

Stefan Thurner

*Big Data und die Folgen
Sind wir noch zu retten?*

Picus Verlag Wien

*Big Data und die Folgen
Sind wir noch zu retten?*

Wir leben im Zeitalter der Daten. Der vielen Daten. Die digitale Revolution hat eben erst begonnen. Daten türmen sich zu Terabytes, Zettabytes. Weil es immer mehr werden, erfinden wir immer neue Worte dafür. Worte wie Yottabyte. Ein Yottabyte sind 1.000.000.000.000.000 oder tausend Billiarden Gigabyte! Offenbar viel, aber groß? Bits und Bytes allein sagen nichts über Big Data.

Wann sind Daten »groß«?

Mitschriften und Aufzeichnungen existieren seit jeher, seit dem Beginn der Geschichte. Aber was der Mensch heute an Daten sammelt, erreicht unvorstellbare Ausmaße. Wir stehen kurz davor, praktisch alles mitzuschreiben, was wer wann wo und wie tut. Wir schreiben auf und erfassen, wer mit wem redet, wer an wen Geld überweist. Wir kennen das Einkaufsverhalten Einzelner und ganzer Bevölkerungsgruppen. Wir wissen, was wen interessiert,

wer was liest, wann welche Filme geschaut werden oder was welche Personen gerade besonders aufregt. Wir kennen die Netzwerke von allen, die miteinander befreundet sind – oder verfeindet –, wissen, wer wen wie beeinflusst, kennen alle Bewegungen aller, wissen, was wer produziert, woran sie arbeitet und in welchen Verhaltensmustern er gefangen ist.

In jedem Auto finden sich inzwischen mehrere SIM-Karten, die alles aufzeichnen und speichern – und bei manchen Modellen auch gleich an den Autohersteller oder an die Versicherung zurückmelden, was mit dem Auto getan wird und was in ihm und in seiner Umgebung passiert. Wir lassen Kühe Sensoren schlucken, die uns in Echtzeit sagen, was im Kuhmagen gerade vorgeht, und die warnen, wenn die Kuh krank zu werden droht. Wir beobachten Verrottungsprozesse in Mülltonnen und monitoren den CO₂-Ausstoß von Müllhalden. Es gibt Uhren, die jedes Mal, wenn das Herz schlägt, oder besonders wenn es nicht schlägt und einen Aussetzer hat, die Daten an Google und Co. schicken, die dann so das Gesundheitsprofil von Personen updaten.

Wir können den Standort und die Fracht jedes Schiffes zu jeder Zeit lokalisieren, überwachen live sämtliche Flugzeuge und Flugrouten, können zusehen, wo Bäume gefällt und Urwald in Palmölplantagen umgewandelt wird. GPS-Daten machen selbst

minimale Veränderungen messbar, wie die Bewegung von Bergen oder den millimeterweisen Anstieg des Meeresspiegels.

Durch den massenhaften Einsatz von Sensoren, die wir auf alles kleben, was sich auf diesem Planeten bewegt oder auch was sich nicht bewegt, durch die Nutzung von mobilen Endgeräten und des Internets bauen wir, bildlich gesprochen, an einer gigantischen Kopiermaschine; indem wir alles mitschreiben, was gerade vorgeht, fertigen wir eine *digitale Kopie* der Welt an – zu jedem Zeitpunkt:

Das ist groß

Daten sind nicht gleich Daten. Der Großteil dessen, was gemessen wird und an Big Data anfällt, ist wertloser Müll und ist nicht direkt verwendbar. Zurzeit sind Daten oft noch viel weniger einheitlich, qualitativvoll oder zugänglich, als man sich das wünschen würde. Die Qualität der gesammelten Informationen ist oft unbrauchbar, nicht nur weil die Formate nicht kompatibel sind, sondern weil Daten oft nicht ausgetauscht werden und in Datensilos nutz- und wertlos herumliegen. Daten werden oft erst wertvoll, wenn sie miteinander kombiniert werden. Zahlreiche Faktoren beeinflussen, welche Art von Daten in welcher

Weise gesammelt werden, also wie die Datensätze letztlich aussehen und wie verwendbar sie sind und wer sie verwenden darf. Zu solchen Faktoren zählen persönliche Interessen, gezielte wissenschaftliche Fragestellungen, das Vorwissen der Sammelnden – was man noch nicht weiß, wird möglicherweise nicht aufgezeichnet – oder auch ganz banal finanzielle und organisatorische Aspekte.

Selbst große Datensätze sind also heute oft noch unvollständig. Tritt plötzlich eine neue Frage auf oder will man eine neue Idee anhand vorhandener Daten überprüfen, passiert es nach wie vor, dass die entscheidenden Informationen zu genau dieser Frage fehlen. Aber die Situation verbessert sich rasend schnell. Hunderttausende Menschen arbeiten täglich daran, bessere und umfassendere Datensätze zu generieren. Um das Problem zu vermeiden, dass man eventuell etwas nicht aufzeichnet, das später relevant werden könnte, setzt sich mehr und mehr der Ansatz durch, einfach möglichst alles aufzuzeichnen. Da Speichermedien fast nichts mehr kosten, steht dem auch nichts mehr im Wege. Das eröffnet ungeahnte Möglichkeiten. Das verbleibende Problem liegt damit in der Datenauswertung, der Interpretation dieser Daten: Wie destilliere ich nutzbringend umsetzbares Wissen aus der vorhandenen Information? Diese Frage ist alles andere als gelöst. Sie bildet den Kern der Komplexitätsforschung.

Hier tut sich ein riesiges Problemfeld auf. Selbst wenn ich alle Informationen über ein (komplexes) System habe, bedeutet das noch lange nicht, dass ich das System verstehe. Es ist oft extrem schwierig, verwendbare und nützliche Erkenntnisse aus Daten zu generieren. Viele Firmen und Institutionen kennen das inzwischen. Man sammelt mit großem Aufwand Daten, aber was hat man dann davon? Einer der mühsamsten und langwierigsten Aspekte an der Arbeit mit Big Data ist es, gesammelte Daten so aufzubereiten, dass sie überhaupt sinnvoll bearbeitet werden können. Immer wieder sind Leute erstaunt, wenn sie erfahren, dass wir mitunter Jahre brauchen, um einzelne Datensätze überhaupt nutzbar zu machen. Aber auch diese Situation wird schnell besser.

Daten und Algorithmen alleine helfen leider noch gar nichts, um Systeme zu verstehen. Um Systeme so zu verstehen, dass man sie schlussendlich kontrollieren und managen kann, braucht man nach wie vor, auch im Zeitalter von Artificial Intelligence, gut durchdachte Fragestellungen, mit denen man an die Daten herangeht. Um große Daten wirklich gesellschaftlich wertvoll zu machen, braucht man neue Methoden, um der Komplexität der Systeme, die sie generieren, Herr zu werden: neue Mathematik, neue Statistik, neue Algorithmen. Der Flaschenhals ist weder die Datenmenge noch die Rechenleistung,

sondern das Fehlen von Methoden, die zum großen Teil erst entwickelt werden müssen. Das ist gemeint, wenn behauptet wird: »Big Data without Big Theory is big bullshit«. Ohne die richtigen Methoden ist Big Data oft kein Datenschatz, sondern eine Datenwüste, die oft mehr Verwirrung als Nutzen stiftet. In solchen Fällen erkennt man, dass man erst wieder wie der Ochs vorm Berg steht, wenn man es mit komplexen Systemen zu tun hat. Daten alleine generieren keine Erkenntnis; Maschinen nehmen einem das Verstehen nicht ab, auch wenn es zurzeit oft so dargestellt wird, als ob das bald geschehen würde. Das verstehen inzwischen auch immer mehr Institutionen, Organisationen und Firmen, die auf ihren Datenbergen sitzen und nicht wissen, wie sie mit ihnen umgehen sollen und wie sie daraus Nutzen ziehen können.

Aber auch hier: Zehntausende Menschen rund um den Globus entwickeln unaufhaltsam neue und bessere Methoden (dass sie dabei durchaus unterschiedliche Ziele verfolgen, werde ich gleich ausführen). Die digitale Kopie der Welt wird immer vollständiger.

Das digitale Zeitalter

Was können wir mit so einer »digitalen Kopie des Planeten« anfangen? Big Data ermöglicht es, uns im wahrsten Sinn des Wortes ein ganz neues Bild von den Dingen zu machen. Und wie jedes Mal, wenn sich Menschen ein neues Bild von der Welt machen, bricht ein neues Zeitalter an. Das aktuelle Zeitalter hat begonnen, viele spüren die ersten Anzeichen bereits – das digitale Zeitalter.

Dieses neue Zeitalter könnte ein großartiges sein! Ich bin wahrscheinlich zu optimistisch, aber ich sehe in meiner Arbeit und der vieler meiner Kollegen, was für fantastische Dinge wir mit Big Data tun können. Wir können vollkommen neuartiges Wissen gewinnen, neue Einsichten, die wirklich und unmittelbar zum Nutzen der Gesellschaft beitragen könnten! Wir bekommen die Möglichkeit, erstmals den Homo sapiens und seine Gesellschaften und Institutionen wirklich zu verstehen, mit einer Qualität, wie sie bisher nur in den »exakten« Naturwissenschaften möglich war.

Stellvertretend für viele Beispiele nenne ich ein paar neue Erkenntnisse, die aus unseren Forschungsgruppen am Complexity Science Hub Vienna hervorgegangen sind. Wir haben den Hub 2016 mit dem Ziel ins Leben gerufen, die Erforschung komplexer

Systeme in Österreich voranzutreiben, um Sinn und gesellschaftlichen Nutzen aus Daten zu generieren.

*Wie wir beginnen,
komplexe Systeme zu verstehen*

Was sind komplexe Systeme? Vereinfacht gesagt sind das all jene Systeme, bei denen das Ganze mehr ist als die Summe seiner Teile. Ein Ameisenhaufen ist ein komplexes System, oder auch der Finanzmarkt. Das Gesundheitssystem ist ein komplexes System, genauso wie das Weltklima oder das Internet. Ein einzelner Organismus ist ein komplexes System, wir sind umgeben von komplexen Systemen, wir bauen komplexe Systeme – die ganze Welt besteht aus miteinander verwobenen, interagierenden, aufeinander einwirkenden, sich (dadurch) ständig verändernden komplexen Systemen. Das klingt kompliziert – und ist es auch. Es ist eben so komplex-kompliziert, dass der Mensch mit Komplexität nicht besonders gut umgehen kann. Wir können komplexe Systeme bis heute nicht kontrollieren und managen. Komplexe Systeme übersteigen die mentale Kapazität eines Einzelnen.

Für ein biologisches Menschengehirn ist es schlicht nicht möglich, die oft riesigen Netzwerke von Abhän-

gigkeiten, gegenseitigen Beeinflussungen und dadurch hervorgerufenen Veränderungen sowohl der Teile als auch der ganzen Systeme bis zur letzten Konsequenz durchzudenken und dann daraus deduktiv Schlüsse zu ziehen. Geht nicht.

Um dennoch mit einer komplexen und vernetzten Welt umzugehen, die sie nicht beherrschen konnten, haben sich Menschen im Laufe ihrer Geschichte allerlei ausgedacht. Lange Zeit haben sie überall übernatürliche Kräfte am Werk gesehen, Geister, Götter und Dämonen, das Schicksal, die Sterne, Bachblüten und die Kraft des Rosenquarz. Mit der Erfindung der wissenschaftlichen Methode hat man erkannt, dass man überraschenderweise doch einige (wenige) Phänomene extrem gut verstehen und in weiterer Folge ungeheuren Nutzen aus diesen Erkenntnissen ziehen kann. Das funktioniert, indem man diese Systeme in immer kleinere Einzelteile zerlegt und diese Einzelteile immer genauer versteht. Bei »einfachen Systemen« kann man dann mit der Kenntnis der Einzelteile auch das große Ganze verstehen. Physik, Chemie und Molekularbiologie sind fantastische Beispiele, wie weit man mit dieser sogenannten »reduktionistischen« Vorgangsweise gekommen ist: Wir ernähren mehr als sieben Milliarden Menschen, alle können miteinander jederzeit kommunizieren, wir verstehen die Funktionsweise von Viren und haben ein Auto

auf den Mond gebracht. Egal wie kompliziert Physik und Chemie auch sein mögen, meist sind die Phänomene, die sie studieren, kompliziert, aber nicht komplex. Komplexe Systeme zeichnen sich dadurch aus, dass die Bauteile durch Wechselwirkungen zusammengehalten werden, die sich ständig ändern. Die Physik und die Chemie sind insoferne »einfach«, als sich die Wechselwirkungen und Kräfte nicht ändern, die Anziehungskraft zwischen Elementarteilchen ändert sich nicht von heute auf morgen, die Gravitation wird nicht stärker, auch wenn Ihre Waage Ihnen das suggeriert. Bei komplexen Systemen ändern sich die Wechselwirkungen ständig. Denken Sie an eine Gesellschaft, in der die Interaktionen zwischen Menschen natürlich jeden Tag vollkommen anders sind. Einige dieser Interaktionen sind wesentlich und können entscheidend dafür sein, wie das System in der Zukunft aussieht; die meisten Interaktionen sind nicht systemrelevant.

Wann immer man mit Systemen zu tun hat, deren Wechselwirkungen sich dynamisch ändern, kommt früher oder später der Moment, an dem sich komplexe Systeme völlig anderes verhalten, als man es erwarten würde. Das ist der Punkt, an dem wir sehr klar erkennen, dass eine simple Erklärung, dass stark vereinfachte Modelle, die in der Physik gut funktionieren (die die verändernden Wechselwirkungen

ignorieren), hier einfach nicht mehr ausreichen – eben zu simpel sind.

Das ändert sich nun, und ich traue mich zu sagen: dramatisch.

Was ist passiert?

Wir haben den Computer erfunden, den *game changer* im Umgang mit komplexen Systemen.

Computer erlauben uns heute nicht nur, Daten zu sammeln und diese zu verarbeiten, es wird erstmals möglich, sämtliche Interaktionen zwischen allen Bauteilen eines komplexen Systems zu erfassen und über die Zeit zu verfolgen. Kein Mensch könnte diese dynamischen Netzwerke überblicken, wie wir sie in der Medizin, der Finanzwelt oder der Wirtschaft vorfinden. Für den Computer mit entsprechenden Algorithmen ist das ein Kinderspiel. Damit wird es dann zum Beispiel erstmals möglich, das Risiko zu berechnen, dass Medizin-, Finanz- oder Wirtschaftsnetzwerke kollabieren.