

## **SoNAR (IDH): Datenschnittstellen für historische Netzwerkanalyse**

### **Bludau, Mark-Jan**

bludau@fh-potsdam.de  
Fachhochschule Potsdam

### **Dörk, Marian**

doerk@fh-potsdam.de  
Fachhochschule Potsdam

### **Fangerau, Heiner**

heiner.fangerau@hhu.de  
Heinrich-Heine-Universität Düsseldorf

### **Halling, Thorsten**

thorsten.halling@hhu.de  
Heinrich-Heine-Universität Düsseldorf

### **Leitner, Elena**

elena.leitner@dfki.de  
Deutsches Forschungszentrum für Künstliche Intelligenz  
(DFKI), Berlin

### **Menzel, Sina**

sina.menzel@hu-berlin.de  
Humboldt Universität Berlin

### **Müller, Gerhard**

gerhard.mueller@sbb.spk-berlin.de  
Staatsbibliothek Berlin

### **Petras, Vivien**

vivien.petras@ibi.hu-berlin.de  
Humboldt Universität Berlin

### **Rehm, Georg**

georg.rehm@dfki.de  
Deutsches Forschungszentrum für Künstliche Intelligenz  
(DFKI), Berlin

### **Neudecker, Clemens**

clemens.neudecker@europeana-newspapers.eu  
Staatsbibliothek Berlin

### **Zellhoefer, David**

david.zellhoefer@sbb.spk-berlin.de  
Staatsbibliothek Berlin

### **Moreno Schneider, Julian**

Julian.Moreno\_Schneider@dfki.de  
Deutsches Forschungszentrum für Künstliche Intelligenz  
(DFKI), Berlin

Die Beziehungen zwischen Menschen etwa in Familien, Organisationen oder Märkten bilden das Gewebe sozialer Ordnungen. Beziehungen konstituieren Möglichkeiten und Zwänge; sie beeinflussen den Zugang zu sozialem Kapital und damit Handlungs- und Wahloptionen (Lin 2001). Die Analyse dieser Beziehungen ist wesentlich für das Verstehen und Erklären von sozialen Phänomenen. Mit der Sozialen Netzwerkanalyse (SNA) entwickelte vor allem die sozialwissenschaftliche Forschung auf der Grundlage der Graphentheorie geeignete Methoden und anschlussfähige empirische Theorien zur Beschreibung und Erklärung dieser Strukturen (Jansen 2007). Die methodischen und theoretischen Ansätze der SNA etwa für die Untersuchung von sozialen Positionen (Kadushin 2012), finden zunehmend auch in Verbindung mit historischen Fragestellungen Anwendung (Bauerfeld/Clemens 2014, Düring et al. 2016). Forschung in diesem Bereich ist aber mit zwei grundlegenden Herausforderungen konfrontiert: Zum einen ist Erhebung und Aufbereitung von Daten für Analysen aus den dezentral, teilweise verstreut überlieferten Archiv- und Bibliotheksbeständen aufwendig. Zum anderen ist die Nutzung der einmal erhobenen Daten für neue Forschungsfragen oder auch nur die Überprüfung der Ergebnisse quantitativer historischer Analysen bisher vor allem von persönlichen Faktoren wie der Kenntnis über Datenbezug, -format und -auswahl sowie technische Verfahren abhängig.

Im Ergebnis der Digitalisierungsprojekte unserer Kultur- und Wissenschaftseinrichtungen stehen inzwischen signifikant große, vielfältig repräsentative Datenkorpora bereit. Durch stete Innovation und Standardisierung in der Aufbereitung digitaler Bestände – beispielhaft genannt seien Optical Character Recognition (OCR) und Named Entity Recognition (NER) – gewinnen diese Daten auch das Interesse einer noch jungen quantitativen Perspektive auf historische Phänomene. Doch trotz der erheblichen Potenziale beruhen bisherige Angebote in erster Linie auf den Logistik- und Nutzungskonzepten für analoge Bestände: So erfolgt die Datennutzung und -generierung über Kataloge, Discovery-Systeme oder digitale Sammlungen einzelner Einrichtungen für die ebenso konventionelle Beschäftigung mit Einzelobjekten. Einrichtungsübergreifende Aggregatoren wie die Deutsche Digitale Bibliothek (DDB) optimieren zwar den zeit- und ortsunabhängigen Zugang, aber die quantitative Verwertung der Daten bleibt hinter den Möglichkeiten zurück.

An dieser Stelle setzt das Projekt SoNAR (IDH), Interfaces to Data for Historical Social Network Analysis and Research, an. In diesem anwendungsbezogenen Forschungs- und Entwicklungsprojekt werden systematisch forschungsorientiert das Aufbereiten, Bereitstellen und Analysieren von Massendaten für den Aufbau einer Forschungstechnologie für die Historische Netzwerkanalyse (HNA), die als ein Zweig der SNA historische Fragestellungen untersucht, erprobt. Ausgangspunkte für das Datenmaterial sind:

- » Kalliope, KPE (Archiv- und archivähnliche Bestände wie Nachlässe und Autographen),
- » Zeitschriftendatenbank, ZDB (fortlaufende Sammelwerke wie Zeitungen und Zeitschriften),
- » Gemeinsame Normdatei, GND (Entitäten wie Personen, Körperschaften und Orte) sowie
- » exemplarische Brief- (Edition Berliner Intellektuelle) und Zeitungsvolltexte (ZEFYS).

Das entstandene und stetig expandierende referenzielle System dieser verteilten Datenangebote bietet der Wissenschaft die Chance, mit statistischen und visuellen Mitteln einen breiten, tiefen Einblick in Genese und Konstellation vergangener sozialer Beziehungen zu gewinnen. Einzelne wissenschaftliche Arbeiten zeigen sehr überzeugend, aber notgedrungen in reduzierter und abstrahierter Form den Wert quantitativer Methoden anhand von Korrespondenzen aus Archivbeständen, wie sie in KPE erfasst sind, und belegen das enorme Erkenntnispotenzial für die historische Forschung (z.B. Mücke und Schnalke 2009, Boschung et al. 2002, Dauser 2008, Fangerau 2010, 2013). Die Titeldaten der ZDB flankieren Aussagen über soziale Netze (KPE, GND) mit Aussagen über Produktions- und Distributionskonstellationen (z.B. Verlag, Herausgeber, Verbreitung, Sprache). Durch das Aufbereiten von Entitäten in Volltexten von Briefen oder Zeitungsartikeln ist es möglich, die formalisierten Aussagen der Metadaten von KPE und ZDB substanziell zu erweitern.

SoNAR (IDH) soll für den breiten, fächerübergreifenden Bedarf Einzellösungen durch ein standardisiertes Angebot ersetzen und so Hürden für die Arbeit mit Methoden der HNA signifikant reduzieren. Im Ergebnis dieses Vorhabens werden die Leistungsfähigkeit bestehender Frameworks und Werkzeuge in einer Prozesskette zur Datenaufbereitung und -bereitstellung sowie die Chancen neuer Visualisierungs- und Interfacekonzepte für eine Forschungsumgebung demonstriert. Mit einem Implementierungs- und Betriebskonzept werden geeignete Ansätze und Konditionen für Aufbau und Betrieb der Forschungstechnologie aufgezeigt. Dieses Vorhaben knüpft an Konzepte der Infometrie, vor allem der Biblio- und der Szientometrie an, wobei jedoch weniger Fragen nach Trends (Tunger 2009), Impact (Hirsch 2005), Wachstum oder Marktwert (Haustein und Tunger 2013, Umstätter 2004) im Vordergrund stehen, sondern z.B. Figurationen (Elias 1970) oder die räumliche und zeitliche Evolution von sozialen Beziehungen und Kontexten (z.B. Themen). Es wird dabei der Umstand berücksichtigt, dass

die Ausgangsdaten nicht für nur ein Forschungsthema erhoben sind, sondern vielfältig nutzbar gemacht werden können. Daher gilt es aber auch, belastbare Aussagen über den Umgang mit fehlenden oder fehlerhaften Daten zu treffen. Die Forschungsumgebung wird durch wissenschaftshistorische Fallstudien begleitet, die im Projekt zu abstrakteren Forschungsdesigns ausgearbeitet werden und so die Potenziale der Technologie für fachwissenschaftliche Fragestellungen demonstrieren.

Erstmals soll ein standardisiertes Instrumentarium zur Verfügung stehen, um mit großen aufbereiteten Datenmengen und einer Forschungsumgebung etwa komplexe, multimodale sozio-historische Kontexte zu untersuchen und Erkenntnisse nach wissenschaftlichen Kriterien in Forschungsprozesse zu integrieren.

Das Poster stellt Konzeption und die einzelnen Teilziele des Projekts vor.

## Bibliographie

**Auer, Sören** (2018): Towards an Open Research Knowledge Graph. Zenodo. (<http://doi.org/10.5281/zenodo.1157185>)

**Bauerfeld, Daniel / Clemens, Lukas** (2014): Gesellschaftliche Umbrüche und religiöse Netzwerke. In: Bauerfeld, Daniel/Clemens, Lukas (Hg.) (2014): Gesellschaftliche Umbrüche und religiöse Netzwerke. Analysen von der Antike bis zur Gegenwart. Bielefeld, 2014

**Boschung, Urs et al.** (Hg.)(2002): Repertorium zu Albrecht von Hallers Korrespondenz 1724-1777. Basel, 2002 (Studia Halleriana ; VII/1)

**Dauser, Regina** (2008): Informationskultur und Beziehungswissen. Das Korrespondenznetz Hans Fuggers. Tübingen.

**Düring, Marten et al.** (Hg.)(2016): Handbuch Historische Netzwerkforschung. Grundlagen und Anwendungen. Münster.

**Elias, Norbert** (1970): Was ist Soziologie? (Gesammelte Schriften in 19 Bänden, 5. Berlin. 2009)

**Fangerau, Heiner** (2010): Spinning the Scientific Web. Jacques Loeb (1859-1924) und sein Programm einer internationalen biomedizinischen Grundlagenforschung. Berlin.

**Fangerau, Heiner** (2013): Evolution of knowledge from a network perspective. Recognition as a selective factor in the history of science. In: Fangerau, Heiner et al. (Hg.): Classification and Evolution in Biology, Linguistics and the History of Science. Concepts, Methods, Visualization. Stuttgart, 11-32

**Haustein, Stefanie / Tunger, Dirk** (2013): Sziento- und bibliometrische Verfahren. In: Grundlagen der Praktischen Information und Dokumentation. Berlin, 479-492

**Hirsch, Jorge E.** (2005): An index to quantify an individuals scientific research output. In: Proceedings of the National Academy of Science of the United States of America. 102, 46. 16569-16572

**Isenberg, Petra et al.** (2008): Grounded Evaluation of information visualizations. In: ACM DL. BELIV '08 Proceedings of the 2008 Workshop on Beyond time and errors: novel evaluation methods for Information Visualization, 56–63.

**Jansen, Dorothea / Wald, Andreas** (2007): Netzwerktheorien. In: Benz, Arthur et al. (Hg.): Handbuch Governance. Theoretische Grundlagen und empirische Anwendungsfelder. Wiesbaden, 188-199

**Kadushin, Charles** (2012): Understanding Social Networks. Theories, Concepts, and Findings. Oxford.

**Kromrey, Helmut** (2002): Empirische Sozialforschung. Opladen, 2002

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

**Luhmann, Niklas** (1987): Soziale Systeme. Grundriß einer allgemeinen Theorie. Frankfurt am Main.

**Moretti, Franco** (2009): Abstrakte Kurven, Karten, Stammbäume. Abstrakte Modelle für die Literaturgeschichte. Frankfurt am Main.

**Munzer, Tamara** (2009): A Nested Model for Visualization Design and Validation. In: IEEE Transactions on Visualization and Computer Graphics (TVCG). 15, 6. 921-928

**Mücke, Marion / Schnalke, Thomas** (2009): Briefnetz Leopoldina. Die Korrespondenz der Deutschen Akademie der Naturforscher um 1750. Berlin, 2009

**Umstätter, Walter** (2004): Szientometrische Verfahren. In: Grundlagen der Information und Dokumentation. Berlin, 237-243

**Tunger, Dirk** (2009): Bibliometrische Verfahren und Methoden als Beitrag zu Trendbeobachtung und Trenderkennung in den Naturwissenschaften. Jülich.