# TechnoScope

by satw 1/20

# Big Data

## Der Umgang mit deinen Daten

www.satw.ch/technoscope

1956 brachte IBM die erste kommerzielle Festplatte auf den Markt. Sie konnte etwas mehr als 4 Megabytes speichern.

Zurzeit hat der grösste im Handel erhältliche USB-Stick eine Kapazität von 2 Terabytes oder 2 Billionen (Millionen Millionen) Megabytes.

Tera kommt vom griechischen Wort für Monster. Diese ungeheuerliche Masseinheit ist bereits zu klein: Google allein verarbeitet pro Tag rund 24 Petabyte.

In ein menschliches Gehirn passen geschätzte 2,5 Petabytes Daten.

### Mega, Giga, Tera, …

| | |
|---|---|
| **1 Bit** | (kleinste digitale Einheit) |
| **1 Byte** | = 8 Bits |
| **1 Kilobyte** | = 1000 Bytes |
| **1 Megabyte** | = $1000^2$ Bytes oder 1000 Kilobytes |
| **1 Gigabyte** | = $1000^3$ Bytes oder 1000 Megabytes |
| **1 Terabyte** | = $1000^4$ Bytes oder 1000 Gigabytes |
| **1 Petabyte** | = $1000^5$ Bytes oder 1000 Terabytes |
| **1 Exabyte** | = $1000^6$ Bytes oder 1000 Petabytes |
| **1 Zettabyte** | = $1000^7$ Bytes oder 1000 Exabytes |
| **1 Yottabyte** | = $1000^8$ Bytes oder 1000 Zettabytes |

Bis 2025 soll die weltweit gespeicherte Datenmenge auf 163 Zettabytes wachsen. Das sind rund zehnmal so viel wie 2016.

satw it's all about technology


**Impressum**
SATW Technoscope 01/20 | Februar 2020
www.satw.ch/technoscope
Konzept und Redaktion: Beatrice Huber | Ester Elices
Redaktionelle Mitarbeit: Christine D'Anna-Huber | Alexandra Rosakis
Grafik: Andy Braun
Bilder: Adobe Stock
Titelbild: Adobe Stock
Lektorat: Ars Linguae
Druck: Egger AG

**Gratisabonnement und Nachbestellungen**
SATW | St. Annagasse 18 | CH-8001 Zürich
technoscope@satw.ch | Tel +41 44 226 50 11

Technoscope 2/20 erscheint im Mai 2020 zum Thema «Food»

# Big Data

## Wissen aus den Daten holen

Smartphones, Kreditkarten, GPS oder Fitnesstracker: Wo wir gehen und stehen hinterlassen wir eine Datenspur. Aber nicht nur wir Menschen, auch immer mehr Sensoren in Geräte und Maschinen hängen am Internet und produzieren Daten: Bei einem selbstfahrenden Auto etwa sind es bis zu 4 Terabytes pro Tag. Ein Terabyte ist eine Eins mit 12 Nullen. Eine schier unvorstellbar grosse Zahl.

Big Data bedeutet zum einen also, dass immer mehr Daten aus ganz verschiedenen Quellen anfallen. Und diese können – auf Festplatten, Flashspeichern und immer häufiger auch in der Cloud – heute viel effizienter und billiger aufbewahrt werden als früher. Experten gehen davon aus, dass die gespeicherte Datenmenge sich in Zukunft etwa all drei Jahre weiter verdoppelt.

**Wertvolle Erkenntnisse aus Daten**

Das Besondere an Big Data sind aber nicht diese riesigen Mengen an Daten an sich. Sondern, dass immer leistungsfähigere Rechner aus diesen Datenbergen wertvolle Erkenntnisse gewinnen können. Die Technologie hat die Datenverarbeitung sozusagen gedopt: Blitzschnell durchforsten spezielle Software und lernfähige Algorithmen die gespeicherten Datenberge und suchen nach Mustern, Zusammenhängen und versteckten Gesetzmässigkeiten. Daraus lassen sich in unzähligen Bereichen – von der Forschung und der Medizin, über das Finanzwesen, das Marketing bis hin zur Landwirtschaft – Schlüsse ziehen und Trends ablesen: Wo droht ein Ernteausfall? Welche Patienten sprechen auf ein Medikament besonders gut an? Welche Kundengruppe dürfte das neue Produkt interessieren? Wie lassen sich Verkehrsströme effizienter managen?

**Jeden Tag gehen 294 Milliarden E-Mails um die Welt. Dazu kommen 5 Milliarden Suchabfragen, 65 Milliarden WhatsApp-Nachrichten und 500 Millionen Tweets.**

Big Data in 3 Minuten erklärt:
**https://www.youtube.com/watch?v=uH813u7_b0s**

## Woher stammen all die Daten für Big Data?

### Eine Spurensuche.

Jede Aktivität im Internet wird registriert: Welchen YouTube-Film hast du angeschaut, welche Produkte gekauft, welche Musik bevorzugst du, wie viele Likes hast du auf Instagram, wer sind deine Freunde, wie fit bist du (Smartwatch sei Dank), wo warst du unterwegs (Mietvelo sei Dank)? Daraus lassen sich Daten ableiten, die du vielleicht selbst nie angegeben hast, z. B. dein Alter, dein Bildungsstand, deine Hobbys. Solche personenbezogenen Daten sind für viele Unternehmen Gold wert, da sie ihre Dienstleistungen und Werbung danach ausrichten können.

### Der Handel mit den Daten

Entsprechend werden Daten rege verkauft und gekauft. Viele Nutzende sehen kein Problem darin, weil sie «nichts zu verbergen haben». Aber Daten können auch missbraucht werden oder man kann aufgrund seiner Daten benachteiligt– z. B. aufgrund ungesunder Gewohnheiten durch eine höhere Krankenkassenprämie– oder sogar Opfer eines Identitätsklaus werden. In China ist die Überwachung durch den Staat sowie die Belohnung oder Bestrafung der Bürgerinnen und Bürger anhand personenbezogener Daten bereits Realität. In der Schweiz regelt ein Bundesgesetz den Datenschutz. Dieses Gesetz wird zurzeit überarbeitet, um den neuen Realitäten im Internet gerecht zu werden.

### Offene Daten als Chance

Big Data ist aber viel mehr als personenbezogene Daten. So genannte Open Data sind frei verfügbar und dürfen von allen genutzt und weiterverarbeitet werden. Dabei handelt es sich z. B. um Wissensportale wie Wikipedia oder Statistiken, wie sie vom Bundesamt für Statistik veröffentlicht werden, Verkehrsinformationen, Daten über aktuelle Umweltereignisse, aber auch Software und künstlerische Arbeiten wie Bilder oder Videos. Ein Beispiel ist die Open-Data-Plattform öV Schweiz **https://opentransportdata.swiss/de/**, die unter anderem Fahrplan- und Haltestellendaten zur Verfügung stellt, sodass damit Apps oder andere Produkte entwickelt oder Statistiken erstellt werden können.

# Wo bin ich?


**Mobilität der Zukunft**
Daten tragen zu einem besseren Verkehrsfluss bei

Stell dir vor, du musst die schnellste ÖV-Verbindung heraussuchen, um noch rechtzeitig zu einem Termin zu kommen, und du hast nur das gedruckte offizielle Kursbuch der Schweiz – eine echte Herausforderung. Zum Glück hat sich seit 1905, als das erste Kursbuch publiziert wurde, viel getan. Verschiedene Apps nehmen einem nicht nur die mühsame Suche nach der besten Verbindung ab, sondern sie verkaufen gleich das passende Billett und verknüpfen die ÖV-Verbindung mit aktuellen Daten wie Verspätungen oder Umleitungen.

**Besserer Verkehrsfluss dank Live-Verbindungen**

Die Mobilität der Zukunft, wie sie z. B. die Stadt Zürich plant, möchte noch weiter gehen und alle Verkehrsteilnehmer – ÖV, Car Sharing, Taxi, Velo-Verleih – live miteinander verbinden **www.stadt-zuerich.ch/vbz/de/index/mobilitaet-der-zukunft**. Auch private Fahrzeuge sollen untereinander und mit der Verkehrsinfrastruktur kommunizieren. Unter anderem soll dies zu einem besseren Verkehrsfluss beitragen. Mittels Google Maps kann man sich bereits über Staus in Echtzeit informieren. Das funktioniert, indem Google die über GPS ermittelte Position von Smartphones erfasst und anhand deren Anzahl und Lokalisierung die Verkehrssituation auf der Landkarte wiedergibt.

**Daten als Wegweiser**

Apropos Landkarte: Dass wir heutzutage so leicht den Weg von A nach B finden können, haben wir Big Data zu verdanken. Google z.B. arbeitet mit unzähligen Partnern zusammen, die detaillierte Koordinaten liefern, um geografische Gegebenheiten präzise und aktuell darzustellen. Street View liefert nebenbei Daten zur Verbesserung der Karte, indem abgebildete Strassenschilder gelesen und mit der Karte abgeglichen werden. Satellitendaten werden eingesetzt, um mögliche geologische oder bauliche Veränderungen zu entdecken. Interessant für uns alle ist zudem, dass wir selbst zum Nutzwert der Karte beitragen können, indem wir Fotos oder Bewertungen abgeben. Algorithmen sowie Mitarbeitende aus Fleisch und Blut verwalten, vergleichen und vernetzen diese mehrschichtigen Datensätze, um so die Karten laufend zu verbessern.

# Big Data by you

Viele wissenschaftliche Fragen bedürfen einer grossen Menge an aufwendig gesammelten Datensätzen. Dabei handelt es sich nicht unbedingt um kompliziert im Labor gemessene Parameter, sondern auch um Daten, die wir alle in unserer unmittelbaren Umgebung festhalten können. Citizen Science – Forschung, die von und schlussendlich für Bürgerinnen und Bürger getrieben wird – schliesst wissenschaftsinteressierte Laien zusammen, um schnell solche Daten zu erheben.

**Alle helfen allen**

Üblicherweise initiieren und leiten Forschende die Projekte, werten die Daten aus und machen sie der Öffentlichkeit zugänglich. Je nach Projekt kann diese in die Datenerfassung und auch bereits in der Planung von Projekten und danach in die Analyse der Resultate mit einbezogen werden. So haben die Universität Zürich und die ETH Zürich das Kompetenzzentrum Citizen Science **https://citizenscience.ch/** ins Leben gerufen.

Die Palette an Forschungsfragen, die mithilfe von Freiwilligen untersucht werden können, ist gross. Auf der Webseite **www.schweiz-forscht.ch** werden verschiedene Projekte vorgestellt, von der Verbreitung des Alpensalamanders über die verschiedenen Schweizer Dialekte bis zur Quantenfehlerkorrektur. Alle Interessierten können sich melden und ihren Teil zur Forschung beitragen, indem sie beobachten, sammeln, fotografieren oder gar ein Game spielen. Über die App **allyscience.ch** zum Beispiel können Heuschnupfengeplagte ihre aktuellen Beschwerden dokumentieren und somit bei der Entwicklung künftiger Frühwarnsysteme und Therapien für Pollenallergiker mithelfen.

**Qualität der Daten muss stimmen**

Eine Schwierigkeit bei dieser variablen Erfassung von Daten ist die Qualitätskontrolle. Besonders bei einer sehr komplexen Datenerfassung kann es leicht passieren, dass Daten unabsichtlich verfälscht werden. Abgesehen von der unmittelbaren Kontrolle über die Community werden Daten deshalb auch durch Algorithmen geprüft und falsche Eingaben herausgefiltert.

# Viel vom Gleichen in der Blase

Was ist Fake und was ist News? Früher führten angesehene Zeitungen durch den Informationsdschungel, erklärten die Nachrichtenlage und ordneten sie ein. Im Internetzeitalter ist der Informationsdschungel noch dichter, sind die Nachrichtenquellen noch zahlreicher geworden, Ursprung und Wahrheitsgehalt von News noch weniger überschaubar. Gleichzeitig haben die klassischen Medien ihre Bedeutung als «Türhüter» eingebüsst. Insbesondere junge Menschen haben ihnen den Rücken gekehrt. Sie informieren sich lieber über Newsaggregatoren wie Reddit, über Blogs und Online-Foren. In der Schweiz nutzen laut der James Studie 2018 rund 51 Prozent aller Jugendlichen soziale Netzwerke täglich als Infoquelle.


**Personalisierte News**

Die Newsaggregatoren basieren auf Algorithmen. Diese tun das, was früher eine Redaktion machte: Sie wählen die Informationen aus, die sie der Leserschaft präsentieren. Doch im Zeitalter von Big Data ist diese Auswahl genau auf den einzelnen Kunden zugeschnitten. Aus allerlei Daten wie Herkunft, Alter und Geschlecht, kombiniert mit den Suchanfragen, Seitenaufrufen und dem Surfverhalten eines Users, erstellen die Algorithmen dessen Persönlichkeitsprofil und leiten daraus ab, welche Nachrichten ihn vermutlich am meisten interessieren oder unterhalten. Alles andere bekommt er gar nicht erst zu Gesicht. Das wird ihm auf die Länge den Eindruck geben, die Welt bestehe aus lauter Gleichgesinnten. Wer nie durch eine andere Meinung herausgefordert oder zum Nachdenken und Dazulernen angeregt wird, der ist in der Filterblase gefangen.

**Raus aus der Blase**

Das beste Gegengift ist Vielfalt: Also herausfinden, wie anderswo über ein Thema berichtet wird – vielleicht zur Abwechslung auch mal in den traditionellen Medien. Schauen, wer hinter einer News steht, wie seriös die Quelle ist und an wen sich die Nachricht richtet. Und sich dann selber eine Meinung bilden.

Weiterführende Information:
**https://www.jugendundmedien.ch/themen/fake-news-manipulation.html**

# Schokolade, Nobelpreise und Fehlschlüsse

Intelligente Algorithmen können aus unterschiedlichen Datensätzen wertvolle Erkenntnisse herausfiltern. So etwa, warum es in der Schweiz so viele Nobelpreisträger gibt. Die Ursache dafür ist erstaunlicherweise darin zu suchen, dass in der Schweiz der Schokoladenkonsum pro Kopf höher ist als anderswo. Das zumindest scheint die Abbildung unten zu belegen. Oder stimmt das vielleicht gar nicht?


**Schokoladenkonsum in Kilogramm** pro Jahr und Einwohner

Nehmen wir ein anderes Beispiel: In der Schweiz sind seit dem Zweiten Weltkrieg sowohl die Anzahl Störche wie auch die Anzahl Babys pro Familie zurückgegangen. Beweist das, dass es eben doch die Störche sind, die die Babys bringen – so wie man früher glaubte? Nein, das tut es nicht. Dass heute weniger Störche auf unseren Dächern klappern ist zwar traurig. Aber es ist nicht die Ursache dafür, dass weniger Kinder geboren werden.

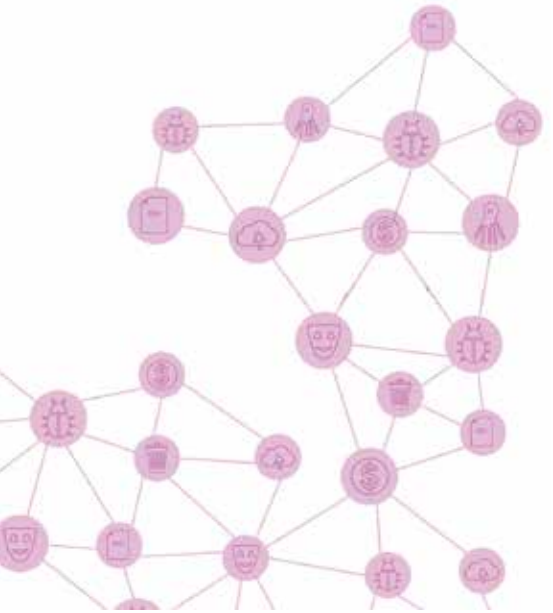
Der Harvard Student Tyler Vigen hat eine ganze Serie von solchen kuriosen Scheinkorrelationen aufgespürt. Viele davon sind zum totlachen. Aber sie belegen auch ein Problem, das Statistiker gut kennen. Es lautet: «Korrelation ist nicht Kausalität». Der Zusammenhang zwischen zwei Dingen heisst noch lange nicht, dass das eine das andere bewirkt.

Es gibt in der Informatik einen hübschen Merksatz: «Garbage in, garbage out» Auf Deutsch heisst das so viel wie: «Mist rein, Mist raus». Auf Big Data bezogen bedeutet es: Wie aussagekräftig das Resultat ist, das ein Algorithmus liefert, kommt nicht nur auf die Daten an, mit denen er gefüttert wird. Sondern auch darauf, ob er so programmiert ist, dass er die richtigen Parameter berechnet.

Kuriose Scheinkorrelationen:
**http://www.tylervigen.com/spurious-correlations**

# Big Data fürs Schulzimmer

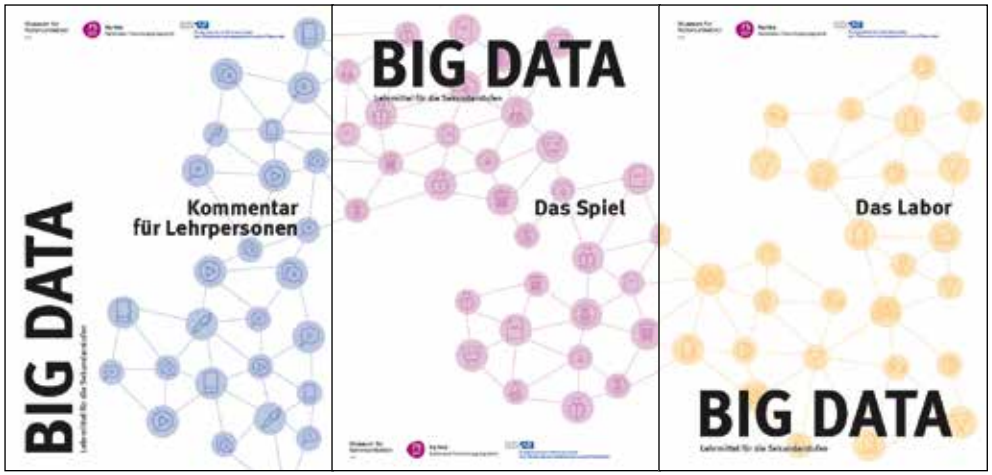
**Big Data ist ein komplexes Thema, aber zu wichtig, um es nicht zu verstehen. Das Nationale Forschungsprogramm NFP 75 «Big Data» und das Museum für Kommunikation haben sich zusammengetan und ein Lehrmittel zu Big Data für Schülerinnen und Schüler der Sekundarstufen I und II entwickelt.**



Mit dem Lehrmittel, das aufgeteilt ist in verschiedene Teile, wird Informatik «be-greifbar» mit einem starken Bezug zur Lebenswelt von Jugendlichen. Die Schülerinnen und Schüler wissen dank dem Lehrmittel, was Big Data ist, und setzen sich kritisch mit dem Thema auseinander. Sie schaffen so ein Bewusstsein für ihre eigene Betroffenheit. Denn auch Jugendliche sind nicht einfach Konsumierende. Es liegt an uns allen, die Zukunft zu gestalten.

Lehrmittel «Big Data»:
**www.mfk.ch/bigdata**






**«Big Data – Das Spiel»**
Als Einführung wird im ersten Teil eine spielerische Methode gewählt. Es gilt, ein Rätsel zu lösen, ähnlich einer Black Story. Ein US-amerikanisches Mädchen wird unverhofft schwanger. Sie sagt es niemandem und versucht, sich so unauffällig wie möglich zu verhalten. Auch der Vater weiss nichts davon. Die Supermarktkette «Target» schickt in dieser Zeit seiner Tochter Gutscheine für den Kauf von Schwangerschaftskleidung und Babyartikeln. Wie kann das sein? Am Anfang der Doppelstunde bekommen die Schülerinnen und Schüler eine rätselhafte Leitfrage gestellt. Im Spiel sammeln sie Informationen, die ihnen helfen, zusammen die Leitfrage zu beantworten und damit das Rätsel zu lösen.

**«Big Data – Das Labor»**
Im zweiten Teil des Lehrmittels werden sieben Aspekte von «Big Data» vertieft. Je nach Niveau und Vorwissen – aber auch Bedürfnis – der Klasse wird aus den sieben Aspekten ausgewählt.

- «Warum Geheimnisse?» – Privatsphäre
- «Nichts ist gratis!» – Wertvolle Daten
- «Was kriegen wir zu sehen?» – Filterblase
- «Denn sie wissen genau, was wir tun…» – Totalüberwachung
- «Wie geht Verhütung im Netz?» – Datenschutz
- «Mit Big Data die Welt retten?» – Chancen von Big Data
- «Gebt uns unsere Daten zurück!» – Open Data

**«Big Data – Der Trail»**
Der letzte Teil des Lehrmittels ist ein multimediales Lernspiel im Museum für Kommunikation in Bern. Inhaltlich ergänzt «Der Trail» die Kapitel von «Das Labor», ist aber auch ohne die vorgängige Bearbeitung des Themas im Unterricht spielbar. «Der Trail» spielt in der Zukunft, im Jahr 2080. Die Menschen haben die Kontrolle über ihre Daten gänzlich an den Grosskonzern Amathron abgegeben. Doch der Untergrund wehrt sich und will die Kontrolle über die Daten zurückgewinnen. Die Spielenden befinden sich in einem Trainingssimulator des Untergrunds und durchlaufen Trainingseinheiten, um sich für die Herausforderungen und Abenteuer in der digitalen Welt fit zu machen. Am Ende der Trainingseinheiten starten die Spielenden als Aktivistinnen und Aktivisten ihre erste Mission. Sie werden in die Vergangenheit – also unsere Gegenwart – geschickt, um dort Einfluss auf wichtige Entscheidungen zu nehmen.

# Forschen mit Big Data

Mit dem Sammeln grosser Mengen an Daten ist die Arbeit noch nicht getan. Die Daten müssen gespeichert, verwaltet, aufbereitet, ausgewertet, dargestellt und vernetzt werden. Während diese Aufgaben für kleine Datensätze überschaubar sind, werden sie für Big Data zu einer grossen Herausforderung. Forschung ist gefragt, zum Beispiel im Nationalen Forschungsprogramm 75 «Big Data», kurz NFP 75. In den NFP werden Forschungsprojekte durchgeführt, die einen Beitrag leisten, um wichtige Gegenwartsprobleme zu lösen oder eben Herausforderungen für die Schweiz zu meistern. Folgende Forschungsfelder werden im NFP 75 unter anderem beackert:

Ein Knackpunkt bei Big Data ist die Zeit, die Algorithmen für die Verarbeitung der Daten benötigen. Insbesondere im Bereich Maschinelles Lernen will man leistungsstärkere und schnellere Algorithmen entwickeln.
**http://bit.ly/nfp75_algorithmen**

---

Eine weitere Herausforderung ist die korrekte und nutzerfreundliche Speicherung riesiger Datenmengen. Ein Ansatz, um den Speicherbedarf zu reduzieren, ist die Echtzeit-Datenanalyse, bei der die eingehenden Daten nicht erst gespeichert, sondern direkt analysiert werden. Diese Datenstrom-Verarbeitungssystemen sollen auch für Fachleute anderer Disziplinen neben der Informatik bedienbar sein.
**http://bit.ly/nfp75_Datenstromanalytik**

---

Datenbankeinträge können mehrere hundert Spalten umfassen, was in aktuellen Programmiersprachen nicht berücksichtigt ist. Das erschwert die Nutzung von Datenbanken. Deshalb zielt ein weiteres Forschungsprojekt darauf ab, das Zusammenspiel von verschiedenen Programmiersprachen und Datenbanken zu verbessern.
**http://bit.ly/nfp75_Datenstromanalytik**

---

Big Data ist für sich selbst ein grosses Forschungsgebiet, bei dem viele Fragen noch ungelöst sind. Nicht nur Datenwissenschaftler, sondern auch Fachleute anderer Disziplinen sind gefragt, damit Big Data sein volles Potential entfalten kann.

# Studien- und Berufswahl

Graziella Dal Maso, Berufs-, Studien- und Laufbahnberatung St. Gallen

**Sehr geehrte Frau Dal Maso**
**Wir waren letzthin in einem Unternehmen, das BIG DATA-Analysen für Firmen anbietet. Ich war sehr fasziniert. Welche Studienrichtung müsste ich studieren für sowas?** (Emma, 13)

Liebe Emma
Was genau hat dich fasziniert? Ist es die Frage, mit welchen Mitteln so grosse Datenmengen gespeichert und strukturiert werden? Oder wie sie auf bestimmte Fragen hin ausgewertet und verfügbar gemacht werden können?

BIG DATA spielt in Zukunft in alle Studienrichtungen hinein, weil damit ganz neue Fragen und darauffolgende Untersuchungen ermöglicht werden. An erster Stelle steht wohl ein Studium wie Data Science, das an Unis und Fachhochschulen studiert werden kann. Im Masterstudium kannst du anschliessend deine Spezialisierung auswählen. Studieninhalte können beispielsweise die Verwaltung, Speicherung und die statistische Modellierung von riesigen Datenmengen oder die Entwicklung von Algorithmen wie auch die Umsetzung und Validierung maschinellen Lernens sein.

**Anwendungen für verschiedenste Wissenschaften kommen dazu**
An der ETH hast du mit einem Bachelor in Elektroingenieurwissenschaften, Informatik, Maschinenbau, Mathematik oder Physik Zugang zum Master Data Science. Bei der Nutzung von BIG DATA sind viele Disziplinen daran beteiligt. Für die Entwicklung von effizienten Algorithmen braucht es die Zusammenarbeit von Datenwissenschaftlerinnen und Absolventen verschiedenster Studiengebiete. Also z. B. die Mitwirkung von Verkehrsplanerinnen, wenn es um Fragen der Verkehrssteuerung geht, von Biomedizinern (Hirnforschung), Ökonominnen und Juristen (beispielsweise Regulierungen im internationalen Handel) oder Sprachwissenschaftlern (maschinelles Übersetzen). Fragestellungen, die BIG DATA betreffen, sind in vielen Fachgebieten denkbar. Einbezogen in die Forschung um BIG DATA sind auch Geistes- und Sozialwissenschaften, wenn es um die Risiken von BIG DATA geht (Überwachungs- und Beeinflussungsmöglichkeiten, Daten-Missbrauch).

Hochschulbesuchstage bieten dir Gelegenheit, Studierende zu befragen. Wenn die Qual der Wahl etwas ratlos macht, unterstützt dich die Studien- und Laufbahnberatung deines Kantons.

## Infos & Links

Information zu allen Studienrichtungen sowie Berufsbeschriebe findest du auf
**www.berufsberatung.ch**

Informationen zum Studium «Data Science» der ETH:
**https://inf.ethz.ch/de/studium/master/master-ds.html.**