir überhaupt bewusst wird, da wird etwas geschrieben. Unsere Idee jetzt - oder wie wir HNA verstehen, ist dieses Explorative. Also weniger das Beschreibende. Das ist auch gut, nur das Explorative, also wenn ich einfach Daten habe, wo ich im Idealfalle - erst eine Frage durch die Visualisierung SEHE, wenn du mal weiter gehst (Anm. an Mark-Jan Bludau mit Bezug auf die Präsentationsfolie. Gezeigt wird eine Folie mit einem Netzwerk, in dem sich viele Verbindungen um einen Knoten ballen.) das ist zum Beispiel so etwas, was jetzt gerade mein Kollege über die Neurologen gemacht hat. Das ist ganz einfach. Das ist ja ganz simpel, ist überhaupt nicht hübsch und gar nicht beeindruckend als Netzwerk - was aber der Punkt ist: Wir haben die einzelnen Biografien zu diesen Leuten zusammengestellt, das sind alles Neurologen, die nach 1945 den Diskurs in der Bundesrepublik bestimmt haben. Und hier kommt es darauf an: Was haben die im NS gemacht und wie hingen die zusammen? Und erst durch diese Visualisierung wurde klar, dass dieser Pitte (Anm. besagter zentraler Knoten: Person), der war - alle anderen haben bei dem gearbeitet. Und jetzt sieht man: Wann ist der in die NSDAP eingetreten und wann alle anderen. Also der war einfach der spiritus rector, die waren bei dem halt irgendwie Mitarbeiter, die haben mehr oder weniger zusammengearbeitet, unter ihm, mit ihm, das ist unterschiedlich. Aber ER hat offensichtlich da die Initiative. Also das ist Interpretation. Aber dieses, dieses Ding - das haben wir vorher aus dem Biografischen nicht gesehen. Das ist niemandem aufgefallen, das ist erst da aufgefallen. Und das ist jetzt nur ein ganz einfaches Beispiel, aber so etwas ist etwas, wo wir denken: Da hat die HNA für uns in unserem Erkenntnisprozess einen echten Mehrwert. Da kommen wir weiter. Also wir produzieren also neue Fragen. Also: Warum ist das so und ja. Also das ist glaube ich das, was uns besonders antreibt. Also, wenn ich von "uns" spreche, dann ist das immer Heiner [Fangerau, nicht anwesend] und ich und auch noch die anderen, die noch bei uns mitarbeiten. Wir sind die Verrücktesten am Institut, was HNA angeht.

3 Okay, wenn du noch mal weiter machst (Anm. an Mark-Jan Bludau mit Bezug auf die Präsentationsfolie). Also hier geht es jetzt darum: Was tun wir in diesem SoNAR-Projekt. Das habe ich ja eben auch schon ein bisschen angedeutet.

4 [Erklärung vom Aufgabenbereich von AP2 im Projekt für die projektexternen Anwesenden]

5 Hier das ist jetzt so ein Physiologenstammbaum, wie man es traditionell macht. Also Schüler-Lehrer-Beziehungen. Aber es ist eben ein Stammbaum. [...] Und hier fehlen uns eben die Verbindungen untereinander. Da kommt man eigentlich nicht rein. Das ist aber das, was man machen konnte relativ einfach.

6 **Vivien Petras:** Wenn du sagst: "Beziehungen unvereinbar", meinst du zwischen den parallelen Leuten, die da sind?

7 **Thorsten Halling:** Ja, genau! Das ist ja auch so, dass hier dieser [...] also auf jeden Fall in der 2. und 3. Reihe stehen Menschen, die auch in den Stammbäumen von anderen Lehrern auch drin sind. Also die haben nicht nur einen akademischen Lehrer, wie wir alle, sondern in der Regel hat man mehrere. Und hier kommt - hier ist es eben auch: Wer hat bei wem mal gearbeitet? Die haben auch mal gewechselt und...also wie gesagt, das ist hier sehr auf dem Stammvater fokussiert und alles andere fällt eben raus bei dieser Art von Darstellung. [...]

- 8 So, also hier das ist das eben, was im Prinzip die Dinge, die uns interessieren: Arten der Beziehungen, was für eine Qualität kann man aus diesen Beziehungsarten herauslesen? Das ist natürlich ein großer Unterschied, ob man jetzt zum Beispiel verwandt ist oder ob man Korrespondenzpartner ist oder ob man sich auf einem Kongress mal bei einer Kaffeepause gesehen hat. Das sind alles Beziehungen, die aber natürlich in der Interpretation unterschiedliches Gewicht haben. Das ist zum Beispiel etwas, was wir in der Visualisierung eben sehr gerne hätten. Dass man eben einfach nicht nur sagt, so wie die hier (Anm. Verweis auf aktuelle Folie), die haben Beziehung A-B, sondern unterschiedliche Gewichtungen, je nach Forschungsfrage. Also, was interessiert mich? Das ist eben auch nicht statisch. Beispielsweise könnte man an das Gelbe wird zum Beispiel dargestellt, wer war zum Beispiel maßgeblich bei Lehrstuhlbesetzungen? Da sind andere Arten der Beziehungen relevant, als, wenn du jetzt sagst: Der ist für die und die Forschungsfrage maßgeblich. Also das ist im Prinzip, das sind dieselben Personen, die du anders gewichtetest. [...]
- 9 Und hier das ist eben auch, das ist wiederum, das hatten wir jetzt eben auch bei in der dritten Spalte ist Nathan Zuntzt, mit dem wir uns jetzt beispielhaft mal auseinandergesetzt haben. Da habe ich jetzt mal durchgespielt. Einfach ganz simpel: Ich habe den Wikipedia-Artikel genommen, der ist zu dem relativ ausführlich, habe da analysiert: Was für Relationen stecken da drin [...] und hier, das ist eine ganz einfache Tabelle, das ist eine Rohfassung. Hier geht es einfach nur darum, was haben wir in dem Wiki-Eintrag für Informationen und wie werden die in der GND/DNB/ZDB abgebildet? Werden sie abgebildet oder werden sie eben nicht abgebildet? Und irgendwie - also das ist für uns eine Probebohrung: Was können die Daten, was können sie eben nicht? Also das - wie gesagt - für euch die Visualisierung (Anm. an AP3 gewandt) das ist unsere Idee: Ihr - auch bei der Datenaufbereitung (Anm. an Elena Leitner gewandt) - könnte eine Möglichkeit geschaffen werden, unterschiedliche Fehlstellen zu benennen? Einerseits die Informationen, die da drin sind werden also visualisiert und dann hat man zum Beispiel in der GND - da hat die Elena die unterschiedlichen Kategorien mit unterschiedlichen Codes herausgearbeitet, also Affiliation, Beziehungen beruflich/familiär und so weiter. Und dann wäre eben zu visualisieren, welche Informationen KÖNNEN drin sein, dazu gibt es aber keinen Eintrag. Aber sie KÖNNEN theoretisch - also als Code sind sie da drin, auch, wenn man sie dort nicht eingetragen hat. Also die GND ist ja kein biografisches Nachschlagewerk, sondern eigentlich nur so ein Identifier. Und je nach dem, wer die GND-Daten eingibt - also wenn es ein Bibliothekar ist, sind sie in der Regel weniger - und wenn es jemand ist, der eben auch Historiker ist, dann sind es vielleicht mehr. Also das ist meine These, die kann ich nicht belegen. Also jedenfalls das wäre die erste Fehlstelle.
- 10 Und die zweite Fehlstelle, die man eventuell visualisieren könnte, sind die, die da auf keinen Fall drinstecken können. Die wir uns zwar wünschen, aber die da nicht angelegt sind. Und ich glaube, wenn man so etwas visualisieren könnte, dann würden wir daraus natürlich auch wieder einen großen Mehrwert ziehen. Aber das ist sicherlich eine echte Herausforderung, also auch an die Datenerhebung schon mal. Also ihr seid da glaube ich Lead schon mal. Also ich glaube, das eigentliche Visualisieren ist vielleicht gar nicht das Problem. Also ich...
- 11 **Marian Dörk:** Ach komm! (schmunzelt, allgemeines Gelächter)
- 12 **Thorsten Halling:** Nein, das man Beziehungen beispielsweise, also einfach eine Kante grau macht - hinterlegt. Da ist eigentlich eine Beziehung oder wir vermuten sie da. Das könnte man eben auch versuchen mit einem Algorithmus zu berechnen: Wie hoch ist die Wahrscheinlichkeit, dass da eine sein könnte? Und die dann einfach irgendwie zu visualisieren, auch wenn wir sie nicht definitiv haben.
- 13 **Sascha Topp:** Darf ich ganz kurz querfragen? Dann müsstet ihr euch ja vorher auf ein bestimmtes Set als Grundlage einigen.

- 14 **Thorsten Halling:** Das haben wir ja.
- 15 **Sascha Topp:** Also nicht nur Datensätze zu soundsovielen Personen, sondern das Kriterienset muss zwischen euch fest ausgehandelt werden. Damit die Visualisierer überhaupt wissen, welche Geister sie suchen.
- 16 **Thorsten Halling:** Ja, das ist die Praktik von Forschungsfragen. Die Idee ist, wir haben eine bestimmte Forschungsfrage eben zum Spielen. Aber das Ziel muss sein, dass es unabhängig von dieser konkreten Forschungsfrage funktioniert. Dass auch du später mit deiner ganz anderen - irgendwas zu [unverständlich] da eingibst und auch ein gutes Ergebnis bekommst. Ohne irgendwelche Physiologen des 19. Jahrhunderts. Das ist einfach nur eine Fingerübung, wie hier auch. Also wir müssen von vorne herein so angehen, dass da ein Algorithmus eben das erkennt, hier gibt es eine theoretische Option, weil dieses Feld, die Relation ist da, aber es ist eben nicht gefüllt. Und eben auch - aber anhand der anderen Daten, die halt da sind, irgendwie zu berechnen, das hat eine bestimmte Wahrscheinlichkeit, dass es diese Relation auch gibt. Weil es die und die Relation gibt. Wie gesagt, ich bin kein Informatiker, das ist ein Wunschkonzert.
- 17 **Susanne Helm:** Ich glaube, also für mich hört sich das auch eher nach statistischer Methodik an, als nach Visualisierungsmethodik, oder?
- 18 **Thorsten Halling:** Ja, ja, das ist der Schritt vor der Visualisierung. Da gucke ich die Elena an (allgemeines Gelächter). Und das ist ja eben das Besondere an diesem Projekt, dass wir eben aus drei unterschiedlichen Perspektiven auf das gleiche Problem gucken.
- 19 **Vivien Petras:** [...] So, das ist ja ein bisschen abstrakter das, was Thorsten gesagt hat, was er sucht (Anm. bezieht sich auf Folie 6 in Thorsten Hallings Präsentation). Und jetzt würde ich euch beide mal fragen (Anm. an Sascha Topp und Dirk Wintergrün gerichtet), ob das mit euch resoniert. Für die Forschungsfragen, die ihr bearbeitet, besonders du, Sascha.
- 20 [...]
- 21 **Dirk Wintergrün:** Also ich habe eine Anmerkung dazu. Also ich stimme dir auch total zu, dass ein Schwerpunkt bei der ersten Visualisierung sozusagen ist [unverständlich]. Dass kann man gar nicht unterbewerten, wie wichtig das ist. Und deshalb meinte ich auch, ich glaube nicht, dass das geht, wie Sascha meinte, nämlich über die Frage, was man dann für Beziehungen zwischen den Daten sucht. Das passiert ganz oft erst in dem Moment, wo du zum ersten Mal auf diesen Haufen draufguckst. Das heißt, das Kriterium, wonach man das sortiert und was man dann zusammenfindet, das weiß man manchmal gar nicht, was wichtig ist oder nicht. Das sieht man in dem Moment, wo man so ein Bild dann sieht. Wie falsch oder gut das ist eine ganz andere Frage. Deshalb kann man sich nicht vorher auf ein Set einigen. Ich denke, was bei uns sehr gut in vielen Fragen passt, ist die Frage nach *missing information*. Das ist wirklich, das in der HNA viel wichtiger ist, als bei den Zahlnetzwerken. Und diese missing - also, auf das, was fehlt aufmerksam zu werden ist häufig viel entscheidender, als das, was man da tatsächlich sieht.
- 22 Und ich glaube eine ganz andere, pragmatische Frage, die wir immer häufiger sehen ist: Fehler in den Daten. Also eine von den wirklich auffälligsten Geschichten, die wir, als wir die ersten großen Netzwerke gemacht haben, ist: Es fallen einem Personen auf und davon sind 90% der [unverständlich] aber falsch. Und es gibt zentrale [unverständlich] die einfach falsch disambiguert (?) werden. Dieser ganze Aspekt...
- 23 **Thorsten Halling:** Jetzt ist die Elena gerade weg (schmunzelt) - die hat das gemerkt - aber das ist das Problem, womit sie sich gerade rumschlägt. Sie schaut sich gerade die Ausgangsdaten an

und da sind halt sehr viele Probleme. Also, die man wahrscheinlich irgendwie beheben kann...also...

- 24 **Dirk Wintergrün:** Manche sind ja auch systematisch, also da weiß man also, man muss damit spielen können, ob diese Personen wirklich dieselben sind oder nicht. Also über solche Möglichkeiten wie potenzielle Links zum Beispiel, das fand ich auch immer wieder ein ganz wichtiges Element mit dem Spielen von Daten. Also dass man sagen kann: "Okay, wenn ich das jetzt wüsste oder wenn das jetzt so wäre, sähe das jetzt systematisch ganz anders aus." oder "Ist das jetzt dieselbe Person?" [unverständlich] Das ist, wo man zumindest - zurzeit noch jedenfalls - am meisten aus den Daten herausholen kann. Und das passt bei uns auch. Ganz deutlich. Zumindest in dem großen Projekt, in dem Sascha ist oder auch in meinem mit den Universitätsleuten zum Beispiel erstmal drauf und dann guckt man: Was fehlt einem da? Was unterscheidet zum Beispiel die unterschiedlichen nationalen Communities? Die aber ganz oft nur deshalb nationale Communities so [unverständlich] darstellen, wo man gleich sieht: Das kann eigentlich nicht sein! Da müsste es Verbindungen eigentlich geben, also es gibt Quellen...
- 25 **Sina Menzel:** Und da meinst du auch die Wahrscheinlichkeiten, wenn du sagst "potenzielle Links" - oder?
- 26 **Dirk Wintergrün:** Ja, ich weiß nur nicht, ob man da mit der Statistik so hin kommt, weil du hast ja nicht - klar, man kann *so missing links*, das klappt manchmal natürlich schon. Wenn man weiß, da ist eine Community und man kennt ganz viele Beteiligte, dann ist das schon klar. Aber manchmal ist das mehr so ein intuitives Problem. Das man sich eigentlich nicht denkt, also aus anderen Quellen ergänzende Quellen. Das muss eigentlich sein, dass die da - also das ist mehr, als Statistik, das ist eher auch so ein, also ich würde sagen so ein heuristisches Experimentieren mit den Daten.
- 27 **Thorsten Halling:** Aber das ist genau das - das ist ja, das ist genau der Punkt, der den Unterschied unseres normalen Arbeitens ausmacht und das in dem Projekt. Das normale Arbeiten ist so, wie du gesagt hast: Wir gucken uns halt alles an, was es gibt zu irgendeinem Thema. Wir suchen, wühlen und gucken und hier noch irgendeine Korrespondenz...
- 28 Jetzt haben wir ja ein - also bis jetzt - ein relativ [unverständlich] Datenset mit einer bestimmten Art von Daten, nämlich aggegierte, hoch normierte Daten. GND-Daten, ZDB- also die sind sehr normiert. Deshalb funktioniert das mit der Statistik eben vielleicht auch besser, als wenn man hier noch ein Archiv und da noch ein Archiv. Und wie gesagt: Wir haben große Lücken, aber die sind von vorne herein angelegt. Die sind einfach da. Ich habe - am Anfang habe ich gedacht: Das ist ja blöd, wenn ich nicht in das Archiv gehen kann. Also dieser typische Historiker: Ich muss doch jeden Brief finden! Ich habe jetzt Kalliope - da sind eben aber nicht alle Briefe drin! Sondern da sind eben aus bestimmten Sammlungen - ABER ich habe mich inzwischen von dieser Idee "Ich will alles sammeln!" - dieser Jäger und Sammler - sondern, ich will jetzt erstmal gucken: Was können DIE? Was können diese Daten?
- 29 Und sie sind eben, weil sie normiert sind, sie sind fehlerhaft, also zum Beispiel bei GND-Einträgen. Ich weiß nicht, wie das bei euch ist, aber zum Beispiel: Ich habe das einem Kollegen demonstriert. Also er wusste überhaupt nicht, was GND-Daten sind, da habe ich gesagt: "Du hast auch einen GND-Eintrag!" Da habe ich reingeguckt und der hatte aber drei! Der gleiche Mann! Aber einfach...
- 30 **Marian Dörk:** Hat der es gut!
- 31 **Thorsten Halling:** ...nee, das war einfach, weil da jemand nicht richtig recherchiert hatte, als er einen GND-Eintrag angelegt hat. Weil, es gab den Mann schon und dann hat noch...

- 32 **Sascha Topp:** Das ist ja peinlich!
- 33 **Thorsten Halling:** ...und der dritte ist eine andere Person, der hat aber den gleichen Namen.
- 34 [Diskussion über die Person, Zwischenfragen von Elena Leitner]
- 35 Also auf jeden Fall, da sind genug fehlende!
- 36 **Dirk Wintergrün:** Das war auch nicht mein Argument, mein Argument war auch nicht: Alles Mögliche! Sondern selbst in diesem - also wir arbeiten ja auch mit diesem Beispiel, auch mit diese [unverständlich]. Weil wir haben eben geografische Daten, wir haben GND, wir haben Zitationsdaten etc. - die will ich gar nicht, die will ich erstmal gar nicht auch durch beliebige Archive ausweiten, ne. Um was es geht, ist ja erstmal die innere Konsistenz dieser Datensätze verstehen.
- 37 **Thorsten Halling:** Ja, genau!
- 38 **Dirk Wintergrün:** Und da glaube ich auch: Man kann einen Teil mathematisch vervollständigen mit hohen Wahrscheinlichen. Aber es gibt eben auch Teile, bei denen das nicht so geht. Und ich glaube das überzeugende Argument, gerade auch für die historische Arbeit ist, dass man beides machen kann. Auch gerade auf diesem Eingeschränken. Gar nicht sagen: "So, ich will jetzt noch mehr!", weil dann kommt man ja immer dahin, dass sowieso alles unsicher ist (schmunzelt). Weil, es gibt ja immer noch etwas. Also insofern stimme ich da - bin ich da völlig auf der Linie. Also man muss nicht jedes Mal alles haben.
- 39 **Sascha Topp:** War denn die Präsentation jetzt zuende? War das die letzte Folie?
- 40 **Marian Dörk:** Es gibt noch jetzt einen zweiten Teil.
- 41 **Thorsten Halling:** Ja, da war halt noch - also da war noch: Mit dem Zunst hätte ich noch eben gezeigt, dass es eben - das, was interessant ist, dass zum Beispiel gerade durch Kalliope, also das sind die Korrespondenzen, eine große Sammlung. Und da haben wir 20 Beziehungen, die nirgendwo anders drin sind. Die habe ich auf einmal dazu! Also die sind auch nicht im Wikipedia-Eintrag. Also ich wollte noch mal sagen: Das ist nicht nur eine Reduktion, sondern es ist gerade durch Kalliope ist es auch noch eine Bereicherung von Dingen, die - das sind halt Nebenbeziehungen - die haben sich halt mal geschrieben. Und wenn sie sich hin und her geschrieben haben, dann ist es natürlich schon enger und wenn sie sich mehrfach hin und her geschrieben haben, dann muss da irgendwas gewesen sein. Und dann: Häufig ist es eben dann - und das ist eben der Historiker, dass das dann auch eben zu Ideen [führt], die wir in den Visualisierungen haben. Uns würde es natürlich sehr viel Spaß machen, wenn das direkt auch einen gewissen Servicecharakter hätte. Also, wenn ich auf einen gewissen Knoten habe, dass ich dann eine Verlinkung habe zu der Quelle. Zum Beispiel zu Kalliope. Oder in der ZDB gibt es Informationen - zum Beispiel Volltexte digitalisiert. Und wenn man die natürlich dann direkt daraus aus dem Ding aufrufen könnte, also das wäre das wäre aus Nutzerperspektive auch ein echter Mehrwert. Nicht nur - also das...
- 42 **Marian Dörk:** Da werden ja schon Features hier gesammelt!
- 43 **Thorsten Halling:** Also weniger die Erkenntnis, als viel mehr der praktische Nutzen. Und eben auch etwas, was du eben auch schon gezeigt hast, Mark-Jan, also, wenn wir schon bei Features sind: Eben ganz viele Filter! Das ist, das wäre eben auch - dass man eben auch relativ mit einem Klick anzeigen könnte, welche Beziehungen gab es 1913, welche gab es in dem Zeitraum 1907-

1940. Dadurch könnte man eben auch dann relativ einfach herausfinden: Welche Beziehungen sind halt dauerhaft oder auch, man kann auch, wenn man solche Ereignisse wie Kongresse und Publikationen - dann hat man häufig prä- also Ereignisse, die lebenslange Beziehungen dann hervorrufen. Also wenn man so etwas dann auch visualisieren könnte, das wäre auch sehr sehr hilfreich für die historische Erkenntnis. Ich habe aber jetzt schon so viel gequatscht - entschuldige!

44 **Mark-Jan Bludau:** Ähm, ja, ich würde jetzt einfach noch mal ganz kurz die Daten beschreiben.

45 [Datenerklärungsteil für externe TeilnehmerInnen folgt]

46 Im Verlauf des Projektes sollen glaube ich aus den Volltexten dann auch Relationen oder zumindest Entitäten extrahiert werden. Ob man dann auch thematische Stichworte herauskriegt, weiß man noch gar nicht...

47 **Elena Leitner:** Meinst du die Relationen?

48 **Thorsten Halling:** Die semantische Suche! Eine semantische Suche.

49 **Mark-Jan Bludau:** Also, dass weiß ich nicht. Ja, zum Beispiel semantisch. Also: In welchem Kontext stehen die...

50 **Elena Leitner:** Ja. Nee, gut, wenn es um Relationen geht, dann naja - also: Ich würde das nicht als thematisch bezeichnen, weil es gibt dann ein Set von Relationen [und] das war's! Wenn [man] etwas thematisch extrahieren will, dann ist das etwas anderes!

51 **Thorsten Halling:** Naja, das ist schon, also um das noch mal - also der zweite Projektschritt ist eben ZEFYS dazuzunehmen. Das steht mit im Antrag drin, aber wir haben schon ganz schnell gelernt, dass das die Königsdisziplin ist. Wir müssen jetzt erstmal die anderen Dinge in den Griff kriegen und beschäftigen uns dann vermutlich in der zweiten Phase mit dem ZEFYS. Was für uns natürlich sehr schade ist. Weil inhaltliche Dinge, also semantische Felder, Evolutionen, Evolution von Wissen, von bestimmten Fragestellungen, kriegt man natürlich mit Volltexten - also, da ist natürlich mehr drin! Nur man, wenn die halt nicht, wenn die halt nicht getaggt sind, dann: Was macht man damit? Ich meine, es gibt ganze Editionen, irgendwelche Nachlässe, wo man, wo es 20 Jahre irgendwelche Nachlässe dann taggen. Um das automatisiert zu machen, wie sich das der Gehrhard [Müller, nicht anwesend] erträumt, das ist beim DFKI auf Stirnrunzeln gestoßen - oder wie war das? Habe ich das...oder habt ihr das anders interpretiert? Also da hat der Georg [Rehm, nicht anwesend] relativ reserviert drauf reagiert.

52 **Sascha Topp:** Wegen der Machbarkeitsfrage?

53 **Thorsten Halling:** Ja, weil es ist halt extrem schwierig, dem Computer oder dem Programm beizubringen, dass er einfach nur, dadurch, dass bestimmte Worte eben häufig mit einer Instution oder mit einer Person zusammen genannt werden - dass das eine Relation dann ist. Aber wenn du das dem nicht vorher sagst: "Hier, das ist jetzt eine Person und das ist irgendwie ein inhaltlicher Begriff, den wir suchen!", dann ist das - theoretisch glaube ich, geht das, aber praktisch ist das eben sehr aufwändig. Aber wie gesagt, wir haben uns glaube ich darauf geeinigt, dass das jetzt erstmal der zweite Schritt ist. Das ist ja auch schon durch die Farbgebung klar, was im Moment der Fokus ist, die dunklen [Bezug unklar] sind erstmal der erste Schritt.

54 **Vivien Petras:** Also momentan steht im Projektantrag, dass wir nur die Relationen identifizieren können, die in der GND sowieso schon definiert sind. Und das ist dann eine kleine, ganz klar

abgegrenzte Liste.

55 **Sascha Topp:** Mit Stichtag!

56 **Vivien Petras:** Ja, also Stichtag in sofern, als dass die GND - da haben wir einen Datendump genommen, ne. Das gesamte Dingens. Und pro Jahr kommen - naja, das kommt darauf an, manchmal hunderttausende von Datensätzen hinzu. Aber ja, mit Stichtag. Die Idee ist natürlich, dass das irgendwann dann automatisiert dann einfach immer die neueste Variante...

57 **Sascha Topp:** Automatisch eingearbeitet wird?! Ua!

58 **Thorsten Halling:** Es soll live sein! Also das ist das Ziel - weil anders wäre es ja unpraktisch, wenn du immer nur mit einer Version, die zwei Jahre alt ist, arbeiten könntest! Nee, das soll schon live sein!

59 **Vivien Petras:** Aber die GND-Relationen sind halt recht - also zwischen Ort und Person und auch zwischen Personen, aber die zwischen Personen die Beziehungen sind schon nicht mehr normiert.

60 **Mark-Jan Bludau:** Ja, da komme ich jetzt auch gleich zu - also ähm...

61 **Thorsten Halling:** Doch, doch, doch, die sind wohl normiert!

62 **Sina Menzel:** Aber nicht bis auf eine tiefere Ebene!

63 **Thorsten Halling:** Aber sie sind schon, also, ob es jetzt beruflich oder familiär ist, das ist schon...

64 **Sina Menzel:** Ja, diese Beschreibungen ja - aber es gibt zum Beispiel keine Norm "ist Sohn von" oder "ist Tochter von" oder "ist Elternteil von". Ja. Das wäre nicht normiert.

65 **Thorsten Halling:** Genau, das ist da zwar vermerkt...

66 **Sina Menzel:** Aber unnormiert vermerkt, genau.

67 **Elena Leitner:** Aber da gibt es ja doch eine Liste, oder? Von diesen Typen [der] Relationen?

68 **Vivien Petras:** Elena kann da vielleicht einiges reißen!

69 **Mark-Jan Bludau:** Ja! Das ist dann vermutlich die Frage, wie viel Prozent dann halt "Vater" anstatt "Beziehung familiär" genannt wird oder so.

70 [Dateneinführung geht weiter; Workshop-Teil-Beschreibung folgt]

Anhang B: 2_Audio_Kollagen_Spekulationen_I

- 1 **Thorsten Halling: [zur eigenen Kollage]** Also, es ist eigentlich so, wie ihr das interpretiert habt, bin ich eigentlich glücklich, dass ihr das so verstanden habt, wie ich das gemeint habe. Also ich habe versucht, die Ideen, die ich eigentlich auch gerade schon formuliert habe, alle komprimiert auf verschiedenen Ebenen hier unterzubringen. Also das ist einmal natürlich die Suche. Wonach suche ich? Kann das einschränken, ich kann dem - ich kann das Ding ändern, indem ich diesen Zeitstrahl eben verschiebe. Dann verändert sich mein Netzwerk in der Darstellung. Ich kann hier eben genau sehen, wie sind diese unterschiedlichen Verbindungen zeitlich, äh, sequenziert. Ich kann dann hier, wenn ich, wenn ich auf die einzelnen, äh, ähm, Nodes gehe, Knoten gehe - ob das jetzt Personen sind oder Institutionen, das ist irrelevant - dann kommen entweder Informationen, entweder Zusatzinformationen - wie biografische - oder es kommen eben - Hier, das wäre GND, hier das wäre ZDB, also, wo ist der Herausgeber. Und hier, das wären, das wäre jetzt ein Volltext, den es irgendwo gibt. Das wäre die Idee.
- 2 Und wie gesagt, hier - habe ich jetzt ganz falsch gemacht (Anm. Die Karte auf der Kollage ist falsch herum aufgeklebt), könnte man eben auch noch eine geografische Ansicht anfordern, wobei die auch noch differenziert ist, dass man auch sagen kann: B ist auch mal gewandert, auch mal nach England ausgewandert. Eben, für drei Jahre. Das macht ja etwas mit der Beziehung, wenn der auf einmal dann nicht mehr um die Ecke wohnt oder einfach - und und hier, das sieht ihr auch, dass, was wir auch im Antrag eben auch schon gesagt haben: Relevant - oder besonders, besonders interessant wird es für den Historiker, wenn er diese, diese Ansicht zitieren kann. Also, wenn er sagt: "Ich habe am 15. habe ich diese Anfrage gemacht." Die ist am, am 20. sieht die unter Umständen anders aus. Und, am 1. vielleicht noch mal anders, weil ein GND-Eintrag hinzugekommen ist. Das ist ja nur ein einziger Eintrag unter Umständen, hat auf einmal darauf Einfluss. Das sind die kleinsten Dinge, also ziemlich viel Input für so ein kleines Papierchen. (Lachen) Vielen Dank!
- 3 **Sascha Topp: [zur Kollage von Mark-Jan Bludau]** Ansonsten mag ich es, weil da über die Rezeptionsgeschichte - wenn man das überhaupt trennen kann - Konjunkturverläufe erkennen kann. Es gibt hier einen Bereich wo etwas stärker aufgegriffen wird, als vorher. Was weiß ich, so wie die Relativitätstheorie, die erst entwickelt, dann gar nicht rezipiert und plötzlich - Zack! Wiederentdeckt wird und dann geht es los, ne.
- 4 Aber diese kleine Unschärfe und dieser kleine Fehler ist total inspirierend, weil daraus andere Modelle möglich sind.
- 5 **Thorsten Halling: [zur Kollage von Susanne Helm]** Oder eben, was bei uns immer relevant ist, dass man unterschiedliche Forschungsdiskurse hat und die sind mit EINEM Akteur miteinander verbunden. Nur einen einzigen...
- 6 **Sascha Topp:** (bestätigend) Ja!
- 7 **Thorsten Halling:** ...Die haben sonst gar nichts miteinander zu tun und ein Einziger ist der Link. Und den muss man identifizieren, weil sonst versteht man überhaupt nicht, warum die doch dann irgendwie einander wahrnehmen.

Anhang B: 3_Audio_Kollagen_Spekulationen_II

- 1 **Sascha Topp:** [über seine eigene Kollage] Ja, da haben wir jetzt sozusagen aus der historischen Forschung ein Beispiel, das mich tatsächlich auch umtreibt. Ich dachte erst, ich nehme, das, das ich tatsächlich am Freitag auch in einer Konferenz vorstellen werde, nämlich die Geschichte des Milgram-Experiments über den Gehorsam, das Mitte der 1960er Jahre in Amerika durchgeführt wurde und dann über den ganzen Globus repliziert wurde. Also man kann dann eben vergleichen, wie die Studien in jedem Land unterschiedlich in den Kontexten [unverständlich], aber dafür hatte ich die Karte nicht. Ähm, das ist eine Konferenz über Crisis of Replication, die Psychologie diskutiert da gerade drüber.
- 2 Ich habe dann ein anderes Beispiel genommen und das ist während des Zweiten Weltkrieges eine Konferenz, die in der Schweiz stattfand - (Anm. An Thorsten Halling gewandt) Du kennst das unter "Max Müller" - über die sogenannten "Heroischen Theorien - äh, Therapien": Elektroschock, Energieschock und so weiter. [unverständlich]schock. Und da gab es eine Konferenz hier in der Schweiz und da waren ganz viele Leute dabei und die haben dann Berichte über diese Konferenz verschickt. So, die sind also nach der Konferenz selbst in ihre Länder zurückgekehrt und haben Berichte darüber geschrieben. Und die Zeitachse soll andeuten, mich interessiert ein bestimmter Rezeptionszeitraum, auf dem ich per Regler tatsächlich das verschieben kann und wo mir aufploppt, in welchem Land ist wann welche Konferenzrezension oder so etwas erschienen. Und dann sieht man auf dieser Karte möglicherweise: Gut, hier waren in der Schweiz zwei Japaner dabei, die berichten in Japan ganz begeistert über den Elektroschock und sich sich einig, tatsächlich, dass das eine fantastische Konferenz war. Äh, und dann, äh, findet aber -
- 3 **Thorsten Halling:** Das kennen manche von euch aus den Quellen - es gibt Berichte über Berichte. Und dann gibt es einen "Stille-Post-Effekt".
- 4 **Sascha Topp:** So. Und dann kann das also sein, dass über den Kontinent über Korrespondenten nach Amerika und nach Südamerika etwas -
- 5 [unklar]: (Einschub) Das sollte hier die Kontinente andeuten.
- 6 **Sascha Topp:** So, also da wir dann ein Bericht nach Amerika geschickt, der publiziert das in so einem amerikanischen Journal. Und das wird von einem Lateinamerikaner aufgegriffen, der überhaupt nicht mehr versteht, was da auf der Konferenz gewesen ist.
- 7 [unklar]: (Zwischenfrage) Und darum ist der unglücklich, der Amerikaner?
- 8 **Sascha Topp:** Weil der Amerikaner - und das ist das, was mich als Historiker interessiert - in, wie soll ich sagen, in dem Kontext, in dem er sich bewegt, sich fragt: "Was soll diese Konferenz überhaupt?" Und die Amerikaner nehmen das ganz anders auf und mich interessiert der Kontext, warum sie es anders aufnehmen. Ja? Und das gilt für jedes einzelne Land, wenn man dann diese Konstruktion *Nationale Geschichtssteuerung* nimmt, aber das hier oben soll tatsächlich andeuten, es gibt - [Ende der Aufnahme, Fortsetzung in Kollagen_Spekulationen_III_Audio]

Anhang B: 4_Audio_Kollagen_Spekulationen_III

1 **Fortsetzung**

2 **Sascha Topp: [über seine eigene Kollage]** Und ich war nicht davon ausgegangen, dass alle raten müssen, sondern dass ich das gleichzeitig beschreibe. Es war aber toll, dass ihr vorher raten musstet.

3 **Thorsten Halling: [zur Kollage von Sascha Topp]** Aber du hast von den - dich als einer der Wenigen völlig, völlig von den Daten gelöst.

4 **Sascha Topp:** Ja, das hätte jetzt auch eine Konferenz über die Psycholanalyse unter Beteiligung des...

5 **Thorsten Halling:** Nee, nee! Nee, nee, das ist ja, das ist einfach ein ganz anderer Ansatz, wir haben jetzt, wir haben alle eigentlich so überlegt: "So, was kann man mit den Daten machen?" Und du hast jetzt gesagt: "Was interessiert mich?" Und was man mit den Daten damit machen kann, das ist völlig irrelevant jetzt erstmal für dieses Konzept. Das finde ich, das hilft uns wahrscheinlich mehr, als wenn wir uns immer nur daran orientieren - weil, wir wissen ja, was in den Daten ist. Aber der User weiß das ja nicht.

6 **Marian Dörk:** (Einwurf) Naja, aber ihr wisst auch nicht alles, was da drin steckt, ne. Also es ist...

7 **Thorsten Halling:** Nein, nein, natürlich nicht, aber wir haben zumindestens - wir bilden uns ein, einen gewissen Überblick inzwischen zu haben, was da drin sein KÖNNTE an Informationen und [unverständlich] das wäre dann ein Handbuch, was man daneben legen müsste und...also die Idee, die Idee ist super!

8 **Sascha Topp: [über die Kollage von Marian Dörk]** Also meine visuelle Vorstellung ist wie in Gephy: Es geht hier von 1700 bis 2015, d.h. dass man über eine Play-Funktion dieses Gebilde von 1700 über 1800 über 1900 und nehmen wir an, dass wir hier das, weiß ich nicht - das Nobelpreiskomitee...

9 **Thorsten Halling:** (schmunzelnd) Ja, ja, ja!

10 **Sascha Topp:** Ihr macht das doch da in Düsseldorf! **Und dann, worum es eigentlich in all diesen Dingen geht: Wir suchen nach Mustern!** Und das würde sich hier tatsächlich zeigen. So wie dieses Gebilde sich nämlich über die Jahrhunderte verändert. Dann gibt es mal so Ausschläge [unverständlich] und die Ausschläge dienen als Anhaltspunkt: Was ist da los? Und das geht nur, wenn man das so fahren lässt oder zurückfahren lässt oder so. Weil das sind ja hier, du hast ja 3-4 Jahrhunderte...

11 **Thorsten Halling:** Da macht das mit der Farbgebung ja eben auch Sinn. Die sind dann eben gerade aktiv.

12 [Diskusison über die Kollage geht kollaborativ weiter]

13 **Thorsten Halling:** Man kann ja auch nach einem Thema suchen, man muss ja nicht zwingend nach einer Person suchen. Also wenn man z.B. "Nobelpreis" eingibt, dann würde das ja absolut Sinn machen.

14 **Dirk Wintergrün: [zur Visualisierung der Orte in der Kollage von Marian Dörk]** Also ich finde die Idee, dass man so etwas wie Lebensdauern für Orte hat, also wenn man historische Beispiele hat, gar nicht so blöd. Also es wird ja nicht sein, dass man jedes Mal - also gut, wenn die Orte halt die ganze Zeit spielen, dann hat man die halt die ganze Zeit laufen, aber das über die

Jahrhunderte die relevanter und weniger relevant werden und sich da ein gutes Maß definiert, ich finde das total einleuchtend! Also jedenfalls - kann ja sein, wenn es ein Beispiel ist, bei dem die Orte die ganze Zeit eine Rolle spielen, dann ist halt die Visualisierung vielleicht nicht so sinnvoll. Aber ich finde das gar nicht so fernliegend, dass die Orte sich über die Zeit verändern können, gerade bei so einer langen Perspektive. Im Gegenteil - ich finde das glaube ich sogar eine ganz gute Idee, dass viele Orte nicht Konstanten sind. Sondern sind sie ja häufig gar nicht. Sondern es wird suggeriert, dass sei eine Konstante. Das ist sie ja häufig gar nicht, weil der Ort einfach keine Rolle mehr spielt...

15 **Thorsten Halling:** Ja, genau das interessiert uns auch!

16 **Dirk Wintergrün:** ...also das finde ich gar nicht schlecht. Auch, wenn das vielleicht gar nicht so gedacht ist zu dem Zeitpunkt (lacht).

17 **Marian Dörk:** Also, aber das, das sozusagen diese Struktur, dass das per Play-Button sozusagen so ein bisschen flackert, finde ich auch sehr interessant! Dass man sagt: "Okay, ich wähle mir die Institution aus und sehe, welche Personen über die Jahre mit ihr in Verbindung standen, welche Orte, was auch immer. Ähm, das ist auch spannend, ja.

18 **Thorsten Halling:** Ja, oder eben die Institution auszuwählen oder den Ort auszuwählen und dann, also den Ankerpunkt kann man auch verändern. Also: Was ist der Ankerpunkt? Worum dreht sich mein Netzwerk?

19 **Dirk Wintergrün:** Das Gute wäre, wenn man die Unterschiede, die Veränderungen deutlich sehen würde, weniger die Start[unverständlich]. Wenn man das deutlich macht.

Anhang B: 5_Audio_Kollagen_Spekulationen_IV_Abschlussstatements

- 1 **Dirk Wintergrün: [über seine eigene Kollage]** Ja, ich wollte wirklich darauf hinaus, dass man, äh, relativ viel, ähm, irgendwie implizit mal wieder - wie du das auch gesagt hast (Anm. an Thorsten Halling gerichtet) implizit in diesen Daten findet. Und das, äh, und das irgendwie visualisieren muss. Also ich weiß eben nur, dass, ähm, aus meinem Netzwerk, dass in irgendeiner Quelle gesagt wird, dass Personen zum gleichen Zeitpunkt da sind. Kann ich daraus ableiten, dass die sich wirklich getroffen haben? Äh, kann ich davon ableiten - weil zwei über dasselbe Thema gesprochen haben - weil das ist mein Quellenhinweis oder aus dem Zeitungsartikel - dass die sich kennen oder nicht kennen? Und das Zweite, was [unverständlich] so eine Zeitleiste da unten
- 2 darzustellen. Nämlich dass Leute sich aus irgendwelchen Gründen treffen und dann tatsächlich entdecken, dass sie über gemeinsame Themen arbeiten. Die Frage ist: Kann man so etwas, also - ich würde so etwas total gerne aus den Daten visualisieren können! Muss man eben sehen: Okay, wo ist das Implizite und wo ist das Explizite?
- 3 Und das Zweite stimmt, das war eben die Idee, dass man am Ende sagen kann: Okay, man hat so etwas wie Steckbrief...zumindest irgendeine Form von linearer Übersicht. Weil das ja doch manchmal hilft zu verstehen, was man da aufgesammelt hat...
- 4 **Thorsten Halling:** (bestätigend) Ja!
- 5 **Dirk Wintergrün:** ...Ne, das ist eben ganz wichtig, dass man nicht immer nur, ähm, immer nur so einen Wust hat. Sondern auch auch [unverständlich] dass man sagt: "Das ist das, was ich feststelle!" Kann auch subjektiv sein. Kann auch so etwas wie eine Ergebnisliste sein. [unverständlich] einfrieren [unverständlich].
- 6 **Sacha Topp: [über die Kollage von Vivien Petras]** Also das kommt mir sehr bekannt vor aus den Visualisierungen, die Aaron zum Beispiel macht. Wenn er rauskriegen will: Wie gut ist die Datendecke zu bestimmten Kategorien? Dann visualisiert der das tatsächlich auch so. Dass man dann erkenn kann: Welche Kategorien sind nicht so gut abgedeckt oder wo sind Lücken? Und das berührt ja das, was Thorsten so ein bisschen auch eingangs thematisiert hat.
- 7 **Marian Dörk:** Ah, dann kann man das - das wäre so eine Art Metasicht, Datensicht auf die Personen, Zeitungen Entitäten! Über welche Jahre man viel weiß und wo man nicht so viel...ist die Frage, ob sozusagen man da wenig weiß oder ob da keine Zeitungen veröffentlicht wurden, ne? Das ist immer so - ne?
- 8 **Thorsten Halling:** Oder ob wir wenig in der Datenbank haben.
- 9 **Sascha Topp:** Darum geht es glaube ich! Daraus kannst du den Arbeitsaufwand ableiten, der noch ansteht, um die Datendecke so zu machen, dass du überhaupt etwa damit anfangen kannst.
- 10 **Marian Dörk:** Achso, einfach auch Metadatendichte oder...
- 11 **Thorsten Halling:** Hmhm (bestätigend). Ja und wieder, um so eine Wahrscheinlichkeit zu berechnen. Wie hoch ist die Wahrscheinlichkeit, dass das einfach nur eine dünne Datendecke ist? Es gibt da etwas zu, die haben nur nicht die nötigen Daten dazu. Oder das ist ganz viel Material zu haben, dann ist die Wahrscheinlichkeit, dass es die Verbindung vielleicht gar nicht gibt, ist höher. Dass man die gar nicht ableiten kann.
- 12 **Marian Dörk:** Ich sehe da irgendwie auch so drei Komplexitätsstufen. Also einerseits erstmal:

Hier sind die Entitäten von A-Z. Dann hängen die aber einzeln alle miteinander zusammen - also nicht alle vielleicht, aber es gibt viele mögliche Verbindungen. Und dann geht es aber in die Genauigkeit, ja, Datenlücken, Unsicherheiten. Das ist dann so, wo dann so: "Urgh!" Wo man dann so daran erinnert wird, dass diese Daten eigentlich alle irgendwie auch ihre Grenzen haben. Das ist so, also rein inelletkuell ist das so (schmunzelt) [unverständlich]. Und wie man das dann auch visalisiert, ne.

13 **Vivien Petras:** [über ihre eigene Kollage] Ich bin begeistert von euch! (allgemeines Gelächter)
Also ja! Ich bin nur noch nicht zur Visualisierung gekommen.

14 **Einwurf aus dem Plenum** "Du hast doch Farben genommen!"

15 **Vivien Petras:** Ja, Farben. Also es sind drei Ebenen. Und ich wollte erstmal anfangen, aufzudeklinieren. Und ich habe erstmal mit dem Linken angefangen (Anm. Gemeint ist die Entitäten-Übersicht). Mark-Jan hatte gefragt: "Welche Einstiegspunkte möchtest du in die Daten?" Und ich komme von den Daten her, also weiß ich, welche Einstiegspunkte da sind. Und dann ist das Zweite der - also im Kopf habe ich mir so überlegt, man klickt irgendwo links auf einen dieser Einstiegspunkte und dann kommt man im zweiten Filter dann auf jeweils eine dieser Ebenen. Ne, also wenn man auf "Ort" klickt, kommt man auf "Ort". Und dann habe ich mir überlegt: Von "Ort", zu welchen Entitäten kannst du dann die Beziehungen knüpfen? Und in meinem Kopf war ich immernoch erstmal angekommen wie: Dass das irgendwie getrennt sein muss und noch nicht zusammen.

16 Und dann bin ich abgelenkt worden durch die Frage "Und wie kann ich jetzt die Lücke in den Daten und die Datenqualität messen?" Und dachte: "Das ist eine ganz andere Ebene!" Und dann habe mich auf dem dritten Blatt ausgelassen und darüber nachgedacht, wie ich feststellen könne, ob da etwas fehlt, genau.

17 **Marian Dörk:** Ah! Das ist sogar - das sind sogar wirklich...das habe ich ja noch gar nicht gesehen, tatsächlich [unverständlich]...

18 **Vivien Petras:** Ja und also - wo ich noch nicht genau weiß ist wie ich auf das, wie ich von links ganz nach dem Roten komme. Ob ich das über die Gesamtdaten mache oder durch das mittlere Feld erstmal etwas eingrenze und dann schaue, ob da Lücken sind.

19 **Thorsten Halling:** Ja, ich glaube, aufgrund der Datenmenge MUSS man das Mittlere dawischennehmen. Weil da haben wir ja drüber nachgedacht: Weil es einfach so groß ist, kriegen wir die Datenlücken jetzt, wie wir es jetzt haben nicht in den Griff.

20 **Vivien Petras:** Also, ich dachte auch, dass das so das Explorative ist [Bezug unklar, vermutlich linker Rand der Kollage] und das [Bezug unklar, vermutlich rechter Rand der Kollage] ist dann das Validierende: Ist denn das, was ich sehe überhaupt abgesichtert durch die Daten? Aber jetzt mache ich das sehr viel intellektueller, als ich eigentlich angefangen habe.

21 [...]

22 **Dirk Wintergrün:** Das macht auch Sinn, das vorher, irgendwie vorher manchmal auch zu wissen in welche - also, ich frage mich auch ein bisschen, wo das hingehört. Denn es macht ja natürlich Sinn, ob man eine Frage überhaupt stellen kann.

23 **Thorsten Halling:** Jaja, das ist klar!

- 24 **Dirk Wintergrün:** Also wenn 1630-1650 keine Daten da sind, kann ich so viel fragen, wie ich will, was in der Periode passiert ist, zumindest an diese Daten. Aber so macht es auch Sinn, deswegen...
- 25 **Marian Dörk:** Ich finde das suggeriert, also das, äh, das lädt aber dazu ein, eine Art Datentransparenz irgendwie herzustellen. Also wenn ich eine Studie gemacht habe, dass ich sage: "Okay, aber pass auf: Ich habe das auf Daten gemacht, die so umfassend oder so dicht sind..." Und wenn man das Paper zehn Jahre später liest, denkt man vielleicht: "Ah, mit GND, mit diesen Milliarden von Einträgen!" - Nein, das war zu einem anderen Zeitpunkt.
- 26 **Thorsten Halling:** Ja, ja, genau. Deshalb haben wir ja auch gesagt: Wir müssen das immer, immer definieren. WANN ist diese Abfrage gemacht worden, um möglichst - im Idealfalle - die Gesamtzahl, dass er die mit rausspuckt. Wie viele Datensätze haben wir zu dem Zeitpunkt. Zumindest die GND-Datensätze.
- 27 **Marian Dörk:** Auch welchen Daten diese Darstellung zugrunde liegt.
- 28 **Thorsten Halling:** Ja, ja, genau.
- 29 **Dirk Wintergrün:** Ja, man braucht so eine Äquivalenz zu den systematischen Fehlern, die man in den Naturwissenschaften macht. Also, wie man das wirklich abbildet.
- 30 **Thorsten Halling:** Ja, ja!
- 31 [...]
- 32 [ABSCHLUSSRUNDE]
- 33 **Marian Dörk:** Ja, vielleicht klappt die Frage mit offenem Ende: Was uns jetzt hier besonders interessiert - ich meine, das ist ja schon selbst ein Kaleidoskop der Möglichen Visualisierungen und Analysen und Daten. Aber was uns jetzt natürlich interessiert ist: Was daran ist besonders vielversprechend auf der Forschendenseite? Was ist vielleicht auch interessant aus einer Gesaltungsseite? Also was bietet sich an, um vielleicht auch neue Formationen, neue Visualisierungen zu erfinden, sage ich mal. Was geben vielleicht auch die Daten her? Also was würdet ihr jetzt so mitnehmen von diesem Austausch, wo ihr sagt: "Das ist echt spannend!" oder "Schwierig!" oder spannend und schwierig zugleich...ähm, interessant.
- 34 [Diskussion über das Vorgehen]
- 35 **Marian Dörk:** Oder wenn ich jetzt mal so frage: Im Modus Hashtags, was fandet ihr jetzt so an Ideen, die aufgekommen sind und den damit verbundenen Fragen besonders vielversprechend?
- 36 **Sascha Topp:** **[Abschluss-Statement]** Also ich könnte mal mit ein paar Punkten anfangen, würde mir aber herausnehmen, stärker die Historikerperspektive da noch mal zu betonen. Was ihr glaube ich an allen Entwürfen seht ist, dass mitgedacht wird: Man braucht eine Timeline-Funktion. Ich weiß aus anderen Projektzusammenhängen aus dem Institut, dass Timeline-Programme inzwischen sehr gut handhabbar sind, aber die Datenpflege wahnsinnig viel Zeit kostet! Man muss einfach in der Analyse dieses ökonomischen Dinge mit bedenken, ne. Das wird euch aber auch beschäftigen. Aber ohne Timelines nützen mir die Visualisierungen nichts! Weder für die Analyse noch für die Vorstellung von Ergebnissen. Das heißt, für den Rechercheweg und die Analyse brauche ich schon die Visualisierung in der Form, dass ich vor und zurück fahren kann. Ja. Ohne das macht es kaum Sinn. Wenn ich ein Netzwerk über 100

Jahre anzeigen: Okay, da kann ich vielleicht irgend etwas erkennen, aber es ist vollkommen artifiziell! Ne.

37 **Thorsten Halling:** Ja, ja das ist ja das Problem, was eigentlich, was wir bis vor ein paar Jahren hatten. Die waren alle statisch.

38 **Sascha Topp:** Die waren statisch, genau.

39 **Thorsten Halling:** Die waren statisch. Und wenn eine Gute war, dann war es eben auf die paar Jahre begrenzt, dass man die anderen dann schon ausgerechnet hat, die Beziehungen. Und wenn es kein Gutes war, dann hat man alle Beziehungen, die man da findet, in eine Grafik gepackt.

40 **Sascha Topp: [Fortsetzung Abschluss-Statement]** Also die Flexibilität begrüße ich auch sehr, die Programme jetzt bieten. Weil ich nämlich dann anhalten kann, Screenshots machen kann und mir das näher anschauen kann. Also ohne das macht wirklich die Netzwerkanalyse kaum Sinn. Dann ist es etwas für die Soziologen, die ich sehr sehr schätze, aber Soziologen sind dann, was Ereignisse betrifft - oft schweben die so über den Dingen, weil sie andere Fragen stellen. Das ist total fruchtbar, mit ihnen zu sprechen, aber sie schweben irgendwie über den historischen Ereignissen und fragen auch nicht immer nach den Kontexten und so weiter. Und das ist das, was mich antreibt. [unverständlich] Soweit.

41 **Thorsten Halling: [Abschluss-Statement]** Ja, danke. Da kann ich mich nur anschließen. Ich fand einige waren ja schon sehr funktional, mein eigener [eingeschlossen], wo ich schon lange darüber nachgedacht habe. Deswegen fand ich die, die das so querdenken, waren für mich heute die interessanten. Also wie ich mir das vorstelle, das weiß ich. Und da waren einige noch Varianten, die fand ich auch total gut, aber zum Beispiel eben die - was du gemacht hast, Marian, das fand ich halt sehr spannend, weil das so eine so eine Mischung ist aus dem, was ich nichtig finde, nämlich die Visualisierungen, die nur schön sind. Schöne, viele Grafen, ja, schön, aber das könnte ich mir auch als Bild einfach an die Wand hängen. Aber das kombiniert das. Dass das auch eine Ästhetik hat aber dass das trotzdem relativ genau identifizierbar ist, was überhaupt für Knoten und Kanten überhaupt noch sind. Denn das scheint mir eine sehr schöne Lösung auch wenn man da ist. Auch wenn man da eben auch diese Timeline-Funktion hat und das ist schon eine sehr gute Lösung.

42 Ja, also ich finde das, um das nochmal zusammenzufassen, ich fand diese unterschiedlichen visuellen Lösungen fand ich halt sehr interessant. Und auch diese Idee, was du gemacht hast, Sascha, eben nicht von den Daten her zu denken. Das, ich glaube, das hilft uns immer mal, darüber nachzudenken, äh, wie der User da rangeht, der eben nicht zwei Jahre mit diesen Daten sich rumgeschlagen hat. Weil sonst haben wir nämlich das, weil sonst arbeiten wir nämlich an dem Nutzer vorbei. Das ist die Gefahr. Also dann ist das, was der Gerhard immer predigt: Veralgemeinern, veralgemeinern! Dann, den Arbeitsauftrag haben wir dann nicht absolut erfüllt. Deshalb fand ich das auch gut, eben weil es sich nicht damit [unverständlich], Sascha. Danke.

43 **Paula Schuster: [Abschluss-Statement]** Ähm, ich habe auch über die User-Perspektive nachgedacht. Und zwar würde mir das auch einfach in der Gestaltung total erleichtern, wenn ich einen konkreten Anwendungsfall hätte. Also: Wer sucht da jetzt wozu Daten? Gleich mit dem dem Zweck dahinter und dem Kontext: Wer ist das, warum tut die Person das, was will die da rausfinden? Damit ich das so gestalten kann. Weil wenn wir haben jetzt so viele verschiedene Ansätze, was da gesucht wird und ja, das hätte es mir glaube ich erleichtert. Und was ich sonst noch interessant fand: Wie manche visuell die Unsicherheit in den Daten dargestellt haben. Das fand ich ziemlich cool. Weil das ja bestimmt öfter vorkommt und das glaube ich auch noch echt

zu wenig gemacht wird in der Visualisierung, also so Unsicherheiten von Daten darzustellen. Das war's.

44 **Sina Menzel: [Abschluss-Statement]** Ja, ich kann es eigentlich kurz machen: Also, meine Hashtags, was in meinen Notizen auch immer wieder auftaucht sind zum Einen - genau - die Data-Insecurities, die Unsicherheiten, wo mir nochmal deutlich geworden ist, wie wichtig das für alle ist und wie präsent aber auch. Also, genau. Und das Andere ist, ähm, Exploration. Also möglichst viel Spielraum für die Daten, möglichst viel rumprobieren, verschiedene Perspektiven und so weiter. Zoom-in, Zoom-out, Makoperspektive, Mikoperspektive und so weiter. Ja. Das war's.

45 **Dirk Wintergrün: [Abschluss-Statement]** Also ich fand den Raum, wie ist das mit dem Umgehen mit Implizit, mit unsichereren Daten. Ich glaube, man muss auch dazu sagen: Uns ist das hier, also uns, die wir hier so reden, ist das irgendwie klar. Aber das ist nicht unbedingt den normalen Rezipienten von solchen Abbildungen klar. Also das ist dann, vielleicht auch das Spannendste, finde ich. Was wir hier mit den Daten, die das im Grunde schon länger machen und den Hintergrund haben tatsächlich auch visualisieren und deutlich machen. Dass das eben nur eine begrenzte Wahrheitsgehalt hat von dem was die Leute dann wirklich machen mit den Daten. Das muss man noch mal betonen.

46 Und was mir noch eingefallen ist - Ja, irgendwie veralgemeinern, aber umgekehrt finde ich es auch ganz wichtig, dass man das, was du gesagt hast (an Paula Schuster gewandt) - die User-Stories, die ihr braucht, um zu designen - aber auch umgekehrt hat. Dass die Leute, die auf so eine Seite kommen oder auf so etwas kommen, auf diese Daten, dass die eine Story haben, an der die sich festhalten können. Was die Leute mit den Daten gemacht haben. Man kann nicht alles, was in den Daten ist beschreiben, aber man kann ganz gut erfahren, was [unverständlich]. Also solche Stories - ja so ein Beispiel, wo du wirklich genau weißt, also, wie Saschas Story - Andeutungen. Also es gab auch so ein paar - die hatten eigentlich immer so ein Beispiel eigentlich erklärt. Also, ich fände es, also ich finde es glaube ich noch viel wichtiger, als so ein tolles User-Interface, wo jeder mal mit rumspielen kann. Dass klar ist, da sind 6-7 Historiker und die haben mit den Daten DAS gemacht. Das leuchtet mir in gewisser Weise auch ein. Also ich weiß nicht, wo man den Schwerpunkt setzt, aber irgendwie ist das mehr wert, als ein Handbuch, glaube ich.

47 **Thorsten Halling:** Ja, also wenn ich da noch kurz drauf antworten - also: Ja und nein. Ich glaube, wenn wir das zu sehr verengen, dann ist der Aufwand, den wir hier treiben für diese Historiker, ist nicht gerechtfertigt. Sondern ich finde schon, wir müssen eine Idee finden, wie wir mit Use-Cases demonstrieren können, was dieses Tool leisten kann.

48 **Dirk Wintergrün:** Genau, absolut. Also ich glaube so lernt, also das ist die Form, wie man kommuniziert, ne, also...

49 **Thorsten Halling:** Ja, und es ist unsere Aufgabe, völlig unterschiedliche Use-Cases durchzuspielen und die dann wieder zu dokumentierten mit Bildern, mit Suchanfragen, wie muss ich suchen, also wie muss ich die Anfrage stellen, damit ich DAS Ergebnis rauskriege, so als Beispiel. Und das zu unterschiedlichen Themen und Personen und auch den richtigen Zeiträumen.

50 **Dirk Wintergrün:** Hmhm, und das ist eine konkrete Designanfrage, weil man das dann festhalten müssen muss. Also (lacht) man muss da durchlaufen können, auch [unverständlich] und nicht nur einfach Screenshots machen, finde ich.

51 **Vivien Petras: [Abschluss-Statement]** Ja, alles, was ihr gesagt habt. Ich bin reingekommen in

den Workshop - und gehe aber mit der gleichen Frage wieder raus: Und das ist die Frage, ob man mit einigen wenigen, simplen, kleinen Use-Cases anfängt - also: Suche über diese Person und das jagen wir jetzt mal durch. Oder ob man gleich mit der eierlegenden Wollmilchsau anfängt. Und gleichzeitig auch schwanke ich immer noch zwischen der Frage, ob man von vorne herein erklärt, was die Daten können oder ob man die Leute erstmal darauf loslässt - mit der Chance, dass sie etwas ganz Neues entdecken, mit dem Risiko, dass sie komplett enttäuscht sind, weil das nichts drin ist. Das sind beides Fragen, die ich hier durch diesen Workshop im Prinzip bestätigt sehe, aber noch nicht beantwortet. Ich habe jetzt - ich würde dir genau widersprechen (an Thorsten Halling gewandt). Ich hätte gedacht, wir müssen die Daten erklären, wir müssen genaue Szenarien zeigen, wie man es macht. Aber dadurch reduziere ich den Gestaltungsspielraum für die Nutzer.

52 **Thorsten Halling:** Ja. Ich habe das ja mit dem Zunzt jetzt mal durchgespielt. Und ich glaube, wir müssen erst noch ein paar andere Sachen machen, aber ich glaube ein MINIMUM an Information kriegen wir für sehr sehr viele Anfragen. Nicht die ideal- die Idealnetzwerke. Aber ich glaube, wenn man sagt, deutscher Sprachraum ist der Schwerpunkt, also wenn man da nicht irgendwelche Australier sucht, dann wird man da etwas finden. Und wenn es nur die 6-7 Basics sind: Wo war der, wann war der wo und hat der irgendetwas publiziert? Und das, glaube ich, können wir versprechen für fast alle Anfragen. Also, ist auf jeden Fall mein erster Eindruck.

53 **Marian Dörk:** Ich würde auch sagen, das ist ja nicht Entweder-Oder. Also man kann ja sowohl kleine Perspektiven, kleinere Experimente machen und trotzdem größere Ansprüche verfolgen. Also wo mehrere Perspektiven zusammenkommen. Also, was so Einfachheit und Komplexität und was man vermitteln kann und was einem konkreten Use-Case folgt und eher open-ended ist. Also, aber ähm, ich glaube, da werden wir Einiges auch einfach noch gemeinsam im Projekt auch gestalten. Also unsere ersten, also wir haben ja das Glück heute Abend, heute Nachmittag auch Projektexterne mit dabei zu haben, aber wir sind uns ja auch gegenseitig Kundinnen und Kunden oder Abnehmer oder Rezipienten oder wie auch immer. Und, äh, das ist ja auch schon sehr wertvoll. Also, die wir jetzt nicht Medizingeschichte zum Beispiel auch auf dem Schirm haben, ne.

54 **Susanne Helm: [Abschluss-Statement]** Ja, also ich finde ich auch, dass [unverständlich] auf jeden Fall wichtig ist. Also, gerade auch die Frage nach dem geografischen Raum, dem würde ich auf jeden Fall auch noch mal transparent machen. Auf der anderen Seite bin ich natürlich - ich bin Sozialwissenschaftlerin, deswegen - was du vorhin meintest (Anm. an Dirk Wintergrün gewandt), sieht man wahrscheinlich auch an meiner Visualisierung. Die Frage, die ich mir jetzt gestellt habe ist: Ist dieses Tool - soll das ein interdisziplinäres Tool sein, also soll es eben auch für Sozialwissenschaftler interessant sein? Und dann fehlt für mich so ein bisschen die Netzwerkperspektive. Also, was kann ich - und auch die Diskussion, wo wir glaube ich auch so ein bisschen angerissen haben: Was ist für euch soziale Netzwerkanalyse? Was wollt ihr - was kann man rausfinden? Du (Anm. Bezug unklar) meintest auch, die Maße, die spielen jetzt überhaupt keine Rolle erstmal. Ist das fest in eurem Projekt? Kann man da vielleicht möglicherweise mit eurem Tool noch irgendetwas rechnen? Weil das ist auch etwas, was ich mir wünschen würde. Wenn ich quasi mir so ein Tool anschauen würde. Also ich würde quasi das Potenzial, dass gerade sozusagen die Netzwerkanalyse hat an Interdisziplinarität vielleicht noch ein bisschen mehr herausholen. Weil man damit glaube ich auch sehr viele User locken kann, die eben nicht aus der Geschichtswissenschaft kommen.

55 **Dirk Wintergrün:** Also das würde ich noch mal komplett getrennt - wäre mein Gefühl sozusagen - das würde ich noch mal komplett getrennt diskutieren. Genau mit unter der Frage mit der Mathematisierung. Welche genauen Aussagen kann ich mit den historischen Daten treffen und wie kann man das visualisieren, denn es macht wenig Sinn...

- 56 **Susanne Helm:** Weil es müsste ja keine - also, man müsste ja keine Forschungsfrage damit beantworten können, aber dass man vielleicht einfach Zentralitätsmaße vielleicht einfach mit reinnimmt.
- 57 **Dirk Wintergrün:** Also, wir wollen ja Forschungsfragen beantworten, historische Forschungsfragen beantworten, das scheint ja doch total sinnvoll zu sein. Es ist ja diese Idee, die wir - wir sind da gar nicht so weit weg. Wenn wir die Nähe [unverständlich]. Wenn wir wissen, wie wir die unscharfen Geraden irgendwie benennen können, dass wir sagen im Zeitraum 1637-1905 hat die Person X eine bestimmte Zentralität und dann können wir eben viel erreichen. Und letztendlich gehts uns ja darum, [unverständlich] übersetzt geht es uns darum: Wie können wir nachvollziehbare Fehlerbalken bauen für die soziale Netzwerkanalyse. [unverständlich] muss auch formulieren. Da bist du (Anm. an Susanne Helm gewandt) vielleicht viel näher an dem, was du dir denkst. Und irgendwo muss man systematische Fehler einfach Mal definieren.
- 58 **Susanne Helm:** Da finde ich einfach deinen Ansatz ganz gut, dass man da wirklich ehrlich mit ist.
- 59 **Elena Leitner:** **[Abschluss-Statement]** Je weiter wir diskutieren, desto weniger Ahnung habe ich, was wir im Endeffekt bekommen. Es gibt ja so viele Vorstellungen, deswegen, wie Vivien schon gesagt hat, ist es ja dringend nötig, so irgendwelche so Basic Use Cases zu haben. Um zum Beispiel ich [mir] das so vorstellen kann, was allgemein wichtig wäre nicht so spezifisch, ich meine so für Benutzer. Ich denke, da kommen ja nicht nur so Historiker da zu dieser Visualisierung, sondern vielleicht ja auch andere Wissenschaftler oder Studenten oder was weiß ich. Die irgendwelche Arbeiten schreiben oder was weiß ich - einfach etwas lesen und möchten sehen, wie zum Beispiel Freud mit anderen Personen verbunden ist oder Orten oder sonst etwas. Deswegen meiner Meinung nach ist es ja wichtig, so ein Basic Fall zu überlegen, wie so eine Suche aussehen kann. Und vielleicht - ich bin mir nicht sicher - ist es ja sehr aufwändig, so Suche, so zum Beispiel nach regulären Ausdrücken oder so etwas [zu] machen. Zum Beispiel ich will wissen, wie Goethe so mit anderen Personen verbunden ist und wie, welche Beziehungen diese zweite Person oder zweiter Kreis hat. So, welche Beziehungen hat dann so dritter Kreis oder sowas. Oder ich will zum Beispiel gucken, wer hat irgendwelche Rezensionen oder sowas - falls es solche Informationen in den Daten gibt - wer hat so irgendwelche Rezensionen zu einem Buch veröffentlicht oder sowas. Ich meine, kann man die Suche so machen, dass so der, so ein Benutzer definiert, was er will und dann bekommt man so eine Visualisierung oder es ist ja sehr aufwändig, wenn man das so in allen Bereichen oder unter allen Merkmalen so das macht. (lachen) Und wie gut dann im Endeffekt ist [die] Performanz.
- 60 **Mark-Jan Bludau:** Bei der Datenmenge wahrscheinlich schwierig, aber das ist jetzt ja für den Workshop erstmal - vielleicht geht es ja in zwei Jahren ganz schnell.
- 61 **Vivien Petras:** In zwei Jahren haben wir die Sache mit den Graphdatenbanken gelöst!
- 62 **Elena Leitner:** Gut. Nee, aber so rein theoretisch es ja nicht besonders effizient dann wird, oder? Wenn man...über alle Daten zu iterieren.
- 63 **Mark-Jan Bludau:** ...über alle Datenpunkte und so iteriert und noch über die Volltexte iterieren will (lacht).
- 64 **Marian Dörk:** Na gut, das - also...aber (Anm. an Elena Leitner gewandt) meintest du reguläre Ausdrücke im Sinne von - um komplexere Filter zu formulieren? Oder meintest du - aber nicht in der Sprache von RegExr?

- 65 **Elena Leitner:** Nee, nee, nee nee! Ich meinte so als Konzept nur.
- 66 **Marian Dörk:** Ich meine man kann schon dann schon...also, und nicht nur den Volltext sondern man kann ja auch nach Entitäten suchen, man könnte ja nach auch womöglich, also, ROLLEN von Entitäten oder von Relationstypen oder so...also das wäre jetzt keine Raketenwissenschaft.
- 67 **Elena Leitner:** Nee klar, aber niemand will dann so 30min warten, bis so ein Ergebnis kommt oder...
- 68 **Marian Dörk:** Nee, wir hoffen da auf die performante Datentechnologie vom DFKI.
- 69 **Elena Leitner:** Aha, okay! (lacht, allgemeines Gelächter)
- 70 **Mark-Jan Bludau: [Abschluss-Statement]** Also von meiner Seite aus war glaube ich die Frage nach dem Ausgangspunkt, also wir haben ja...vier - sind bei einer bestimmten Ansicht schon gelandet, wo ich mich dann frage: Wie kommt man dann tatsächlich zu dieser Selektion erstmal hin. Weshalb ich das Letzte von Vivien dann auch sehr interessant fand, wo man - dass man tatsächlich vielleicht erstmal wie so einen Fragebogen - Was interessiert dich eigentlich? - oder so. Das irgendwie eingrenzt. Aber da weiß ich von meiner Seite halt auch nicht genau, wie tatsächlich so Fragestellungen stattfinden. Ob das von einer Person ausgeht oder man tatsächlich eher große Komplexe sich angucken will oder so.
- 71 **Dirk Wintergrün:** Das kann ich gleich von außen sagen: Wenn ein Fragebogen kommt, dann...
- 72 **Mark-Jan Bludau:** Achso, das war jetzt mehr so...
- 73 **Dirk Wintergrün:** Ja, ja, aber auch, ich meine auch - ich meine [unverständlich]. Ich würde auch glaube ich auch gar nicht - also als Außensicht jetzt mal - ich würde auch glaube ich nicht so viel [unverständlich], sondern man denkt sich immer Fragestellungen dann aus, die niemand anders hat, wenn man über das Veralgemeinern nachdenkt. Aber man muss glaube ich auch am Ende - entweder wird man super generisch und dann interessiert es gar keinen mehr oder man ist so - man denkt über so viele kleine Schritte nach, dass man auch nicht mehr [unverständlich].
- 74 **Marian Dörk: [Abschluss-Statement]** Und das ist ja auch - auch was du gesagt hast, Paula - das ist ja auch das Privileg, eben, dass wir sozusagen die Nutzer*innen, die Expert*innen im Team haben sozusagen und dass wir sozusagen, dass ist auch eben würde ich sagen ein besonderes Privileg, in der frühen Konzeptentwicklung, dass man immer wieder iterativ miteinander bespricht [unverständlich], das man sozusagen über diese Qualitäten oder Kategorien oder Datenaspekte hinaus dann tatsächlich in diesen Modus kommt auch ein Szenario mal beschreiben zu können: Wie geht man vor, was ist der Einstiegspunkt, wie gehe ich fort, wie fahre ich fort im Interface oder so, was teile ich vielleicht anschließend? Also ich glaube, da werden wir noch viel Kommunikation auch brauchen um das so zu justieren, dass es für uns intern schon mal interessant ist und gewinnbringend oder erkenntnisreich. Und da wird schon etwas für die da draußen dann wiederverwendbar sein, denke ich. Ich fand total spannend - weil ich glaube, das kommt aus den Kollagen vielleicht nicht ganz so raus - aber, wir haben die ja erst interpretiert aber dann wurden die ja noch mal erklärt und das poppte ja eben auch auf wie so eine [...] Story für Besucher, dass wir im Moment eben noch viel über Tools nachdenken und so experimentelle Darstellungen, dass es aber perspektivisch das glaube ich schon hilfreich sein könnte, darüber nachzudenken, wie man das dann vermittelt. Also, dass man nicht einfach nur einen Link zu einem Prototypen hat, sondern womöglich auch eine Beschreibung. **Keine Visualisierung funktioniert ohne Text!** Und welche Rolle würde denn Text irgendwie spielen. Also ist das ein Text - ein veröffentlichtes Paper irgendwo oder ist das auch im Rahmen von so einem - fände ich interessant mitzudenken, weil sich Visualisierungen nicht so einfach selbst

vermitteln lassen. Also, dass das auch erklärt wird.

75 **Thorsten Halling:** Also bei den Visualisierungen bin ich bei dir, aber das Tool - die Datenbankabfrage. Wie muss ich da agieren [unverständlich] - irgendetwas, was ich bis dahin schon gemacht habe also da darf jetzt nicht, da darf jetzt nicht irgendwie etwas ganz Neues kommen! Es ist genauso wie, wie bei Windows.

76 **Marian Dörk:** (Ablehnender Laut)

77 **Thorsten Halling:** Ja, nee, das ist das einfach, wenn man von Windows auf irgendetwas anderes umsteigt, braucht man erstmal!

78 **Marian Dörk:** Ja.

79 **Thorsten Halling:** Und das ist nicht der Sinn, also eigentlich musst du die, die Gewohnheiten aufgreifen.

80 **Marian Dörk:** Genau. Wir können nicht bei Null anfangen. Aber - aufgreifen, aber auch andere, neue Praxen entwickeln, ne.

81 **Thorsten Halling:** Ja, aber dass es möglichst erklärungsfrei ist, zumindestens die Abfragen zu machen.

82 **Marian Dörk:** Nee, sorry, mir ging es jetzt nicht um die Erklärung des Interfaces im Sinne von einem Handbuch, sondern um die Erklärung der Erkenntnisse. Also wie vermittelst du das der Commu-

83 **Thorsten Halling:** Ja, okay. Ja, ja. Aber das macht doch unser Beispiel.

84 **Marian Dörk:** Genau. Aber auch das muss erklärt werden. Also ich meine jetzt der Fall mit dem Schweizer Kongress (Anm. gemeint ist das Anwendungsbeispiel als der Kollage von Sascha Topp, siehe Audio II, 1-2), der wird sich auch in einer komplexeren oder [unverständlich] Visualisierung nicht einfach erklären. Also der braucht - man muss da auch einfach an die Hand genommen werden. Aber das müssen wir noch entwickeln, welchen Raum das nachher dann einfach einnehmen wird.

85 Ich würde mal fragen - nee - vielleicht erstmal "Danke" sagen!

86 [Workshop Ende]

Anhang B: Visualisierungsworkshop_Notizen_Petas

SoNAR Workshop Explorative Perspektiven auf soziosemantische Netzwerke 16.9.2019

1. Einwilligungserklärung
2. Kurzeinführung in den Workshop Marian

Normalerweise werden Schauobjekte benutzt, heute nicht
3. Vorstellung der Teilnehmer
Marian Dörk, Mark-Jan Bludau, Elena Leitner, ... (Design, Urban Complexity Lab),
Vivien Petras, Dirk Wintergrün (MPWIG), Sina Menzel, ... (Designerin, Urban
Complexity Lab) Thorsten Halling (HHU), Sascha Top (MPWIG)
4. MD: Ablauf, keine Pausen geplant, aber zwischendurch kann man sich rausklinken
5. MJB: Kurzvortrag (Präsentation): Netzwerke visualisiert
Verschiedene Visualisierungen mit Netzwerken werden gezeigt (nature video, charting
culture)
6. Fingerübung:
MD: Details vergessen der vergangenen Visualisierungen, aber Fidelity und Spektrum
der Formen im Kopf behalten, auch Modellierungs- und Konzepte merken

Kleines SN mit 10 AkteurInnen, Aufgabe für 10 Minuten: Visualisierung des SN, um
etwas Interessantes zu zeigen

Zwischenfrage: ST: was passiert heute eigentlich noch, was wird erwartet?

MJB: Bezug zu der Wirklichkeit (den Fachwissenschaftlern herstellen), kritische
Punkte, die die Design-Gruppe nicht sehen würde

Ablauf: Fingerübung, Einführung in die Daten, Einführung in den Forschungsprozess
Thorsten

MD: ins Gespräch kommen mit Artefakten

MD: kurz Skizze vorstellen und sagen, welche Aspekte hervorgehoben wird
Jeder hat leicht andere Vorstellungen, viel Netzwerkzentralität, strukturelle Löcher,
Verbindungen über weak ties, typographische Verbindung, welche Verbindungen sind
vorhanden

Netzwerk hat einige komische Punkte: jemand mag jemanden nicht und der andere
kennt denjenigen gar nicht

7. TH: Historische soziale Netzwerkanalyse (Präsentation HHU): Kritik: ästhetische
Visualisierungen, aber kein Erkenntnisgewinn ohne Lesen der Erklärungen, das
benötigen sie nicht
warum also HNA: exploratives Erkennen, Fragen werden erst durch Visualisierung
klar oder ersichtlich, Sachen, die aus der biographischen Erklärung vorher nicht
erkennbar waren
traditionell: Stammbaum – hier fehlen Verbindungen untereinander (zwischen
Menschen untereinander, z.B. Menschen, die bei anderen berühmten Leuten
vorkommen würden)
Folie 6: was wir in der Visualisierung gern hätten, unterschiedliche Gewichtungen je
nach Forschungsfrage
Folie 8: Probebohrung, was können die Daten und was können sie nicht, Vergleich
zwischen Informationen in Wikipedia und GND – Visualisierung sollten auch
Fehlstellen benennen können (in der GND: welche Informationen könnten drin sein,
GND ist ja eigentlich kein biographisches Nachschlagewerk, sondern eigentlich nur für
Identifier, auch welche Informationen können auf keinen Fall sein)
ST: Kriterienset, wonach Ihr sucht, muss dann vorher festgelegt sein

TH: naja, es muss auch unabhängig von einer konkreten Forschungsfrage funktionieren können – Physiologen im 19. Jh. ist im Prinzip auch nur eine Fingerübung

Könnte man berechnen, ob es eine Relation gibt, weil es die andere Relation gibt?

DW: Zustimmung, Schwerpunkt muss das heuristische, explorative (Was ist überhaupt da) sein, das kann man gar nicht überbewerten.

Was man sucht, kommt oft erst, wenn man zum ersten Mal auf die Daten raufschaut

Für uns auch wichtig: missing information, data – die Aufmerksamkeit ist auch entscheidend

Andere pragmatische Sache ist: Fehler in den Daten, kann man diese identifizieren
TH: Elena sieht viele Probleme in den Ausgangsdaten

DW: manche Fehler sind auch systematisch, potenzielle Links ist interessant im Spielen mit den Daten, ganz oft schaut man, was fehlt einem eigentlich, Exploration zeigt, „das kann eigentlich gar nicht sein“

TH: jetzt haben wir eine bestimmte Art von Daten, normalerweise schauen wir uns alles an, was es gibt – also hier andere Vorzeichen, große Lücken sind vornherein angelegt; „Arbeiten mit der Lücke“ „Arbeiten mit bestimmten Art von Daten“ (GND, ZDB, Kalliope)

Bsp: Jörg Vögele kommt 3mal in der GND vor, einer ist jemand anderes

Gerade durch Kalliope haben wir 20 Beziehungen, die nirgendwo anders drin sind, nicht in Wikipedia – nicht nur eine Reduktion, sondern auch eine Bereicherung
Schön wäre Verlinkung zur Quelle der Referenz, z.B. aus Kalliope aufrufbar

MJB: Vorstellung der Daten (Präsentation SBB)

Beziehungen zwischen Normdaten, Beziehung durch Brief (Empfänger, Sender), Agent A ist am Datum am Ort B; Beziehungen, die möglich sind

TH: wir nutzen Daten, die zu einem anderen Zweck aggregiert wurden, es ist höchst wahrscheinlich, dass wir im Laufe des Projekts weitere Relationen hinzukommen

8. Arbeitsphase

MJB: was wären interessante Aspekte, was sind gute Einstiegspunkte für die Visualisierung, gibt es bestimmte Dimensionen / semantische Beziehungen, die besonders wichtig sind?

Wir gehen dies bastelnd an: A2 Blatt, Grafiken (Personen, Karte, Screenshot alte Zeitungen, GND Normdatensätze

MD: es geht hauptsächlich um Personen, deren Präsenz sich geographisch, in Zeitungen / ZS, Konferenzen, Institutionen, niederschlägt

DW: Welche Rolle spielen die Zeitschriften?

MJB: da könnten Entitäten drinstecken

TH: das Prinzip muss funktionieren, nicht der Inhalt muss immer da sein – andere Inhalte können hinzugefügt werden

[VP: Fragen, die durch Daten beantwortet werden könnten:

- welche Personen befanden sich zum Zeitpunkt x am Ort y?
- welche Personen arbeiteten zum Zeitpunkt x an Fach z?
- Welche Orte hatten zu Zeitpunkt x möglichst viele Wissenschaftler aus Fach z?
- Kannten sich Person A und B?
- Konnten sich Person A und B gekannt haben?
- Wo sind Person As Schüler hingekommen (Stammbaum)?
- Welche Personen werden in Ort y zum Zeitpunkt x besonders häufig benannt?

- Welche Beziehungen hatte Person A grundsätzlich zu anderen Personen?

VP Einstiegspunkte: Zeit & Ort, Person, Zeitschrift / Zeitung, Profession, Institution]

Alle schauen alle Entwürfe an, dann alle: was sieht man in der Collage

Datensicht muss zum Datenauswertungszeitpunkt für die Zitierung fixiert werden (welche Daten – wieviele – wurden wann ausgewertet)

9. Abschlussdiskussion

Was nehmen wir als Spannendes mit aus dieser Arbeitsphase

ST: man benötigt eine Timeline

TH: schließt sich an, die querdenkenden waren interessant, nicht von den Daten her denken

Designerin Paula: Nutzerperspektive, es wäre viel besser, wenn man eine Anwenderperspektive hätte – was möchte die Person finden, wie kann man visuell die Unsicherheit in den Daten darstellen

SM: Datenunsicherheiten, Exploration – möglichst viel Spielraum für die Daten

DW: Elefant im Raum – Umgang mit den unsicheren Daten, das ist nicht unbedingt den „normalen“ Rezipienten klar im Umgang mit solchen Daten; irgendwie verallgemeinern, wichtig sind aber auch die User Stories, man kann nicht alles mit den Daten machen, z.B. durch ein Beispiel zu erklären – besser als ein Handbuch durch Nutzerszenarien darstellen/demonstrieren, was das System kann

VP: einige kleine Nutzerszenarien oder eierlegende Wollmilchsau, Daten mehr oder weniger erklären

Andere Studentin im Lab: datenethische Frage ist wichtig, soll das Tool interdisziplinär sein, mir fehlt die Netzwerkperspektive, die Netzwerkmaße scheinen keine Rolle zu spielen, kann man mit dem Tool etwas berechnen für die soziale Netzwerkanalyse – kann man damit Nutzer locken, die nicht aus der Geschichte kommen

EL: je weiter wir diskutieren, desto weniger kann ich mir vorstellen, was am Ende rauskommt. Es ist dringend nötig, ein paar basic use cases zu haben, insbes. auch für WissenschaftlerInnen aus anderen Disziplinen; es ist schwierig, Suche mit regulären Ausdrücken zu machen, z.B. Beziehungen von Goethe und dann der zweite Kreis darum

MJB: Frage nach dem Ausgangspunkt, wie kommt man zu einer Selektion tatsächlich hin

DW: nehmt doch erstmal die Szenarios von HHU

MD: wir denken im Moment noch sehr über Tools nach und sollten aber mehr darüber nachdenken, wie man dies vermittelt – keine Visualisierung funktioniert ohne Text und wie sieht dies aus

[Anhang B: Visualisierungsworkshop_Notizen_Menzel](#)

Visualisierungsworkshop

- Visuelle Anteile am Projekt werden kreativ behandelt
- Brainstormingteil wird nicht aufgezeichnet
- Was sieht man nicht direkt?
- Vorstellung Sascha Topp:
 - o „Vermittlungsarbeit“
 - o Narrationsänderung, methodische Änderung auf dem Feld der Geschichtswissenschaft
 - Diese Änderung ist nicht nur positiv
 - Das Potenzial in den Daten wird oft nicht erkannt
 - Fähigkeiten der Informatik werden gebraucht, um das Potenzial auszuschöpfen
 - Methodischer Wandel vor allem von HHU-Fangerau-Lehrstuhl
- Datendimensionen in eigene Netzwerke unterbringen und verknüpfen
- Personen als Kanten, nicht als Knoten

Visualisierungsübung

- Diskussionspunkte finden aus der HNA der Vergleichswerte
- Beobachtung: Langes Überlegen der Historiker (Topp, Wintergrün, Halling), bis ganz konkret etwas aufs Papier gebracht wird. Wenig explorativ, sehr ergebnisorientiert in den Handlungen, die Vorarbeit erfolgt rein kognitiv. Visualisierungsnahe WiMi (Schuster) fängt direkt an, etwas zu Papier zu bringen und brainstormt dabei scheinbar, es wird durchgestrichen und neu gemacht...
- Fast alle Teilnehmenden überlegen noch während der Vorstellungsrunde der Ergebnisse der anderen permanent weiter
- Schuster wählt eine individuen-konzentrierte Herangehensweise, alle anderen eine personenübergreifende (z.B. generalistische Herangehensweise bei Petras)
- Halling sucht nach der „Spinne im Netz“, das Erkenntnisinteresse ist hier auf dem Drahtzieher im Netzwerk, was konkret an seiner Forschungsarbeit auch vorkommt und später auch in seiner Präsentation so noch einmal als besonders interessant dargestellt wird. „Hub“-Gedanke; Interessiert an Individuen, aber dennoch generalistischer Ansatz. Hallings Herangehensweise ist darüber hinaus diskursinteressiert („Wer spricht mit wem über was?“)
- Topp ist besonders interessiert an den *weak ties*, also den schwachen Verbindungen zwischen einzelnen stark vernetzten Gruppen. Hier seien für ihn neuralgische Kommunikationspunkte zwischen verschiedenen sozialen Parteien. Er sieht in den schwachen Verbindungen die Abgrenzungen zwischen Polen („Bälle im Gegensatz zu anderen Bällen“; „Intersektionelle Verbindungen“). Topp legt viel Wert auf Interaktion („Dann will ich das auch anfassen und drehen können!“). Visuelle Spielmöglichkeiten werden als wichtig eingestuft.

Präsentation von Thorsten Halling: Einführung in die Arbeit der HNA

- Mehrwert von HNA
 - o Erkenntnisgewinn muss da sein!
 - o Grafik muss mehr als Ästhetik sein!
 - o Erfassung von mehr Inhalten, als man aus dem Text auf Anhieb entnehmen kann (oder überhaupt entnehmen kann mit menschlicher Wahrnehmung aus Text)
 - o Exploration ist sehr wichtig!

- Die Frage soll durch die Visualisierung getriggert werden. Visualisierung als Pool neuer Fragestellungen und Perspektiven auf hist. Quellen.
- Neue Erkenntnisse durch Punkte, die sich zeigen.
- Beziehungsarten und -qualität/-grad soll möglichst hervorgehen
- Visualisierung der Fehlstellen sind wichtig! Also die Lücken im Netz und Unsicherheiten in den Daten. Wer fehlt, wo ist durch Quellenmangel keine klare Aussage zu treffen, wo muss gerade deswegen geforscht werden?
- Wichtig ist es, Grenzen der Daten aufzuzeigen, damit keine Fehlinterpretationen entstehen.
- GND ist nicht als lexikalisches Nachschlagewerk angelegt, sondern als Identifier-Datenbank. Halling sieht eine Schwierigkeit in der Differenz zwischen bibliothekarischem und geschichtswissenschaftlichem Vokabular („Zu einem ganz anderen Zweck aggregiert!“)
- Wichtig ist für Halling, auszudifferenzieren:
 - o Was ist potenziell in den Daten, aber nicht immer ausgefüllt?
 - o Was ist auf keinen Fall in den Daten, weil hierfür gar kein Feld/Erfassungsmotivation besteht?
- Halling sieht auch in der Errechnung von Wahrscheinlichkeiten in Netzwerken viel Potenzial, Dinge aufzuzeigen, die Geisteswissenschaftler so nicht auf Anhieb sehen. Statistik kann hier helfen, zumindest bedingte Aussagen über Verbindungen treffen zu können.
- Heuristisches Experimentieren mit Logischem. Teile, die nicht mathematisch zu vervollständigen sind anders ansehen. Möglichkeiten, die laut Halling „über das Gesamte hinausgehen!“
- Fehler und Datenlücken sind von vorneherein klar, da wir feste Daten haben. Halling ist stark dafür, diese klar zu ergründen. Typologisierung der Merkmale, Klassifizierung und Gewichtung.
- Links zu Volltexten und Normdaten sind Halling sehr wichtig, hier wäre genau die Arbeitersparnis zu sehen.
- Temporäre Filter findet Halling zentral, auch aus explorativen Gründen.
- Laut Halling zählt auch die Erprobung des Systems und die darauffolgende Anpassung ganz besonders, d.h. die Algorithmen der Forschungsinfrastruktur müssen immer wieder erweiterbar und anpassbar sein.

Brainstorming und Diskussion

- Halling möchte auf jeden Fall Speicheroptionen und die Möglichkeit, den Freeze von Ansichten zitierfähig aus dem System auszugeben.
- Zeitstrahl-Option ist für Halling von besonderer Wichtigkeit!
- Halling möchte analog zu Topp auch eine Filteroption von Forschungsakteuren, die nur über eine einzige Kante mit anderen verbunden ist (häufiges Forschungsinteresse).
- Halling: Unterschiedliche Verbindungen in Farben, je nach Art der Verbindung
- Topp: Hat jüngst eine Forschungsarbeit zu einer Konferenz gemacht, bei der er das soziale Zentrum der Konferenz für die Außenrezeption herausfinden wollte („Stille Post Effekt“)
- Topp: Animationsoption über gewissen Zeitraum wäre gewünscht. Wichtigkeit der Zeitleiste wird betont! („Ohne Timelines nützen mir die Visualisierungen nichts! Ohne das macht das kaum Sinn!“). Dabei sind auch konkrete Zeitpunkte wichtig mit konkretem, punktualem Kontext.

Abschlussrunde

- Halling: „Nicht von den Daten her denken!“; „Wir müssen für die Suche Gewohnheiten aufgreifen!“

- Wintergrün: Es ist nicht allen klar, dass Unsicherheiten in den Daten existieren. Um zu vermeiden, dass invalide Forschung entsteht, müssen die Unsicherheiten zuverlässig gekennzeichnet sein!
- Helm: Sozialwissenschaftliche Perspektive fehlt! SozialwissenschaftlerInnen sollten in Case Studies einbezogen werden.
- Helm: Mathematisierung der Daten wäre gut, also statistische Ausgaben.
- Dörk: „Keine Visualisierung funktioniert ohne Text!“

Anhang B: Mail_Auswertungsansätze_AP3_Doerk

Betreff:

Re: [sonar-idh] Co-Design Workshop am 16.09. um 14 Uhr

Von:

Marian Dörk <doerk@fh-potsdam.de>

Datum:

18.09.2019, 16:58

An:

sonar-idh@listserv.dfn.de

Liebe alle,

herzlichen Dank noch einmal an diejenigen von Euch, die für den Workshop am Montag bei uns in Potsdam zu Gast waren. Es hat uns Spaß gemacht, mit Euch gemeinsam über Visualisierung zur historischen Netzwerkanalyse zu 'brainstormen' und zu diskutieren. Gern mehr davon :)

Neben ersten konkreten Ideen und Ansätzen sind wir in den Gesprächen auf wichtige Themen und Fragestellungen gestoßen (bzw. haben diese noch einmal bekräftigt), wie zum Beispiel

- Relevanz der Zeitleiste,
- Potenzial dynamischer Ansichten,
- Perspektive der Nutzer*innen
- Unsicherheiten und Fehlstellen,
- Datenethik und Sichtbarmachung von Fehlern und
- Use-cases und Anwendungsfälle

Mark-Jan ist zwischendurch leider krank geworden; ich überlasse ihm aber die detaillierte Analyse und Zusammenfassung, wenn er zurück ist.

Her sind die Folien, die Mark-Jan für uns vorbereitet hatte:

[https://docs.google.com/presentation/d/12RGxQKppR_YtAyGnzMXMX92NYGrEqacnBrgRELW1iz4/edit?
usp=sharing](https://docs.google.com/presentation/d/12RGxQKppR_YtAyGnzMXMX92NYGrEqacnBrgRELW1iz4/edit?usp=sharing)

Ich hatte im anschließenden Gespräch Links zum DNB-Projekt versprochen:

- Prototyp: <https://dnbvis.fh-potsdam.de>
- Artikel: https://uclab.fh-potsdam.de/wp/wp-content/uploads/Dialog_mit_Bibliotheken_1-2018_Herseniet_al.pdf

Schöne Grüße

Marian

--

Prof. Dr. Marian Dörk
Information Visualization

Fachhochschule Potsdam
Institute for Urban Futures
Urban Complexity Lab

+49 331 580-2511
doerk@fh-potsdam.de
<https://uclab.fh-potsdam.de>
<https://mariandoerk.de>

>> On 9. Sep 2019, at 13:28, Mark-Jan Bludau <mark-jan.bludau@fh-potsdam.de> wrote:
>>
>> Liebe Kolleginnen und Kollegen,

> >
> > Wie bei der letzten Video-Konferenz besprochen werden wir einen Co-Design Workshop durchführen, hauptsächlich in Koordination mit der HHU und der HU. Insgesamt geht es bei dem Workshop darum sinnvolle Ansätze und Zugänge für die Visualisierung von Netzwerken zu finden unter der Einbeziehung von unterschiedlichen Perspektiven auf die Thematik. Wir haben uns bereits auf nächsten Montag (16.09) um 14 Uhr geeignet und zusätzliche Leute eingeladen. Der Workshop wird an der FH Potsdam stattfinden und soll bis ca. 17 Uhr gehen.
> >
> > Bisher bestätigte Teilnehmer*innen:
> >
> > -Mark-Jan (FHP)
> > -Marian (FHP)
> > -Thorsten (HHU)
> > -Sina (HU)
> > -Vivien (HU)
> > -Elena (DFKI)
> > -Sascha Topp (Max-Planck-Institut für Wissenschaftsgeschichte)
> > -Dirk Wintergrün (Max-Planck-Institut für Wissenschaftsgeschichte)
> > -Susanne Helm (FHP/FU)
> >
> > +möglicherweise 1-2 weitere von uns angefragte Teilnehmer*innen
> >
> > Sonst sind natürlich auch alle anderen SoNAR-Teammitglieder bei Interesse/Zeit eingeladen ebenfalls teilzunehmen! Sagt doch einfach kurz Bescheid, falls ihr Interesse habt auch noch dazu zu kommen.
> >
> > Viele Grüße
> > Mark-Jan
> >
> > ---
> > Mark-Jan Bludau, M.A.
> > Data Visualization
> > Research Associate - UCLAB
> >
> > Fachhochschule Potsdam – University of Applied Sciences
> > Institute for Urban Futures
> > Urban Complexity Lab (HG 111)
> > Kiepenheuerallee 5, D-14469 Potsdam
> >
> > +49 331 580-2517
> > mark-jan.bludau@fh-potsdam.de
> > <https://uclab.fh-potsdam.de>
> > <https://markjanbludau.de>
> >