

Gesellschaftlicher Mehrwert von Open Data aus der Zivilgesellschaft

Claudia Juech

Ich habe in den letzten fünf Jahren für Stiftungen gearbeitet, die ausschließlich auf die gemeinnützige Anwendung von Daten und datenwissenschaftlichen Methoden spezialisiert waren. Im Rahmen dieser Arbeit habe ich u.a. ausgewählte gemeinnützige Organisationen konkret und praktisch mithilfe eines Teams von Dateningenieur:innen, Cloud-Ingenieur:innen und Datenwissenschaftler:innen unterstützt. Zusätzlich erhielten diese Non-Profit-Organisationen Fördermittel für Personal- und Technologiekosten. In einzelnen Fällen haben wir auch Technologielizenzen bereitgestellt.

Einige der Projekte wurden mit US-amerikanischen Organisationen durchgeführt, teilweise mit Forschungseinrichtungen, andere mit Organisationen in Europa, Afrika, Asien und Lateinamerika. Thematisch waren die Projekte weit gestreut und reichten von Gesundheit zu Umweltfragen und Menschenrechtsthemen.

Was habe ich gelernt?

Non-Profit-Organisationen, die keine Datenorganisationen sind, d. h. deren Geschäftszweck nicht die Bereitstellung von Daten ist, erheben in der Regel zwei Arten von programmatischen Daten. Programmatische Daten sind in meiner Definition Daten, die sich auf die inhaltliche Arbeit des Non-Profits beziehen, anders als Finanz- oder Personaldaten oder allgemeine Unternehmensdaten.

Die zwei Arten von programmatischen Daten sind:

- 1) Projektbezogene Daten – werden im Rahmen von Fördermittelprojekten zur Dokumentation benötigt, um gegenüber dem Geldgeber Rechenschaft abzulegen (wie viele Personen haben eine Impfung erhalten in welchem Zeitraum), und manchmal auch genutzt, um weiterreichende Wirkungsanalysen (z. B. Rückgang von Todesfällen und schweren Verläufen bei Masern) durchzuführen
- 2) Forschungsdaten – sind inhaltliche Daten, die im Rahmen der Arbeit anfallen oder spezifisch zum Zwecke der Programmentwicklung erhoben werden.

Zwei Beispiele für Forschungsdaten:

Crisis Text Line ist einer der weltweit bekanntesten Notrufdienste für psychische Erkrankungen. Die gemeinnützige Organisation nutzt Big Data und Künstliche Intelligenz (KI), um Menschen bei der Bewältigung von Traumata wie Selbstverletzung, emotionalem Missbrauch und Selbstmordgedanken zu helfen.

Nach eigenen Angaben verfügt Crisis Text Line über „den weltweit größten Datensatz zur psychischen Gesundheit“, basierend auf anonymisierten Daten, die die Hotline aus ihren Textkonversationen mit Menschen in Krisen gesammelt hat. Ich habe mich einige Male mit dem

Team von Crisis Text Line getroffen und war beeindruckt, wie sie Daten nutzen, um die Arbeit ihrer ehrenamtlichen Berater:innen zu unterstützen, z. B. bei der Triage von Hilfeanfragen oder indem automatisch Informationen über spezifische lokale Hilfsangebote bereitgestellt werden. Zunächst wurde diese Datensammlung kostenlos auf ihrer Website zur Verfügung gestellt. Später gründete Crisis Text Line einen gewinnorientierten Ableger, um zusätzliche Einnahmen aus den Daten zu erzielen – ein Schritt, über den auch andere gemeinnützige Organisationen im Datenbereich nachdenken. Diese Tochtergesellschaft von Crisis Text Line verwendete eine abgewandelte Version der Informationen, um Software für den Kundenservice zu entwickeln.

Im April 2021 erklärte das Unternehmen, dass die Erkenntnisse, die aus fast 200 Millionen Nachrichten von Crisis Text Line gewonnen wurden, geholfen hätten, KI-gestützte Kundenservice-Tools zu entwickeln, die Kundendienstmitarbeiter:innen bei Live-Chats mit aufgebracht Kunden unterstützen sollen. Alle Daten, die mit dem gewinnorientierten Unternehmen geteilt wurden, seien vollständig „anonymisiert“ worden, d. h. von allen Details befreit, die zur Identifizierung von Menschen verwendet werden könnten, die die Beratungsstelle in Not kontaktiert hatten. Rechtlich gesehen waren die Aktivitäten beider Organisationen nicht zu beanstanden. Dennoch löste diese Tatsache ein erhebliches negatives Medienecho aus. Die meisten Datenexpert:innen verurteilten die Vorgehensweise von datenethischem Standpunkt und Crisis Text Line stellte nachfolgend die Weitergabe von Daten ein.

Ein weiteres Beispiel im Bereich Forschungsdaten:

Eine Organisation, mit der ich seit Langem zusammenarbeite, ist das Schweizer Kinderhilfswerk terre des hommes. Eines der vielen Projekte terre des hommes ist ein digitales Gesundheitsprojekt in Burkina Faso. Im Rahmen des Projekts wurden im Laufe der Zeit Daten über mehr als zehn Millionen medizinische Konsultationen gesammelt. Der Schwerpunkt lag zunächst auf den Konsultationen, bei denen bestimmte Krankheiten wie Malaria diagnostiziert und behandelt wurden. Als das Team von terre des hommes jedoch immer mehr Daten sammelte, erkannten es, dass diese Daten ihm auch dabei helfen könnten, andere Probleme anzugehen. Das ist im Übrigen sehr verbreitet. Oft werden Daten als Nebenprodukt eines Prozesses gesammelt und erst später stellen die Organisationen fest, dass die Datensätze auch in anderen Zusammenhängen von Nutzen sein könnten.

In diesem Fall hatte terre des hommes die Vermutung, dass die Daten hilfreich sein könnten, um Diagnose- und Behandlungsfehler in diesen ländlichen Kliniken, in denen es keinen Arzt gibt, aufzudecken und korrespondierende Trainingsmaßnahmen zu entwickeln. Sie sahen in den Daten auch die Möglichkeit, das Gesundheitsministerium bei der Verbesserung des Infektionsschutzes und der Verwaltung von Medikamentenbeständen zu unterstützen.

In Zusammenarbeit mit der London School of Hygiene and Tropical Medicine bereinigte terre des hommes die Daten und entwickelte eine Reihe von Datenmodellen, um diese Themen anzugehen.

Terre des hommes hat diesen Datensatz Universitäten, wie z. B. der Universität Genf, zu akademischen Forschungszwecken zur Verfügung gestellt. Ein erklärtes Ziel bleibt außerdem, die Daten und die Analyse dem Gesundheitsministerium in Burkina Faso zu übergeben, dieses hat sich allerdings aus Kapazitätsgründen und angesichts der fragilen Lage im Land vorerst nicht realisieren lassen.

Doch erstmal genug zu Forschungsdaten. Mein letztes Beispiel bezieht sich auf die Nutzung von Projektdaten, also solchen Daten, die in erster Linie zur Erfolgsmessung und Berichterstattung gegenüber Förderern gesammelt werden. In meiner Praxis habe ich nur selten gesehen, dass diese Daten für andere Zwecke genutzt werden. Eine Möglichkeit besteht darin, die Daten der vielen verschiedenen gemeinnützigen Organisationen, die in einer Region tätig sind und dort teilweise denselben Menschen auf unterschiedliche oder sogar dieselbe Weise helfen, zusammenzuführen, um ein vollständigeres Bild der Lage in der Region zu erhalten. So hat z. B. der International Civil Society Centre (ICS Centre) in Berlin zahlreiche Hilfsorganisationen zusammengebracht, die alle in Bangladesch, Vietnam und anderen Ländern aktiv sind, um gemeinsam zu schauen, welche Fortschritte bei der Umsetzung der Sustainable Development Goals gemacht worden sind, wo die Lücken sind und welche Bevölkerungsgruppen systematisch nicht berücksichtigt werden. Keine der einzelnen Organisationen hat eine ausreichende Abdeckung in diesen Ländern, geographisch oder thematisch. Gemeinsam aber sind sie in der Lage, viel aussagekräftiger die Belange bestimmter Bevölkerungsgruppen gegenüber der Regierung und dem öffentlichen Sektor zu vertreten.

Was zeigen diese Beispiele:

- Daten, die gemeinnützige Organisationen sammeln, sind teilweise nirgendwo anders erhältlich und/oder ergänzen bestehende Datenquellen und füllen Lücken insbesondere in der Abdeckung von ausgegrenzten Gruppen.
- Diese Daten können besonders wertvoll sein, wenn generell nur wenige Daten zur Verfügung stehen, z. B. in Ländern mit niedrigem Einkommen oder wenn Statistiken nur mit großer Zeitverzögerung zur Verfügung stehen.
- Die Daten, die gemeinnützige Organisationen sammeln, können großen Nutzen für die Arbeit der Organisationen bringen, aber auch für andere externe Akteure – z. B. den Privatsektor, wie im ersten Beispiel, Wissenschaftler:innen und Regierungen, wie im Beispiel von terre des hommes, oder anderen Non-Profit-Organisationen. Letztlich bringen diese Daten Mehrwert für unsere Gesellschaften.

Diese Beispiele zeigen auch die Barrieren: Die Messlatte für die ethische Datennutzung liegt für gemeinnützige Organisationen zu Recht höher, sodass diese oft davor zurückschrecken, ein Öffnen ihrer Daten überhaupt in Betracht zu ziehen. Die Datennutzungskapazität ist generell gering, was das Angebot und die Nachfrage erschwert, und die Qualität der Daten lässt – zumindest im internationalen Bereich – oft zu wünschen übrig und ist nicht immer geeignet, um verlässliche Rückschlüsse zu ziehen.

Die bisherigen Beispiele zeigen das Potenzial der Wiederverwendung von Daten durch andere Akteure in neuem Zusammenhängen auf. Wie sieht es aus, wenn wir uns gezielte Datenpartnerschaften anschauen, d. h. Fälle, wo Organisationen zusammenkommen, um Daten gemeinsam zu erheben und/oder gemeinsam zu nutzen?

Die Datennutzung ist im gemeinnützigen Bereich erst in den Anfängen, dies trifft umso mehr zu, wenn wir uns Non-Profit-Datenpartnerschaften ansehen: Viele dieser Organisationen haben erhebliche Probleme, ihre eigene Datenentwicklung voranzutreiben und die Zusammenarbeit mit anderen, die ähnliche Probleme haben, verstärkt diese Schwierigkeiten noch. Dennoch ist dies

meiner Meinung nach ein Bereich mit enormem Potenzial. Tim O'Reilly, Gründer eines der einflussreichsten Technologieverlage, der z. B. den Begriff Open Source populär gemacht hat, hat einmal gesagt: Daten sind kein Öl – sie sind Sand. Individuelle Daten sind wertlos, sie haben nur in der Kombination von Millionen einen Wert. Das stimmt zwar heute nicht mehr ganz, in den Zeiten wo wir mit „small data“ wesentlich mehr machen können, als das noch vor einigen Jahren der Fall war. Grundsätzlich stimmt es aber, dass wir die neuen Möglichkeiten, die Datenerhebung und Datenanalysen uns bieten, nur dann beginnen auszuschöpfen, wenn wir Daten kombinieren, oder – mit anderen Worten – Daten mit anderen Daten anreichern – „data enrichment“.

Jurist:innen und Technolog:innen sind gemeinsam mit Repräsentant:innen der Zivilgesellschaft und Politik, dabei, neue Modelle des sicheren und ethischen Datenaustausches und der gemeinsamen Datennutzung zu entwickeln und zu testen, z. B. in Datenkooperativen. Als Organisationsform befähigen Datenkooperativen den Einzelnen, indem sie ihm die vollständige Kontrolle über die Qualität und Quantität der von ihm freigegebenen Daten geben. Der Nutzen dieser Modelle ist nicht nur, dass sie technisch und rechtlich die Grundlage für einen breiteren Datenaustausch schaffen. Sie zeigen auch neue Wege auf, wie Daten fair und transparent genutzt werden können – ein kritischer Beitrag, den die Zivilgesellschaft zur politischen Datendiskussion einbringen kann und muss.

Dass das selbst unter schwierigsten Bedingungen möglich ist, zeigt das Beispiel von Slumdweller's International, einer Organisation, die Slumbewohner:innen in 32 Ländern repräsentiert. Überall auf der Welt sammeln lokale Repräsentant:innen dieser Organisation systematisch Daten über die Bedingungen in ihren Siedlungen, um kritische Wissenslücken zu schließen. Sie müssen sich vorstellen, dass diese Siedlungen – da sie informell sind – oft nicht in öffentlichen Stadtplänen erscheinen. Sie werden ignoriert, sind nicht Teil der offiziellen Stadt, auch wenn sie seit mehr als 20 Jahren existieren und Zehntausende von Menschen beherbergen. Es gibt keine Polizeistationen, keinen offiziellen Zugang zum Wasser oder Elektrizität. Die eigenständigen Datensammlungen schaffen alternative Wissenssysteme, die sich im Besitz der Gemeinschaften befinden und die in Fällen wie dem Dharavi-Slum in Mumbai zu einer veränderten Stadtplanung und zu neuen Partnerschaften mit Versorgungsunternehmen geführt haben.

Angesichts des mangelnden Vertrauens der Slumbewohner:innen in die Bürokratie war diese eigenständige Datensammlung die einzige Möglichkeit, diese Daten verlässlich zu erheben. Ihr Wert allerdings entsteht durch die Nutzung und Teilung mit anderen.

Ich freue mich darauf, diese Chancen, die die vermehrte Öffnung von Daten der Zivilgesellschaft darstellt, in den nächsten Stunden tiefergehend mit Ihnen zu diskutieren.