

## Bericht zur IfL Forschungswerkstatt #4

### Big Data, Open Geography?

Reflektierter Umgang mit Big Data in den raumbezogenen Wissenschaften

Leipzig, 26. – 27. März 2018

Daten sind überall. Ob wir wollen oder nicht, wir alle produzieren und konsumieren Daten. Von der Verkehrsplanung über das Gesundheitswesen bis zu unseren sozialen Netzwerken - unsere Welt ist durchzogen von Gelegenheiten, Daten über unser Verhalten, unsere Vorlieben, Gewohnheiten und Lebensweisen zu verbreiten und zu sammeln. Den *data-trails*, die wir beispielsweise im Netz hinterlassen - immer unter dem Versprechen der Optimierung und Personalisierung von Inhalten - schenken wir oftmals keine große Beachtung. Durch die massenhafte, automatisierte und in Echtzeit mögliche Sammlung dieser Daten wurde ein gesamtgesellschaftlicher Prozess angestoßen, dessen Auswirkungen wir in den nächsten Jahren zunehmend spüren werden. Auch und gerade für die Wissenschaft stellen sich neue Herausforderungen im Umgang mit solchen großen Datensätzen, weshalb die IfL Forschungswerkstatt #4 in diesem Jahr unter dem Thema „Big Data, Open **Geography** – Reflektierter Umgang mit Big Data in den raumbezogenen Wissenschaften“ stattfand. Neben den neuen Möglichkeiten der Datenakquise erfordern besonders die oben genannten Aspekte der Automatisierung und Echtzeitgewinnung von Daten ganz neue wissenschaftliche Zugänge. Während klassische Methoden den Umgang mit statischen Datensets ermöglichen, sind Big-Data-Sets dynamisch und volatil. Es findet ein Paradigmenwechsel von theoriebasierter Forschung hin zu datenbasierter Forschung statt. Die wissenschaftstheoretischen Auswirkungen spüren wir in der Forschungsgemeinde schon jetzt, doch sind dahinter gesamtgesellschaftliche Prozesse zu erkennen, welche unseren erkenntnistheoretischen Horizont einerseits erweitern, andererseits aber auch einengen können. Gerade die Rolle, welche Algorithmen bei der Produktion und Auswertung von Big Data zukommt, betrachtet die wissenschaftliche Gemeinde mit Skepsis. Um der Reflexion der theoretischen Aspekte, wie etwa den wissenschaftstheoretischen Implikationen oder Datenschutzaspekten einerseits, und den anwendungsbezogenen Herausforderungen wie beispielsweise der Akquise von Social-Media-basierten Datensets andererseits jeweils genug Raum in unserer Veranstaltung zu bieten, wurden die Formate der IfL Forschungswerkstatt #4 unter diesen beiden Gesichtspunkten ausgewählt. Zu Beginn leiteten fünf Experten im SynPodium mit ihren Erfahrungen aus der Feldforschung sowie der Praxis eine angeregte Diskussion zum Thema ein. Anschließend boten fünf Diskussionsforen den Teilnehmenden die Möglichkeit, verschiedene Aspekte der Nutzung von Big Data zu vertiefen. Die ExperiSpaces am zweiten Veranstaltungstag machten unterschiedliche Anwendungsbeispiele für die Teilnehmenden erfahrbar und ermöglichten somit die Exploration des Forschungsalltags mit Big Data.

### Das SynPodium

Den Auftakt der Veranstaltung bildete in diesem Jahr wieder das SynPodium, eine interaktive Podiumsdiskussion, in der fünf Diskutanten mit ihren Erfahrungen im Umgang mit Big Data die Diskussion eröffneten. Die Podiumsteilnehmer waren **Dirk Burghardt** (*Technische Universität Dresden, Institut für Kartographie*), **Henning Füller** (*Humboldt-Universität zu Berlin, Geographisches*

Institut), **Michael Klemm** (Universität Koblenz-Landau, Institut für Kulturwissenschaften), **Jörg Scheffer** (Universität Passau, Fachbereich Geographie) sowie **Till Straube** (Goethe-Universität Frankfurt am Main, Institut für Humangeographie). **Tim Leibert** (Leibniz-Institut für Länderkunde) moderierte das SynPodium. Nach einer kurzen Vorstellungsrunde und Präsentation der Berührungspunkte der Podiumsteilnehmer mit Big Data, begann die Diskussion mit einer begrifflichen Einordnung. Die Diskutanten hoben dabei jeweils unterschiedliche Aspekte von Big Data hervor, etwa das (mangelnde) Vertrauen in die Algorithmen, die der Datenakquise sowie Auswertung zugrunde liegen, das Entstehen von Metadaten welche ebenfalls in die Analyse mit einfließen, die Zusammenführung von Daten aus unterschiedlichen Quellen sowie neue statistische Auswertungsmethoden, welche den dynamischen Charakter von Big Data abzubilden vermögen. Im weiteren Verlauf der Diskussion berichteten die Podiumsteilnehmer aus dem breiten Anwendungsspektrum von Big Data. Eine kartographische Perspektive brachte **Dirk Burghardt** in die Diskussion ein. Dabei ging er beispielsweise auf die Überschneidung von klassischer Datenerhebung, etwa mittels Satellit- oder Laserscannerdaten, sowie neuen Methoden der Datenerhebung, zum Beispiel via Smartphone oder Social-Media-basierten Daten ein. Auch das der Nutzung von Big Data zugrunde liegende technische Knowhow, der damit verbundene Aufwand in Bezug auf die Speicherung und Handhabung der Daten sowie die Abhängigkeit von Dritten bei der Datenakquise wurden thematisiert. **Henning Füller** berichtete von den Erfahrungen seines Habilitationsprojekts, welches sich mit Big Data Anwendungsmöglichkeiten im Gesundheitswesen befasst. Neben dem Problem der Intransparenz der Big Data und Machine Learning zugrunde liegenden Algorithmen, sprach er Steuerungsformen von Big Data, die Disparitäten zwischen dem theoretischen Verständnis eines Prozesses sowie dessen Erfassung mittels Daten und die Frage nach der Erfassbarkeit und Prognostizierung von sozialem Verhalten an. Einen Einblick in die diskursanalytische Praxis gab **Michael Klemm**, der Mediendiskurse auf sozialen Plattformen und dort auftretende Phänomene wie Hashtagkampagnen oder Shit- und Candystorms untersucht. Qualitative Auswertungen von Big Data am Fallbeispiel sozialer Medien, Möglichkeiten der politischen Einflussnahme mittels Botnetzwerken, die Gefahr ein zu großes Augenmerk auf die Erhebung rein digitaler Daten zu legen, sowie Schwerpunktverschiebungen im Rahmen wissenschaftlicher Argumentationen wurden dabei in die Diskussion eingebracht. **Jörg Scheffer** thematisierte die Auswirkungen von Digitalisierung und Big Data im Zusammenhang mit sozialen Gruppen, datendurchsetzten Räumen und Auswirkungen auf das gesellschaftliche Zusammenleben. Dabei beschrieb er das mangelnde Wissen über kontinuierliche Datensammlung unter den Nutzenden von ICTs, die Sachzwänge, welche Menschen zur Nutzung von datensammelnder Technologie anreizt, die geographische Perspektive auf neue digitale Räume sowie die Kommerzialisierung von Big Data und das dadurch entstehende Machtgefälle zwischen Unternehmen und jenen, deren Daten gesammelt wurden. Von seinen Praxiserfahrungen im Rahmen seiner laufenden Promotion zur Nutzung von Big Data durch die Polizei sowie Erfahrungen in der Lehre berichtete **Till Straube**. Neben dem Big Data zugrundeliegenden statistischen Paradigmenwechsel, dem Verschneiden von Datensets aus unterschiedlichen Quellen und dem sich ändernden Verständnis von Wissen und Wahrheit wies er besonders auf die Lücke zwischen Theorie und Praxis in der Handhabung von Daten hin.

Im weiteren Verlauf des SynPodiums schalteten sich auch die weiteren Teilnehmenden der IfL Forschungswerkstatt #4 in die Diskussion ein und sprachen dabei unterschiedliche Aspekte der Big Data Thematik an. So wurde etwa auf die Problematik von Prognosen, welche ohne Theoriebezug entstehen und somit zu einer Steuerung ohne Verständnis über die Prozesse führen könnten hingewiesen. Vorteile von Big Data und Machine Learning in Bezug auf den Umgang mit Unsicherheiten oder unvollständigen Daten wurde Anhand von Fallbeispielen wie etwa dem Katastrophenschutz oder Polizeieinsätzen thematisiert. Auch Fragen des Zugangs zu sowie die kritische

Betrachtung von Daten, welche nicht selbst erhoben wurden, wie etwa Daten sozialer Medien wurden diskutiert.

### *Die Diskussionsforen*

Im Anschluss an das SynPodium fanden jeweils unter Leitung eines Podiumsteilnehmers Diskussionsforen statt, welche die Möglichkeit boten, unterschiedliche Themenkomplex von Big Data tiefgehend zu diskutieren. **Dirk Burghardt** widmete sich in diesem Format den Visualisierungsmöglichkeiten von Big Data. Hierbei wurden Aspekte wie die Wechselwirkungen zwischen Mensch und Computer und die damit verbundene optimale Aufgabenteilung, Visualisierungen als Möglichkeit, komplexe Informationen einfach zu vermitteln ebenso diskutiert wie mögliche interaktive oder gamifizierte Visualisierungsformate. Eine kritische Perspektive auf Big Data in der raumbezogenen Forschung brachte **Henning Füller** mit seinem Diskussionsforum ein. Dabei wurde die Distinktion der Begriffe Big Data und Machine Learning, mangelnde Open Source Infrastrukturen für die Forschung sowie mögliche Nutzungsformen von Machine Learning Algorithmen für kritische Forschung thematisiert. **Michael Klemm** diskutierte mit den Teilnehmenden unter der Überschrift „Big Data, Qualitative Forschung und Forschungsethik“ über die Ergebnisse seiner Diskursforschung im Bereich der sozialen Medien. Mit Hilfe von Twitonomy, einem Tool zur Exploration von via sozialen Medien gewonnenen Daten, ließe sich ein dynamische und spielerische Erkundung von Tweets umsetzen, welche systematischere Datenexplorationsansätze, beispielsweise Untersuchungen zu „Social-TV Diskursen“, unterstützen kann. Die Gruppe um **Jörg Scheffer** erörterte geographische Perspektiven auf Raum und Raumtransformationen im Zeitalter der Digitalisierung. Neben den Folgen der Nutzung digitaler Daten im Alltag und Veränderungen der Qualität digitaler und physischer Räume, fand auch die sich wandelnde Definitionsmacht über Räume und Begriffe ihren Platz in der Diskussion. Das von **Till Straube** geleitete Diskussionsforum widmete sich dem Prozess der Reproduktion von Verräumlichungen über Big Data. Während sich die Raumkonzepte im letzten Jahrhundert stark gewandelt haben, findet die Interpretation von Big Data meist unter den Gesichtspunkten des klassisch euklidischen Container-raums statt, beispielsweise über GeoTags. Das Spannungsfeld Mensch-Raum-Maschine ist jedoch weitaus komplexer und erfordert eine dieser Komplexität angemessene Forschungsweise, etwa über eine ethnographische Betrachtungsweise der technisch komplexen Apparate, die die Basis für Big Data bilden. Das Ende des ersten Veranstaltungstages markierte die Reflexion der Diskussionsforen Reflexion der Diskussionsforen entlang von kurzen Posterpräsentationen aus den einzelnen Gruppen.

### *Die ExperiSpaces*

Neben den spannenden theoretischen Implikationen der zunehmenden Verbreitung von Big Data ist auch die praktische Anwendung in der Forschung ein wichtiger Teil des wissenschaftlichen Prozesses. Daher wurden am zweiten Veranstaltungstag ExperiSpaces angeboten, Workshops mit Experiment-Charakter, die verschiedene Aspekte des Forschungsalltags mit Big Data für die Teilnehmenden zugänglich machten. Die Teilnehmenden konnten die am Vortag besprochenen Themen unter praktischen Gesichtspunkten aufgreifen und weiterentwickeln. So stellten **Alexander Dunkel** und **Dirk Burghardt** (*Technische Universität Dresden, Institut für Kartographie*) ihre Arbeit zu Tag Maps, einer Visualisierungsmethode zu via Social Media gewonnener Bild- und Meta-

daten, und der darauf basierenden Auswertung von kollektiver Wahrnehmung vor. Unter Anwendung selbst erstellter Tools und geographischer Informationssysteme ermöglichen Tag Maps eine Visualisierung von Fototags auf der Karte eines vorher ausgewählten Gebiets. Die Größe der Tags auf der Karte steht als Variable für die Häufigkeit der Nennungen des jeweiligen Begriffs. Standortbasierte Social Media Daten können somit explorativ untersucht werden, wobei die Reflexion der Grenzen der zugrundeliegenden Datensätze ein wichtiger Teil des Erkenntnisprozesses bleibt. Zur Exploration gamifizierter Ansätze von Big Data luden **Laura Ziegler** und **Max A. Irmer** (*Bauhaus-Universität Weimar, Institut für Europäische Urbanistik*) ein. Über die Website MatchMyCity.org konnten die Teilnehmenden ein interaktives Städte-Speed-Dating erleben, auf dessen Basis im Anschluss die methodischen und methodologischen Möglichkeiten und Herausforderungen des Ansatzes sowie ethische Fragestellungen diskutiert wurden. Da die Methode sich auf digital erhobene Daten stützt, ist der Zugang zu entsprechenden Erhebungswerkzeugen ein limitierender Faktor, ebenso die Kompetenz der Nutzer\*innen. Andererseits können die abwechslungsreichen Aspekte der Datenerhebung sowie ein Belohnungssystem eine höhere Teilnahmebereitschaft (zumindest bestimmter Gruppen) hervorrufen. Besonders geeignet ist die Methode für die Erhebung quantitativer Daten, wobei immer die Notwendigkeit der kritischen Reflexion der Validität sowie Repräsentativität des Datensatzes besteht. Insbesondere wenn die Daten nicht von den Forschenden selbst erhoben werden, sollte größtmögliche Transparenz ein integraler Teil der Datenanalyse sein. **Thilo Wiertz** und **Tobias Schopper** (*Albert-Ludwigs-Universität Freiburg, Institut für Umweltsozialwissenschaften und Geographie*) boten die Möglichkeit, praktische Erfahrungen im Umgang mit quantitativen Verfahren der Textanalyse mittels CPQWeb und R zu sammeln. Die Teilnehmenden konnten im Verlauf des Workshops eigene Ideen für die Anwendung von korpuslinguistischen Analysen entwerfen und anhand von bestehenden Datensätzen die Methode ausprobieren. Dabei wurden die Vorzüge dieser Methode insbesondere für die Mustererkennung in großen Textkorpora ebenso diskutiert wie Vergleiche zwischen Teilkorpora sowie zeitliche Verschiebungen innerhalb von Diskursen. Im ExperiSpace wurden aber auch limitierende Faktoren der korpuslinguistischen Analyse herausgearbeitet. So ist die Methode nicht für schnelle Ergebnisse, die Analyse dynamischer Medien oder sehr kleine Datensätze geeignet.

### Ausblick

Wir möchten allen Teilnehmenden und Mitwirkenden der IfL Forschungswerkstatt #4 für die anregenden Diskussionen und Beiträge im Verlaufe der zwei Veranstaltungstage danken. Auch in diesem Jahr haben wir eine insgesamt sehr positive Resonanz im Nachgang der Veranstaltung widerspiegelt bekommen. Insbesondere die praktische Arbeit im Rahmen der ExperiSpaces sowie die informelle Gesprächsatmosphäre wurden dabei hervorgehoben. Ebenfalls geäußerte konstruktive Kritikpunkte werden wir natürlich für die Planung der IfL Forschungswerkstatt #5 berücksichtigen.

**Thomas Tillmann**