SSOAR

Open Access Repository
www.ssoar.info

# Die Integration von Befragungs- und Prozessdaten einer Online-Kontaktbörse

Schmitz, Andreas; Klein, Doreen; Skopek, Jan; Schulz, Florian; Blossfeld, Hans-Peter

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

**Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:**
GESIS - Leibniz-Institut für Sozialwissenschaften

**Empfohlene Zitierung / Suggested Citation:**
Schmitz, A., Klein, D., Skopek, J., Schulz, F., & Blossfeld, H.-P. (2009). Die Integration von Befragungs- und Prozessdaten einer Online-Kontaktbörse. *Sozialwissenschaftlicher Fachinformationsdienst soFid*, Methoden und Instrumente der Sozialwissenschaften 2009/1, 31-44. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-205115

**Nutzungsbedingungen:**
*Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.*
*Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.*

**Terms of use:**
*This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.*
*By using this particular document, you accept the above-stated conditions of use.*

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

# Die Integration von Befragungs- und Prozessdaten einer Online-Kontaktbörse

*Andreas Schmitz, Doreen Klein, Jan Skopek, Florian Schulz & Hans-Peter Blossfeld*

## Zusammenfassung

*Online-Befragungen stellen mittlerweile ein Standardwerkzeug der empirischen Sozialforschung dar, dessen Potentiale und Grenzen seit längerem diskutiert werden. Im Kontext vieler Internetanwendungen fallen jedoch zusätzliche Daten an, deren Verwendung für sozialwissenschaftliche und methodische Fragestellungen noch keine vergleichbare Verbreitung gefunden hat. Es handelt sich hierbei um prozessproduzierte Daten, die durch (soziale) Handlungen von Nutzern erzeugt und vom jeweiligen Provider ohne einen ursprünglich wissenschaftlichen Verwendungszweck aufgezeichnet werden. Am Beispiel des Projektes „Prozesse der Partnerwahl auf Online-Kontaktbörsen“[1] wird argumentiert, dass Verhaltensspuren, die durch Such-, Interaktions- und Kommunikationshandlungen in Social-Web-Applikationen generiert und aufgezeichnet werden, gewinnbringend für die sozialwissenschaftliche Forschung genutzt werden können. Neben einem empirischen Anwendungsbeispiel des ergänzenden bzw. komplementären Gebrauchs nichtreaktiver Beobachtungsdaten und Befragungsdaten, werden methodische Möglichkeiten vorgestellt, mit denen prozessproduzierte Daten supplementär zur Aufwertung der Qualität von Befragungsdaten verwendet werden können.*

## 1 Einleitung

Die sozialwissenschaftliche Forschung bedient sich des Internets als Erhebungsmedium insbesondere in Form von E-Mail- und Web-Befragungen. Im Kontext vieler Internetanwendungen fallen jedoch zusätzliche Daten an, deren Verwendung für sozialwissenschaftliche und methodische Fragstellungen noch keine vergleichbare Verbreitung gefunden hat. Es handelt sich hierbei um prozessproduzierte Daten, die durch (soziale) Handlungen von Nutzern erzeugt und vom jeweiligen Provider ohne einen ursprünglich wissenschaftlichen Verwendungszweck aufgezeichnet werden. So können beispielsweise im Kontext von Applikationen die unter dem Namen Social Web bzw. Web 2.0 firmieren[2], recht einfach Daten aufgezeichnet werden, die Auskunft über Interaktionen, Relationen und soziale Netzwerke erteilen können (vgl. Stegbauer 2006). Die detaillierten Aufzeichnungen derartiger prozessproduzierter Daten, die durch den Gebrauch von sozialen Online-Applikationen möglich werden, enthalten damit oft reichhaltige Informationen über deren Nutzer.

---

1 Das DFG-geförderte Projekt „Prozesse der Partnerwahl bei Online-Kontaktbörsen“ wird unter Leitung von Prof. Dr. Hans-Peter Blossfeld am Lehrstuhl für Soziologie I der Otto-Friedrich-Universität Bamberg durchgeführt.

2 Richter und Koch (2007:5f.) fassen unter dem Begriff Web 2.0 die spezifische Kombination aus neuen Techniken, neuen sozialen Anwendungsmöglichkeiten, neuen Geschäftsmodellen und schließlich einer sozialen Bewegung zusammen, die sich in technischer Hinsicht über die Verfügbarkeit kostengünstiger, hoher Bandbreiten und aus der Perspektive der Anwender über die leichtere Bedienbarkeit und aus dem daraus resultierenden größeren Umfang an Inhalten beschreiben lässt. Schmidt (2006:2 zitiert nach Richter und Koch 2007:8) definiert Social Software, also aus dem Kontext des Web 2.0 entstehende Anwendungen, als „internetbasierte(n) Anwendungen, die Informations-, Identitäts- und Beziehungsmanagement in den (Teil-) Öffentlichkeiten hypertextueller und sozialer Netzwerke unterstützen“. Gemein ist den unterschiedlichen Formen sozialer Software, dass sie die Verlagerung sozialer Aktivitäten in das Internet unterstützen, gleichsam vorantreiben und sie der Erzeugung von Inhalten durch die Nutzer selbst bedürfen. Bei der spezifischen Form „Onlinedating“ der Social Software wird der Nutzer selbst mit seinen öffentlich gemachten Attributen zum zentralen Inhalt.

Die beiden prinzipiell unterscheidbaren Datenformen „Befragungsdaten“ und „prozessproduzierte Daten“ werden jedoch im Zusammenhang der Internetforschung zumeist nicht gemeinsam ausgewertet. Im Folgenden wird argumentiert, dass die Möglichkeiten der Datenauswertung entscheidend verbessert werden können, wenn beide Datenquellen miteinander kombiniert werden.

Der vorliegende Artikel diskutiert daher Möglichkeiten der Verknüpfung von Befragungsdaten und prozessproduzierten Daten, die auf einer Internetkontaktbörse erhoben wurden. In diesem Zusammenhang wird aufgezeigt, welche Vorzüge sich aus einem komplementären und supplementären Gebrauch dieser beiden Datenquellen in inhaltlicher als auch methodischer Sicht ergeben. Unter einer komplementären oder triangulierenden Verwendung verstehen wir dabei die wechselseitige Ergänzung von Daten aus verschiedenen Quellen in informationeller Hinsicht, unter einer supplementären Verwendung die Ergänzung in Hinblick auf die Datenqualität.

Im Folgenden werden zunächst Stärken und Schwächen von online erhobenen Befragungs- und Prozessdaten beschrieben, um dann zunächst die Möglichkeiten der komplementären Verknüpfung dieser beiden Datenquellen zu diskutieren und an einem empirischen Beispiel vorzustellen. Im Anschluss daran geben wir einen Überblick über Verfahren mit denen Nonresponse als klassisches Problem der Umfrageforschung durch den supplementären Einsatz von prozessproduzierten Daten korrigiert werden kann.

# 2 Datenstruktur des Projektes „Prozesse der Partnerwahl von Online-Kontaktbörsen“

## 2.1 Befragungsdaten

Im Unterschied zu klassischen Datenerhebungstechniken wie schriftlichen, persönlichen oder telefonischen Befragungen stieg seit Ende der neunziger Jahre die Verbreitung von Online-Befragungen stetig an. Mittlerweile gehören Online-Befragungen zum Standardwerkzeug der empirischen Sozialforschung, dessen Potentiale und Grenzen seit längerem diskutiert werden (vgl. z.B. Couper/Coutts 2006, Welker/Wenzel 2007 oder Baur/Florian 2008).

Zu den Vorteilen von Online-Befragungen gegenüber klassischen Befragungen in Offline-Kontexten zählen die Aspekte der relativ schnellen, flexiblen und preisgünstigen Umsetzung von Forschungsvorhaben, des Wegfalls von Beeinflussungseffekten durch den Interviewer und der damit verbundenen Abmilderung sozialer Erwünschtheitseffekte (vgl. Baur/Florian 2008) sowie dem sofortigen Vorliegen der Daten in digitaler Form. Weitere Vorteile finden sich in der zeitunabhängigen Alokalität und den größeren Standardisierungsmöglichkeiten dieses Befragungsverfahrens (vgl. Bosnjak 2003). Zudem besteht bei Web-Befragungen die Möglichkeit der exakten Aufzeichnung des Befragtenverhaltens, also etwa der Anzahl von Kontaktierungsversuchen, der Befragungsdauer oder des Abbruchverhaltens. Diese so genannten Paradaten (vgl. Kaczmirek/Neubarth 2007), die den Erhebungsprozess dokumentieren, finden beispielsweise Verwendung in Nonresponse-Analysen, in Analysen von Befragungsartefakten sowie in Analysen der Panelattrition (vgl. de Keulenaer 2005).

Demgegenüber stehen die Nachteile von Online-Befragungen. Das augenfälligste Problem stellt das der Generalisierbarkeit der auf Basis der erhobenen Daten gewonnen Schlüsse dar. So sind

Befragungsdaten durch Antwortverweigerungen seitens der Befragten oft nicht vollständig.[3] Zudem bilden sie (messfehlerbehaftete) subjektive Präferenzen und Einstellungen nicht aber objektive Handlungen oder Handlungsresultate ab. Da sich die untersuchten Akteure kaum sämtliche überindividuellen Prozesse vergegenwärtigen können, stellt sich beispielsweise die Frage, inwieweit mit der Datenerhebungsmethode der Befragung, die im Wesentlichen auf bewusste Handlungen fokussiert, beispielsweise unbeabsichtigte Nebenfolgen menschlicher Handlungen eine adäquate methodische Erfassung erhalten können (vgl. Baur 2008:4). Das Problem der Generalisierbarkeit korrespondiert weiterhin mit den Schwierigkeiten einer zufälligen Stichprobenziehung, die die Grundlage inferenzstatistischer Verallgemeinerungen darstellt. Einer Stichprobenziehung geht die Definition der Grundgesamtheit voraus, die bei Erhebungen mithilfe von Web-Befragungen in aller Regel schwierig ist. Aber selbst bei einer klar definierten Grundgesamtheit können durch systematischen Nonresponse-Selection-Bias Probleme der Datenqualität auftreten, die Analysen und Schlüsse zum Teil erheblich in Frage stellen.

## 2.2 Prozessproduzierte Daten

Unter prozessproduzierten Daten werden ganz allgemein Daten verstanden, die „als Aufzeichnungen öffentlicher oder privater Organisationen im Rahmen ihrer Tätigkeit anfallen oder gesammelt werden“ (Müller 1977: 1). Beispiele hierfür sind die IAB Beschäftigtenstichprobe (Haman et al. 2004), Analysen von Rechtsakten, administrativen Erhebungen (Zanutto/Zaslavsky 2002) oder auch Analysen von Forschungsföderungen der Deutschen Forschungsgemeinschaft (vgl. Bovelet 2006). Bergmann/Meier (2000:429) definieren nun *elektronische* Prozessdaten als „alle Daten, die im Verlauf von computergestützten Kommunikationsprozessen und Arbeitstätigkeiten generiert werden - entweder automatisch oder aufgrund von Einstellungen der Nutzer.“ Auch diese elektronischen Prozessdaten stellen im wesentlichen Nebenprodukte der Datenorganisation des jeweiligen Anbieters dar.

Im Unterschied zu offline gewonnenen Prozessdaten, deren Aufbereitung sich oft über einen erheblichen Zeitraum erstrecken kann (vgl. Hartmann/Krug 2007), können online gewonnene Prozessdaten zeitnah erfasst und verarbeitet werden.

Nicht nur die den Befragungsprozess beschreibenden Paradaten können aufgezeichnet werden, sondern online vorgenommene Handlungen in Form von „Klicks“ überhaupt, da bei fast allen Online-Aktivitäten Daten anfallen, die für ganz unterschiedliche Forschungszwecke dienlich sein können. Bei online realisierter Kommunikation können somit insbesondere Kommunikations*beziehungen* automatisch aufgezeichnet werden. So untersuchten Stegbauer und Rausch (1999) Beziehungsstrukturen mit Hilfe von archivierten Nachrichten einer Mailingliste. Berker (1999: 234) bezeichnet die Erhebung dieser Daten als „geplante Beobachtung“ und hebt die Objektivität und Ausschaltung der Messfehler als besondere Vorzüge hervor. Im Gegensatz zu Offline-Beobachtungen sind elektronische Prozessdaten viel stärker an die „physische Spur“ ihrer Verursacher gebunden (Berker 1999: 234). Elektronische Prozessdaten können deshalb als *nichtreaktive, objektive* Verhaltensdaten beschrieben werden, die zudem die Temporalstruktur der ihnen zugrunde liegenden Handlungen abzubilden im Stande sind (vgl. Bergmann/Meier 2000).

Elektronische Prozessdaten sind im Vergleich zu Befragungsdaten in geringerem Umfang von Problemen der selektiven Teilnahme, sozial erwünschtem Antwortverhalten, Antwortverweige-

3 Typischerweise unterscheidet man zwischen *Unit-Nonresponse* und *Item-Nonresponse*. Während Unit-Nonresponse den vollständigen Ausfall des Befragten durch Verweigerung der Befragungsteilnahme beschreibt, bedeutet Item-Nonresponse lediglich das Fehlen von Antworten auf spezifische Fragen.

rungen oder inkorrekten Angaben betroffen. Um beispielsweise Rückschlüsse bezüglich der Einwirkungen eines Marktes auf die Handlungen der in ihm agierenden Akteure ziehen zu können (etwa durch differentielle Auswirkungen von Gelegenheitsstrukturen), bedarf es der kompletten Abbildung des Marktes (etwa der Verteilung relevanter Merkmale). Hierbei können die elektronischen Prozessdaten in ihrer Vollständigkeit von besonderem Wert sein.

Prozessproduzierte Daten entstammen allerdings nicht-operationalisierten Konstrukten und decken damit nicht all jene Bereiche ab, die von inhaltlichem Interesse sind. Insbesondere die mit den jeweiligen Handlungen verbundenen Wahrnehmungen und Bewertungen seitens der Akteure, aber auch Dispositionen im Allgemeinen schlagen sich in der strukturalistischen Abbildung prozessgenerierter Interaktionen nicht nieder. Weiterhin benennt Schnell (2005:255) als Probleme offline-gewonnener prozessproduzierter Daten mögliche Verzerrungen, die durch Behördenmitarbeiter und deren implizite oder explizite Absichten hervorgerufen werden. Er empfiehlt eine Offenlegung des „Entstehungsprozesses der Daten“ (ebenda), was im Zusammenhang mit prozessproduzierten Daten einer Web-Applikation problemlos gewährleistet werden kann, da bei der jeweiligen Aufzeichnung einer Beobachtung die Ermessensspielräume verschiedener Personen nicht einfließen.

## 3 Design des Projektes „Prozesse der Partnerwahl auf Online-Kontaktbörsen“

Die im Rahmen des Projektes erhobenen Daten entstehen aus der Kooperation mit einem großen privatwirtschaftlichen Online-Dating-Anbieter und setzen sich zum einen aus der Aufzeichnung nichtreaktiver plattforminterner Vorgänge, der sog. Verhaltensspuren der Nutzer, und zum anderen aus reaktiven Befragungsdaten unterschiedlicher Online-Fragebögen zusammen. Die prozessproduzierten Datenbankauszüge werden seitens des Anbieters in anonymisierter Form als SQL-Dateien für die vorgesehenen Analysezwecke unter Aufsicht eines Datenschutzbeauftragten des Providers zugänglich gemacht.[4] Die Erhebung der Befragungsdaten erfolgt mittels verschiedener Online-Fragebögen. Die Nutzer der Internetkontaktbörse werden über eine E-Mail, die einen Link zum jeweiligen Fragebogen enthält, über das plattforminterne Nachrichtensystem sowie über Befragungslinks auf der Plattform zur Teilnahme an den Befragungen eingeladen. Die Fragebogeneingaben werden dann in einer projekteigenen Datenbank ohne Zugriffsmöglichkeit des Anbieters gespeichert.

4 Die Daten werden dabei anonymisiert weitergegeben, d.h. sie enthalten keine persönlichen Informationen über den Nutzer wie beispielsweise den realen Vor- und Nachnamen, die E-Mail- oder Postadresse. Weiterhin werden die Nutzer der Internetkontaktbörse in den Allgemeinen Geschäftbedingungen darüber informiert, dass Profildaten sowie plattforminterne Vorgänge im Zuge wissenschaftlicher Analysen zur Partnersuche und Partnerwahl ausgewertet werden.

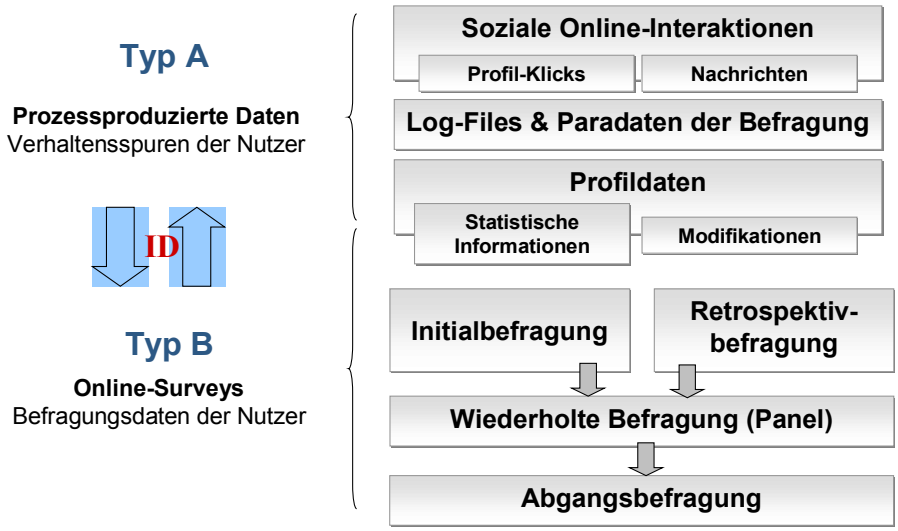


*Abbildung 1:* Datenstruktur des Projektes "Prozesse der Partnerwahl auf Online-Kontaktbörsen"

Abbildung 1 veranschaulicht die resultierende Datenstruktur. Jeder Nutzer kann über eine Identifikationsnummer eindeutig den prozessproduzierten Daten (Typ A) und den Befragungsdaten verschiedener Befragungen (Typ B) zugeordnet werden. Die Profildaten nehmen dabei einen Zwischenstatus ein.[5] Im Ergebnis entsteht eine relationale Datenstruktur aus der je nach Fragestellung unterschiedliche Datensätze, wie beispielsweise aggregierte Querschnittsdaten oder Längsschnittsdaten, generiert und in einem herkömmlichen Statistikpaket ausgewertet werden können.

Befragungsdaten werden im Rahmen des Projektes über verschiedene Online-Fragebögen gewonnen, die *allen* Nutzern der Internetkontaktbörse jeweils zu unterschiedlichen Zeitpunkten über eine E-Maileinladung und in Form von personalisierten Links zugänglich gemacht werden. Da somit keine Stichprobenziehung aus der Grundgesamtheit der Nutzer dieser Internetkontaktbörse vorgenommen wird, handelt es sich um den Versuch einer Vollerhebung mit dem Resultat einer selektiven Stichprobe. Mit Eintritt in die Dating-Plattform werden zunächst die für die Nutzung des Web-Angebotes notwendigen Profilinformationen wie Geschlecht, Geburtsdatum, Bildungsabschluss, Kinderzahl, Art der gewünschten Beziehung, Angaben zu Gewicht und Körpergröße sowie Haar- und Augenfarbe, Rauchgewohnheiten und Postleitzahl abgefragt. Über diese Profilinformationen stellen sich die Nutzer gegenüber anderen Nutzer auf der Internetkontaktbörse dar.

Mit der Registrierung auf der Kontaktbörse erhalten die Nutzer zunächst eine Einladung zur so genannten *Initialbefragung*. Hier werden Bewertungen, subjektive Wahrnehmungen und Einstellungen zur Partnersuche im Allgemeinen sowie speziell zur Partnersuche im Internet abgefragt. Ausgewählte Aspekte des Initialfragebogens werden im 14-tägigen Rhythmus bei denselben Nutzern

5 Die Profilinformationen sind einerseits den Befragungsdaten zuzuordnen, da sich die Profilmaske den Nutzern der Kontaktbörse als Profil*fragebogen* darstellt. Andererseits können diese Profilinformationen den prozessproduzierten Daten zugeordnet werden, da diese Daten ohne ein ursprünglich wissenschaftliches Interesse im Rahmen der Tätigkeit der Kontaktbörse selbst aufgezeichnet werden.

wiederholt abgefragt *(Panelbefragung)*. Im Unterschied zu anderen Panels fallen Non-Respondenten einer Welle dabei nicht aus der Erhebung, sondern werden auch bei Nicht-Teilnahme für darauf folgende Befragungswellen kontaktiert (vgl. Lepowski/Couper 2002).[6] Weiterhin werden die Partnerschafts-, Erwerbs- und Bildungsbiographie der Nutzer mittels einer *Retrospektivbefragung* erhoben. Bei längerer Inaktivität auf der Internetkontaktbörse oder bei endgültiger Abmeldung von der Plattform in Form der Löschung des eigenen Profils erhalten die Nutzer weiterhin eine *Abgangsbefragung*. Hier werden Gründe der Abmeldung und mögliche Erfolge der Partnersuche erfragt.

Zusammenfassend sind die Befragungen des Projektes in inhaltlicher Hinsicht motiviert als Abbildung individueller Lebensverläufe und aktueller Wahrnehmungen in Bezug auf Partnerschaften und weitere Dispositionen; dabei sind auch externe Ereignisse wie beispielsweise das Eingehen von Offline-Partnerschaften und Gelegenheitsstrukturen der Partnersuche außerhalb des Internets relevant.

Die kontinuierlich archivierten Prozessdaten beinhalten demgegenüber die Paradaten der verschiedenen Befragungen ebenso wie die Clickstreams der Nutzer. Insbesondere lassen sich die beschrittenen „Wege“ auf der Plattform, also die Typen und Reihenfolge der unterschiedlichen Seiten des Webangebotes, Veränderungen der Profilinformationen, Interaktionen bzw. Kommunikationen mit anderen Nutzern und Suchhandlungen aufzeichnen.

Die Möglichkeiten der Datenanalyse können nun entscheidend verbessert werden, wenn beide Datenquellen miteinander verknüpft werden. Ähnlich der Datenlage von offline-gewonnenen Prozessdaten und Befragungsdaten ist eine Kombination der beiden Datenquellen vielversprechend, da somit jeder Nutzer über reichhaltigere Informationen beschrieben werden kann. „Durch die Verknüpfung von prozessproduzierten Daten mit speziell für die Forschungsfrage erhobenen Befragungsdaten kann ein Maximum an Aussagekraft und Validität der Ergebnisse erreicht werden“ (Hartmann/Krug 2007: 10). Ebenso sehen Schnell et al. (2005: 418f) die Fruchtbarkeit von nichtreaktiven Beobachtungsdaten insbesondere im triangulativen Ansatz gegeben, also der komplementären Verwendung von Beobachtungsdaten und beispielsweise Befragungsdaten (vgl. auch Flick 2004) oder in der Verknüpfung mit anderen Datenquellen.

Bei der für die gemeinsame Verwendung notwendigen Datenintegration unterscheidet man exakte und nichtexakte Verknüpfungen. Im Unterschied zur Datenfusion, bei der Fälle basierend auf Ähnlichkeiten unter Verwendung spezifischer Algorithmen zusammengeführt werden, werden im vorliegenden Fall im Zuge von Datenintegrationen exakt verknüpfte Datensätze erstellt (vgl. Kiesl/Rässler 2006). Da eine eindeutige Nutzer-ID anbieterseitig geniert wird, ist das im Offline-Kontext virulente Problem fehlerhafter Werte der Identifikationsvariablen im vorliegenden Design nicht gegeben (vgl. Kiesl/Rässler 2005). Im Folgenden liefern wir ein empirisches Beispiel für eine komplementäre Verwendung der integrierten Daten. Im Anschluss diskutieren wir die Möglichkeiten der supplementären Datenintegration, also der methodischen Möglichkeiten der Verwendung integrierter Daten.

6 Den Befragten wird dabei die Möglichkeit gegeben, eine erneute Kontaktierung generell abzulehnen. In diesem Fall erfolgt keine weitere Einladung, an der Befragung teilzunehmen.

## 4 Komplementäre Datenintegration

Die bisherigen Ausführungen zeigen, dass Online-Befragungsdaten und elektronische Prozessdaten unterschiedliche informationelle Dimensionen abbilden und spezifische Schwächen und Stärken auf unterschiedlichen Gebieten aufweisen. Bei Befragungen können die im Rahmen der Forschungsfrage notwendigen Konstrukte zielgerichtet abgebildet und die subjektiven Wahrnehmungen der Akteure erhoben werden, mit Hilfe prozessproduzierter Daten hingegen ist es möglich, tatsächliche Handlungen und entstehende Strukturen zu erfassen. Im folgenden empirischen Anwendungsbeispiel aus der Partnerwahlforschung stellen wir eine komplementäre Verwendung der beiden Datentypen vor.

Eine zentrale Frage der klassischen soziologischen Partnerwahlforschung widmet sich den relevanten Bestimmungsgründen für das Zustandekommen von Paarbeziehungen. Neben strukturellen Faktoren werden dabei insbesondere individuelle Faktoren, v.a. Präferenzstrukturen in der Literatur diskutiert. Problematisch ist eine präferenzbasierte Erklärung des Zustandekommens von Paarbeziehungen jedoch dann, wenn sie post hoc vorgenommen wird, das heißt, theoretische Rückschlüsse auf in der Vergangenheit liegende Wahlhandlungen auf Basis der Beobachtung bereits existierender Beziehungen gezogen werden (vgl. Blossfeld & Timm 1997, 2003). Das hier vorgestellte Projektdesign erlaubt im Unterschied zu herkömmlichen Querschnitts- aber auch Längsschnittdesigns eine Beobachtung und Modellierung von Kontaktereignissen im Zeitverlauf vom Moment ersten virtuellen Begegnung an.

Unser Beispiel widmet sich der Erklärung der Anzahl der eingehenden Erstkontakte eines Nutzers anhand der im Profil gemachten Angaben. Dafür verwenden wir die Daten der Internetkontaktbörse im Zeitraum vom 01.01.2008 bis zum 29.06.2008. Wir verzichten an dieser Stelle auf eine theoretische Herleitung und inhaltliche Interpretation der Befunde. Das Beispiel dient lediglich der Illustration einer komplementären Verknüpfungsmöglichkeit der vorhandenen Daten. Wir verwenden in unserem Beispiel die prozessproduzierte Information „erstmalige Kontaktierung eines Empfängers durch einen Sender“ als abhängige Variable komplementär mit Befragungsinformationen, wie sie auf der Profilseite erfragt wurden. Die Anzahl der bei einem Empfänger eingehenden, voneinander unabhängigen Erstkontaktierungsereignisse wird dabei mit Hilfe eines einfachen Poissonmodells beschrieben[7] Tabelle 1 gibt die Ergebnisse für vier hierarchische Modelle wieder. Erwartungsgemäß sinkt die Anzahl eingehender Erstkontakte mit steigendem Alter, sinkender Bildung[8] sowie von Frauen zu Männern (siehe Modell (1) in Tabelle 1). Ebenfalls aus Modell (1) ist ersichtlich, dass die Angabe einer unverbindlichen Webfreundschaft als Beziehungsabsicht relativ zur Angabe einer festen Beziehung zu höherer Eingangsrate und die Angabe von Flirt und Heirat zu niedrigerer Kontaktrate führen. Bezüglich der familialen Situation lässt sich festhalten, dass die Angabe verheiratet, aber getrennt lebend oder verwitwet zu sein, zu einer signifikant geringeren Erstkontaktierung im Vergleich zu ledigen Personen führt. Modell (2) beinhaltet zusätzlich die Information, ob auf die Frage nach dem Vorhandensein von Kindern geantwortet wurde oder nicht. Die Auslassung dieser Information wirkt deutlich negativ auf die Anzahl der erhaltenen Erstkontakte. In Modell (3) wird zusätzlich der Einfluss der Vollständigkeit der im Profil gemachten Angaben in den „Freitextfeldern“[9] als unabhängige Variable modelliert. Der Umfang der angegebe-

7 Bei dieser Modellierung wird angenommen, dass die Anzahl der Ereignisse (hier: der eingehenden Nachrichten) einer Poisson-Verteilung folgt.

8 Das Bildungsniveau wurde ordinal skaliert und hat einen Wertebereich von 0 (geringes Bildungsniveau, ohne Ausbildung) bis 5 (hohes Bildungsniveau, Promotion).

9 Dieser Indikator wurde konfirmatorisch mit Hilfe einer Faktorenanalyse aus der jeweiligen Vollständigkeit von drei textuellen Profilangaben („Ich über mich“) definiert,

nen Informationen weist in diesem Modell keinen signifikanten Einfluss auf die Anzahl eingehender Erstkontakte auf. Differenziert man den Einfluss der Vollständigkeit in den Freitextfeldern nach Geschlecht (Interaktionsterm in Modell 4) ergibt sich jedoch ein deutlicher Befund. Für Frauen sinkt mit dem Grad der persönlichen, qualitativen Angaben die Anzahl der eingehenden Erstkontakte, während für männliche Nutzer die Eingangsrate wesentlich unberührt bleibt.

*Tabelle 1:* Erklärung der eingegangenen Kontaktanzahl (Zählratenmodell)

| | | Modell (1) b (s) | Modell (2) 'b (s) | Modell (3) b (s) | Modell (4) b (s) |
|---|---|---|---|---|---|
| **Alter** | | -.007** | -.008*** | -.008** | -.007** |
| | | .002 | .002 | .003 | .003 |
| **Geschlecht** | | -.322*** | -.305*** | -.335*** | -.577*** |
| (Ref.: Frau) | | .035 | .035 | .040 | .092 |
| **Bildung** | | .111*** | .110*** | .134*** | .134*** |
| | | .012 | .011 | .013 | .013 |
| **Beziehungswunsch** | Flirt | -.231** | -.207*** | -.106 | -.095 |
| (Ref.: Feste Bez.) | | .072 | .072 | .091 | .091 |
| | Webfreundschaft | .263*** | .289*** | .334*** | .327*** |
| | | .055 | .055 | .066 | .066 |
| | Heirat | -.164 | -.233 | -.364 | -.366 |
| | | .237 | .237 | .237 | .237 |
| **Familiäre Situation** | Verheiratet | .177 | .098 | .237 | .218 |
| (Ref.:Ledig) | | .124 | .124 | .138 | .138 |
| | Geschieden | -.017 | -.095 | -.166* | -.176* |
| | | .068 | .068 | .073 | .073 |
| | Verheiratet, aber getrennt lebend | -.353* | -.460** | -.585*** | -.588*** |
| | | .169 | .169 | .175 | .175 |
| | Verwitwet | -.347* | -.413* | -.520* | -.537** |
| | | .169 | .169 | .204 | .204 |
| **Keine Angabe zu Kindern** | | - | -.274*** | -.226*** | -.222*** |
| | | | .035 | .055 | .055 |

| | Modell (1) b (s) | Modell (2) 'b (s) | Modell (3) b (s) | Modell (4) b (s) |
|---|---|---|---|---|
| **Vollständigkeit des Profils (Decentile)** | - | - | -.002 | -.056** |
| | | | .008 | .020 |
| **Interaktion:** | - | - | - | .039** |
| **Geschlecht*Vollständigkeit** | | | | .013 |
| Konstante | 1.050** | 1.167*** | 1.183*** | 1.501*** |
| | .089 | .089 | .124 | .164 |
| N | 1517 | 1517 | 1011 | 1011 |
| Freiheitsgrade | 10 | 11 | 12 | 13 |
| McFadden's $R^2$ | 0.030 | 0.037 | 0.043 | 0.045 |
| Cragg & Uhler's $R^2$ | 0.157 | 0.192 | 0.251 | 0.258 |

*Quelle:* Datenbankauszug einer großen deutschsprachigen Internetkontaktbörse; eigene Berechnungen.
*Signifikanzniveaus:* * $p \leq 0{,}1$; ** $p \leq 0{,}05$; *** $p \leq 0{,}01$.

Das empirische Anwendungsbeispiel veranschaulicht verschiedene Möglichkeiten einer komplementären Verwendung objektiver Prozessdaten und subjektiver Befragungsdaten. Zum einen werden *Prozessdaten in Form beobachteter Interaktionen* (Kontaktierungsereignisse) und (*subjektive) Befragungsdaten* (Fragen zur Soziodemographie, familiären Situation sowie zum Beziehungswunsch) herangezogen. Die relationale Datenstruktur ermöglicht es zusätzlich, subjektive Angaben durch die Beobachtung der Reaktionen anderer Nutzer zu *objektivieren*. So können Befragungsdaten (Grad der Profilvollständigkeit) in objektivierter Form verwendet werden. Ähnlich lassen sich Prozessdaten nun auch *subjektivieren*, also hinsichtlich ihrer subjektiven Wahrnehmung diskutieren. Die objektive Nicht-Antwort auf die Frage nach dem Vorhandensein von Kindern wird von den potentiellen Kontaktpartnern wahrgenommen und beurteilt, bevor diese (wiederum objektiv beobachtbar) mit Handlung reagieren. Die komplementäre Verwendungsweise ermöglicht also auch fehlende Werte inhaltlich nutzbar zu machen, z.B. um die Reaktionen auf die Abwesenheit eines relevanten Signals abzubilden. Schließlich sind *Interaktionsterme* von Befragungs- und Prozessdaten (Interaktion von Geschlecht und Profilvollständigkeit) modellierbar und, wie in unserem Beispiel, recht aufschlussreich.

Neben der Beschränkung auf den exemplarischen Charakter unserer Analyse liegt eine prinzipielle methodische Schwäche von Modellen, die auf derartigen Social-Web-Daten fußen darin, dass die Online-Befragungsdaten extrem selektiv sind. Das Problem der selektiven Befragungsteilnahme und des methodischen Umgangs mit dieser wird im folgenden Abschnitt diskutiert.

## 5 Supplementäre Datenintegration

Beim Vergleich der unterschiedlichen Stärken und Schwächen prozessproduzierter Daten und Befragungsdaten haben wir bereits darauf verwiesen, dass sich beide Datensätze in einem wesentlichen methodischen Gesichtspunkt unterscheiden: dem Grad ihrer Vollständigkeit. Da die Nutzer stets Daten im Vollzug einer online getätigten Handlung produzieren, haben sie nicht die Möglichkeit, die Teilnahme zu verweigern. Sind die prozessproduzierten Daten also vollständig und erschöpfend, so sind die Befragungsdaten demgegenüber als unvollständig und zudem selektiv zu bezeichnen.

Das Problem fehlender Daten, das in Unit- oder Item-Nonresponse unterschieden werden kann, besteht nicht nur in dem zahlenmäßigen Ausfall eines Teils der Befragten, sondern auch und gerade in der verzerrten Stichprobe die sich aufgrund systematisch beobachteter und nicht-beobachteter Unterschiede ergibt. Rubin (2004) untergliedert die unterschiedlichen Ausfallmechanismen in Missing Completely at Random (MCAR), Missing at Random (MAR) und Not Missing at Random (NMAR). Ist die Annahme eines völlig zufälligen Auswahlmechanismus nicht haltbar, so führt dies gegebenenfalls zu verzerrten Schätzern, da die zur Analyse herangezogenen vollständigen Fälle sich von den nicht-berücksichtigten Fällen und damit von der Aussagegesamtheit systematisch unterscheiden können. Das Problem des Nonresponse-(Selection)-Bias, also des systematischen Ausfalls von Befragungsteilnehmern, wird in einem Paneldesgin verschärft, da mit jeder Welle nicht nur weniger Respondenten vorliegen, sondern diese auch zunehmend spezifische Eigenschaften besitzen, was inferenzstatistische Verallgemeinerungen weiter erschwert.

Um derartige, systematische Ausfälle handhabbar zu machen, existieren unterschiedliche methodische Ansätze, die herangezogen werden können, wenn die MCAR-Annahme fälschlicherweise getroffen wurde, also fehlende Werte bei den Befragungen von (unbeobachteten) Variablen abhängen (vgl. Schnell 2005:469).

Vollständige Prozessdaten können zunächst deskriptiv verwendet werden, um zu vergleichen, ob sich die Respondenten von den Nicht-Respondenten unterscheiden. Diese Unterscheidung bezieht sich dann nicht nur auf die zumeist ebenfalls lückenhaften Profilangaben der Nutzer z.B. zu Alter, Geschlecht und Bildungsstand oder auf Beobachtungen wie bspw. individuelle Login-Häufigkeiten, sondern insbesondere auf die sinnhaften Aspekte des beobachteten Verhaltens, wie beispielsweise die sozioökonomische Zusammensetzung der Kontakte eines Nutzers bzw. dessen offenbarte Präferenzen. Auch wenn für einen Nutzer keine Befragungsdaten vorliegen, verfügen wir so über vollständige und valide Records für diesen Nonrespondenten im Zeitverlauf, was für den Umgang mit fehlenden Werten einen grossen Vorteil darstellt. Diese Datenlage ist recht vorteilhaft, da nun ein Grundstock an Informationen für jeden Nutzer vorliegt, was dem methodischen Postulat Bethlehems (2002: 277) entspricht: “To be able to asses and remove a nonresponse bias, it is important to collect as much auxiliary information as possible”.

Im Falle von Antwortverweigerungen können im vorliegenden Fall nun Selektivitäten basierend auf prozessproduzierten Daten behandelt werden. Insbesondere die Propensity Score Gewichtung (Rosenbaum/ Rubin 1984), Selektionskorrekturen (Heckmann 1979), und die multiple Imputation (Rubin 1987), werden als aktuelle Möglichkeiten der statistischen Behandlung fehlender Werte diskutiert.

Der Propensity Score stellt eine konditionale Wahrscheinlichkeit der Befragungsteilnahme bei gegebenem Vektor beobachteter Kovariaten dar und wird als Kehrwert der Teilnahmewahrscheinlichkeit als Beobachtungsgewicht verwendet. Die prozessproduzierten Daten können dann als ein zusätzlicher Teil dieses Kovariatenvektors verwendet werden. Tritt der Selektionsbias aufgrund

beobachteter Charakteristika auf, ist dann eine Adjustierung über Propensity Gewichtung möglich. Die Propensity Gewichtung erzeugt jedoch inkonsistente Schätzer, wenn nach Kontrolle der Kovariaten noch „Selection on Unobservables“ vorliegt .

Beim zweiten Korrekturverfahren, der Treatment Selektionskorrektur, wird die Befragung als Treatment aufgefasst. Dabei werden simultan ein Outcome und die Selektion in die Befragung unter Verwendung eines umfangreichen Datenvektors modelliert. Die prozessproduzierten Variablen können hier insbesondere als Pre-Treatment-Variablen, also als Kovariaten, die den Selektionsprozess in die Befragung modellieren, dienlich sein. Vorteilhaft an diesem Ansatz ist, dass auch hohe Ausfallraten kompensiert werden können und ein Test für „Selection on Unobservables“ möglich ist. Dem steht der Nachteil gegenüber, dass für jedes (Regressions-) Modell eine eigene Selektionsmodellierung erstellt werden muss.

Das dritte Verfahren der multiplen Imputation ist ein auf der Bayes-Statistik beruhendes Datenmehrungsverfahren, bei dem die fehlenden Daten von Nonrespondenten auf Basis theoretischer Annahmen über den Ausfallmechanismus ersetzt werden können. Im vorliegenden Fall können die Befragungs- und Prozessdaten vorheriger und zeitgleicher Wellen für Imputationen herangezogen werden. Die Hinzunahme prozessproduzierter Informationen macht dabei die Missing at Random-Annahme (MAR) ggf. plausibler, da der Kovariatenvektor umfangreicher wird. Rässler (2000) hat mit Hilfe einer Simulationsstudie darüber hinaus die relative Robustheit multipler Imputationen gegenüber der Verletzung der MAR-Vorraussetzung zeigen können. Ein weiterer Vorteil liegt darin, dass gute Ergebnisse auch bei nicht normalverteilten Daten erzielt werden (Graham/Hofer 2000).[10]

Das Ziel dieser drei Verfahren besteht darin, die zuvor aufgezeigten Probleme systematischer Selektivität unter Verwendung eines Maximums an vorliegenden Informationen zu entschärfen, um so inferenzstatistische Schlüsse zu ermöglichen.

## 6 Ausblick

Die Analyse von befragungsbezogenen Paradaten ist ebenso (vgl. Kaczmirek 2008) wie die strukturalistische Verwendung derartiger Daten (vgl. Stegbauer/Rausch 2006) verbreitet, während die komplementäre und supplementäre Verwendung der Inhalte von Befragungsdaten und elektronischen Beobachtungsdaten ihren Zenit noch nicht erreicht hat. Dies lässt sich darüber verstehen, dass erst mit Hilfe des Internets und der zunehmenden Verlagerung sozialer Aktivitäten ins Internet die technischen und forschungslogischen Möglichkeiten entstehen, die diese bislang wenig berücksichtigte Form sozialwissenschaftlicher Datenerhebung und -integration betreffen. Gerade im Social Web-Kontext scheinen jedoch besondere Chancen für die Triangulation von Befragungs- und Prozessdaten zu liegen, da Online-Aktivitäten ohnehin digitale Verhaltensspuren hinterlassen, deren Erhebung im Vergleich zu Offline-Kontexten weniger aufwendig ist. Im Kontext der sozialwissenschaftlichen Erforschung sozialer Formen des Internets sollte dabei in Zukunft vermehrt geprüft werden, welche prozessproduzierten Daten bereits vorliegen und nutzbar gemacht werden können.

Es ist zu erwarten, dass im Zuge der Ausweitung des Internets auf immer weitere Teile des menschlichen Alltags auch die Möglichkeiten der komplementären und supplementären Verwen-

10 Für eine detaillierte Diskussion der Multiplen Imputation vor dem Hintergrund der Datenstruktur des Projektes vgl. Schmitz et al. (2008).

dung elektronischer Prozessdaten und klassischer Befragungsdaten in den Werkzeugkoffer der empirischen Sozialwissenschaften zunehmend Eingang finden.

## Literaturverzeichnis

*Baur, Nina und Michael J. Florian* (2008): Stichprobenprobleme bei Online-Umfragen. In: *Nikolaus Jackob, Harald Schoen, Thomas Zerback* (Hrsg.): Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung. Wiesbaden: VS Verlag für Sozialwissenschaften.

*Baur, Nina* (2008): Mixing Process-Generated Data in Market Sociology. Eingereichtes Manuskript.

*Bergmann, Jörg und Christoph Meier* (2000): Elektronische Prozessdaten und ihre Analyse, S. 429-437 in: *Uwe Flick, Ernst von Kardorff und Ines Steinke* (Hrsg.): Qualitative Sozialforschung. Ein Handbuch. Reinbek: Rowohlt.

*Berker, Thomas* (1999): WWW-Nutzung an einer deutschen Hochschule - Computer, Sex und eingeführte Namen. Ergebnisse einer Protokolldateienanalyse, In: *Bernad Batinic, Andreas Werner, Lorenz Gräf und Wolfgang Bandilla* (Hrsg.): Online Research. Methoden, Anwendungen und Ergebnisse. Göttingen: Hogrefe, Verlag für Psychologie .

*Bethlehem, Jelke G.* (2002): Weighting Nonresponse Adjustments Based on Auxiliary Information. In: *Robert. M. Groves, Don A. Dillman, John L. Eltinge und Roderick J.A. Little* (Hrsg.) Survey Nonresponse. New York:Wiley.

*Blossfeld, Hans-Peter und Andreas Timm* (2003): Educational systems as marriage markets in modern societies: A conceptual framework. In: Hans Peter Blossfeld & Andreas Timm (Hg.): Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies. Dordrecht, S. 1-18.

*Blossfeld, Hans-Peter und Andreas Timm* (1997): Der Einfluss des Bildungssystems auf den Heiratsmarkt – Eine Längsschnittanalyse der Wahl des ersten Ehepartners im Lebenslauf. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 49 (3), S.440-476.

*Bosnjak, Michael* (2003): Web-basierte Fragebogenuntersuchungen - Methodische Möglichkeiten, aktuelle Themen und Erweiterungen, S. 109-133 in: *Informationszentrum Sozialwissenschaften (IZ) & Arbeitskreis Sozialwissenschaftlicher Institute e.V. (ASI)* (Hrsg.) Online-Erhebungen. Bonn: IZ

*Couper, Mick und Elisabeth Coutts* (2004): Online-Befragung. Probleme und Chancen verschiedener Arten von Online-Erhebungen, S. 217-243 in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 44.

*de Keulenaer, Femke* (2005): Using process data to predict attrition from a panel survey: a case study. S. 3844-3851 in: Proceedings of the American Statistical Association, Survey Research and Methods Section. Alexandria, Va.: American Statistical Association, 2005.

*Flick, Uwe* (2004) : Triangulation. Eine Einführung. Reihe Qualitative Sozialforschung – Bd. 12, *Ralf Bohnsack, Christian Lüders & Jo Reichertz* (Hrsg.)Wiesbaden: VS Verlag für Sozialwissenschaften.

*Graham, John. W. und Scott. M. Hofer* (2000): Multiple imputation in multivariate research. S. 201-218 in: *Todd. D. Little, Kai. U. Schnabel, & Jürgen. Baumert* (Hrsg.) Modeling longitud-

inal and multiple-group data: Practical issues, applied approaches, and specific examples. Hillsdale: NJ.

*Hartmann, Josef und Gerhard Krug* (2007): Verknüpfung von Befragungs- und Prozessdaten- Selektivität durch fehlende Zustimmung der Befragten? IAB Discussion Paper, 13.

*Heckman, James*. (1979) Sample selection bias as a specification error. Econometrica, 47, 153–61.

*Kaczmirek, Lars und Wolfgang Neubarth* (2007): Nicht-reaktive Datenerhebung: Teilnahmeverhalten bei Befragungen mit Paradaten evaluieren, in: *Deutsche Gesellschaft für Online-Forschung* (Hrsg.): Online-Forschung 2007: Grundlagen und Fallstudien, Neue Schriften zur Online-Forschung, 1, Köln.

*Kiesl, Hans und Susanne Rässler* (2005): Techniken und Einsatzgebiete von Datenintegration und Datenfusion, in: *König, Christian; Matthias Stahl, Erich Wiegand* (Hrsg.) (2005): Datenfusion und Datenintegration, 6. Wissenschaftliche Tagung, Tagungsberichte Band 10.

*Kiesl, Hans und Susanne Rässler* (2006): How valid can data fusion be, Journal of Official Statistics .

*Lepowski, James M. und Mick P. Couper* (2002) : Nonresponse in the second wave of longitudinal household surveys. S. 259-272 in: *Groves, Robert. M.; Don A. Dillman, John L. Eltinge und Roderick J.A. Little* (Hrsg.) Survey Nonresponse. Chichester, UK: Wiley.

*Müller, Paul J.* (1977): Die Analyse prozeß-produzierter Daten. Stuttgart: Klett-Cotta.

*Rässler, Susanne* (2000): Ergänzung fehlender Daten in Umfragen. In: Jahrbücher für Nationalökonomie und Statistik. Bd. 220/1 Stuttgart.

Rosenbaum,Paul R. und Rubin, Donald B. (1984). "Reducing bias in observational studies using sub-classification on the propensity score". JASA, 79:516-24.

Rubin, Donald B. (1987) Multiple Imputation for Nonresponse in Surveys. J. Wiley & Sons, New York.

*Rubin, Donald B.* (2004): Multiple Imputation for Nonresponse in Surveys. Reprinted with appendices as a "Wiley Classic." New York: John Wiley and Sons. Appendix 1: (1977) "TheDesign of a General and Flexible System for Handling Non-Response in Sample Surveys."; Appendix 2: (1983) "Progress Report on Project For Multiple Imputation of 1980 Codes." Manuscript distributed to the U.S. Bureau of the Census, the U.S. National Science Foundation
and the Social Science Research Foundation.

*Stegbauer, Christian und Alexander Rausch* (2006): Strukturalistische Internetforschung- Netzwerkanalysen internetbasierter Kommunikationsräume. Wiesbaden: VS Verlag für Sozialwissenschaften.

*Schmidt, Jan* (2006): Social Software: Onlinegestütztes Informations, Identitäts und Beziehungsmanagement. S. 37-46 in: Forschungsjournal Neue Soziale Bewegungen, Nr. 2/2006.

*Schmitz, Andreas, Messingschlager, Martin, Blossfeld, Hans-Peter und Rässler, Susanne (2009): Multiple Imputation mit web-generierten Prozessdaten zur Behandlung fehlender Werte von Web-Befragungen.( Arbeitspapier)*

*Schnell, Rainer, Hill,Paul B.und Esser, Elke (2005) :* Methoden der empirischen Sozialforschung. Oldenburg Verlag.

*Schulz, Florian; Jan Skopek, Doreen Klein, Andreas Schmitz* (2008): Wer nutzt Online-Dating in Deutschland? Zeitschrift für Familienforschung (im Druck).

*Welker, Martin und Olaf Wenzel* (2007): Online-Forschung 2007. Grundlagen und Fallstudien. Köln: Halem-Verlag.

*Zanutto, Elaine und Alan M. Zaslavsky* (2002): Using Administrative Records to Impute for Nonresponse. S. 403-415 in: Survey Nonresponse. *Robert Groves, Don A. Dillman, Roderick J.A. Little, und John Eltinge* (Hrsg.) New York: John Wiley.

## Zu den Autoren

Andreas Schmitz, M.A., Dipl. Soz. Doreen Klein, Dipl. Kfm. Jan Skopek,
Dpl. Soz. Florian Schulz, Prof. Dr. Hans Peter Blossfeld

Lehrstuhl für Soziologie I
Otto-Friedrich-Universität Bamberg
*Kontakt:* andreas.schmitz@uni-bamberg.de