



ARBEITSMATERIAL FÜR MIGRANTEN- ORGANISATIONEN

**Forum der Kulturen Stuttgart e. V.
House of Resources**

Marktplatz 4 · 70173 Stuttgart
Tel. 0711/248 48 08-26
Fax 0711/248 48 08-88

hor-stuttgart@forum-der-kulturen.de
house-of-resources-stuttgart.de

Modul 4 Daten

(aus der Seminarreihe Digitalverbunden 2020, Modul 4: Daten)¹

Inhalt

1. Erste Schritte im Bereich Daten
2. Datensicherheit und Datenschutz
3. Visualisierung von Daten mit Tableau

Heutzutage ist ein guter Umgang mit Daten in den Vereinen wichtiger denn je. Sie finden Daten immer und überall. Können Sie sich aber im Wirrwarr der Begriffe wie Big Data, Künstliche Intelligenz, Industrie 4.0 zurechtfinden? Erkennen Sie, wie sie Daten für Ihre Organisation gewinnbringend nutzen können? In diesem Modul dreht sich alles um das Thema „Daten“.

Im vorliegenden Modul erhalten Sie ein Verständnis für das Potenzial von Daten und die Datenanalyse für die Zivilgesellschaft und für Ihre Organisation. Sie erfahren, wie die Datenanalyt*innen arbeiten und was aus Sicht der Expert*innen für das Gelingen von Datenanalyseprojekten wichtig ist. Sie

¹ **Hintergrundinformation zum vorliegenden Text aus der Reihe *Digitalverbunden 2020*:**

„Digitalverbunden 2020“ ist eine fünfteilige Reihe (5 Module) und beantwortet Fragen rund um das Thema Digitalisierung. Vereinsaktive erhalten Informationen über digitale Lösungen, die in der Praxis in verschiedenen Arbeitsbereichen angewendet werden können.

Die digitale Seminarreihe *Digitalverbunden 2020* wurde von Berliner Verein youvo e. V. – der Engagement-Plattform für Kreative - und dem Bundesverband Netzwerke von Migrantenorganisationen (NeMO) e. V. ins Leben gerufen. Das Forum der Kulturen Stuttgart e. V. leitet in gegenseitiger Absprache die Inhalte aus den Workshops als Multiplikator im Rahmen der Bildungsarbeit an die Stuttgarter migrantischen Organisationen weiter mit dem Ziel, diese durch diese Wissensvermittlung in dem wichtigen Prozess der Digitalisierung zu unterstützen.

Der Inhalt dieser Arbeitsunterlage stammt von

Nina Hauser von CorrelAid e.V., einem Netzwerk von über 1400 ehrenamtlichen Datenwissenschaftler*innen, die ihre Kenntnisse für den guten Zweck einsetzen wollen. In gemeinsamen Projekten mit gemeinnützigen Organisationen nutzen wir Daten und Datenanalyse, um soziale Herausforderungen zu lösen. Weiterhin vernetzt CorrelAid gesellschaftlich-denkende Datenanalyt*innen und stößt einen Dialog mit der Zivilgesellschaft über das Potenzial von Daten an.

erhalten Tools, mit denen Sie die Prozesse in Ihrer Organisation identifizieren können, in denen Daten generiert werden. Außerdem verschaffen Sie sich einen Überblick über die (bereits vorhandene) Dateninfrastruktur in Ihrer Organisation. Dabei berücksichtigen Sie bei der Arbeit die wichtigen Prinzipien der *Datensicherheit und des Datenschutzes*. Im letzten Kapitel lernen Sie Tools kennen, mit denen Sie selbst bereits einfache, aber wirkungsvolle Analysen und Visualisierungen Ihrer Daten erstellen können.

1. Erste Schritte im Bereich der Daten

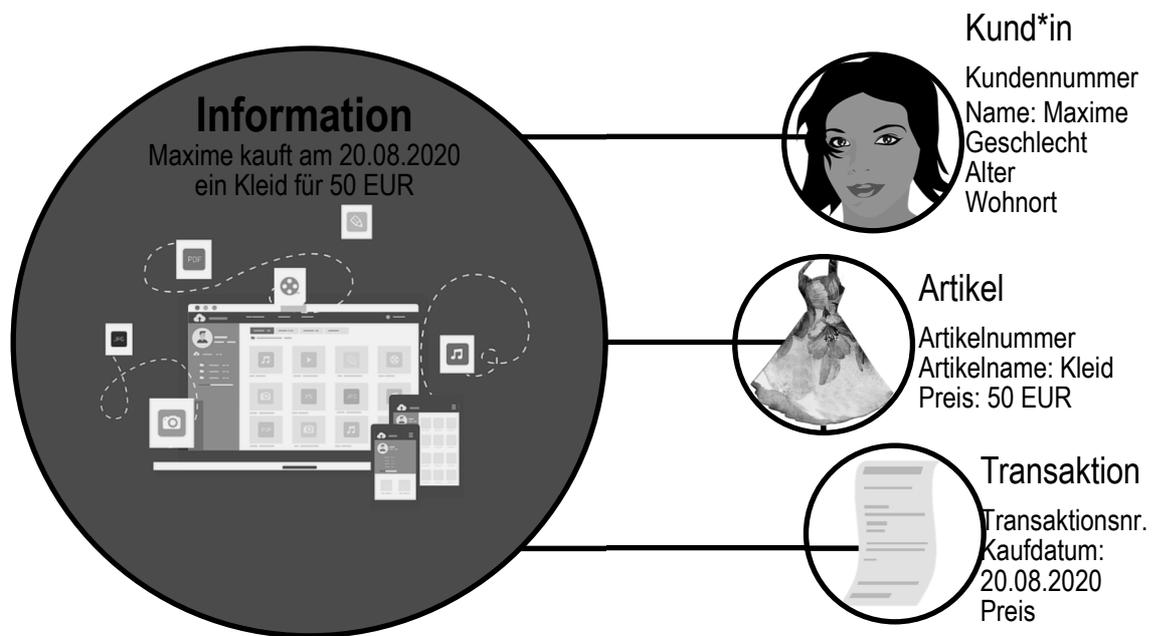
Daten vs. Information



Foto: Pixabay

Daten sind **Elemente**, die Informationen formal darstellen.
Erst im **Zusammenhang/Kontext** stellen Daten Informationen dar.
Je mehr Daten wir haben, umso mehr Informationen stehen uns zur Verfügung
(vorausgesetzt, dass die Daten gut angelegt sind).

Um zu veranschaulichen, was Daten und was Informationen sind, betrachten Sie bitte die folgende Grafik:



Grafik: CorrelAid e. V.

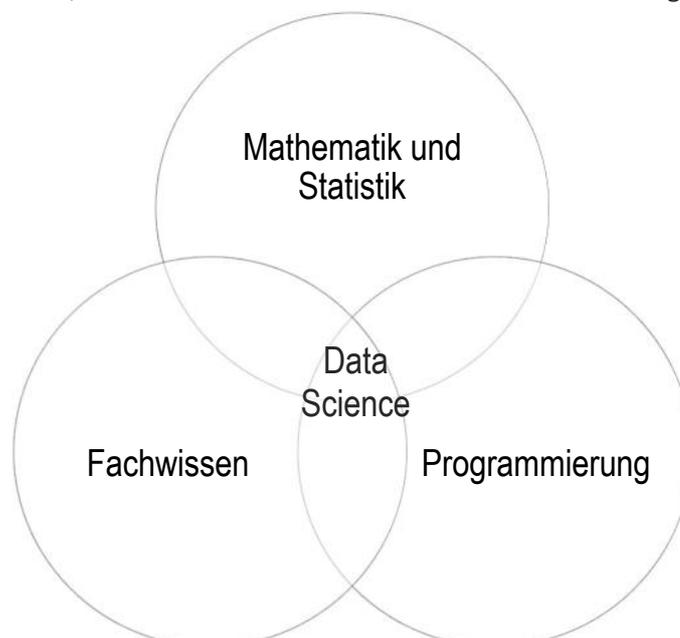
Der Name, das Alter, Geschlecht und der Wohnort der Kundin sowie der gekaufte Artikel und die Kaufdaten, also Kaufdatum und Preis, sind die Daten, die uns zur Verfügung stehen. Daraus ergibt sich die Information auf der linken Seite: Maxime kauft am 20.08.2020 ein Kleid für 50 Euro. Erst aus der gelungenen Verbindung der einzelnen Daten (Elemente) zur Kundin, zum Artikel und zu der erfolgten Transaktion (dem Kauf) ergibt sich die *Information*.



Foto: Pixabay

Data Science

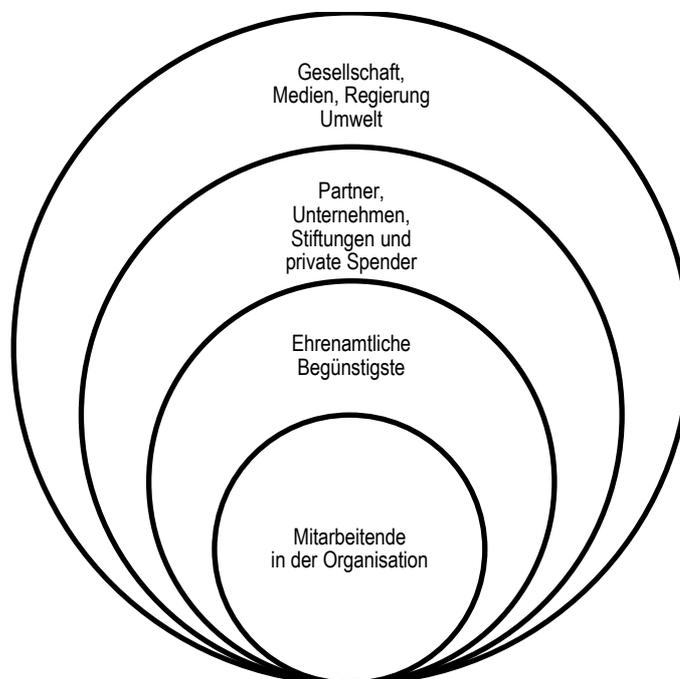
Data Science ermöglicht die Verarbeitung von Daten. Sie bezeichnet die Gewinnung von Erkenntnissen aus verschiedenen, gesammelten Daten. Data Science ist ein interdisziplinäres Fachgebiet mit vielen Anwendungsgebieten, das bedeutet, Forscher und Mitarbeitende unterschiedlichster Fachrichtungen arbeiten im Bereich der Data Science. Data Scientists, so bezeichnet man die dazugehörige Berufsgruppe, fragen sich, wie wir auf Basis von Daten bessere Entscheidungen treffen können.



Grafik: CorrelAid e. V.

Netzwerk

Migrantische Organisationen arbeiten mit und für ein Netzwerk. Das können unterschiedliche Akteur*innen und die Umwelt sein. Jede neue Herausforderung (z. B. Projekt, Veranstaltung, Bildungsmaßnahme etc.), die Ihre Organisation annimmt, bringt neue Kontakte. Zwischen Ihnen als Vertreter*in der Organisation und weiteren Akteur*innen aus der Gesellschaft, Kultur, Wirtschaft und Politik entstehen *Beziehungen* und *Abhängigkeiten*. Diese sind unterschiedlich intensiv. Deshalb ist der Level ausgetauschter Informationen – die Menge ausgetauschter Daten – unterschiedlich groß.



Grafik: CorrelAid e. V.

Alle Daten, die durch die Kontakte und unterschiedlich stark ausgeprägte Beziehungen entstehen, sind *Potenziale für die Organisation* und können für die Verbesserung der Prozesse und IT-Programme genutzt werden! Dabei können für jede Zielgruppe folgende Fragestellungen hilfreich sein:

Mitarbeitende in der Organisation:

- Welche Herausforderungen beschäftigen unsere Mitarbeitenden?
- Wie können wir unsere Prozesse verbessern und Qualität sichern?

Ehrenamtliche und Begünstigte:

- Was zeichnet unsere Begünstigten und Ehrenamtlichen aus?
- Was benötigen unsere Begünstigten?
- Wie können wir Ehrenamtliche besser einbinden?

*Private Spender*innen und Stiftungen:*

- Was zeichnet unsere Spender*innen aus?
- Wie können wir potenzielle Spender*innen/Stiftungen erreichen?

Partner und Unternehmen:

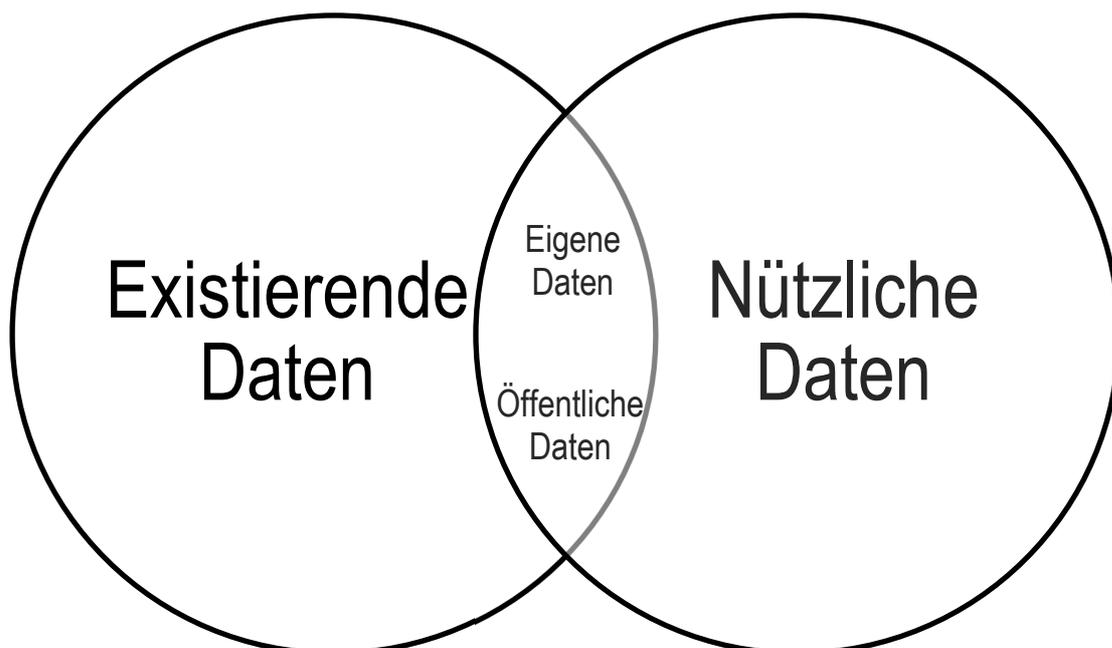
- Wie hat die Zusammenarbeit uns bei der Erreichung unserer Ziele vorangebracht?
- Wie können wir Lieferketten effizient gestalten?

Gesellschaft, Medien, Regierung und Umwelt:

- Wie sieht die Gesellschaft unseren Aktivitätsbereich?
- Wie berichten Medien über uns/unseren Aktivitätsbereich?
- Welche gesetzlichen Regelungen gelten für uns?
- Wie wird über unsere Themenstellung im Parlament diskutiert?
- Welche Informationen haben wir über unsere gesellschaftliche Herausforderung?

Bitte beachten Sie, dass Daten nichts bringen, wenn man sie nicht nutzt!

Es gibt also Daten, die bereits existieren und solche, die für die Organisation nützlich sind:



Besonders nützlich sind Daten, die bei operativen Vorgängen (also der Arbeit, die gemacht wird) entstehen. Das können z. B. Formulare, E-Mails, Chats etc. sein.

Man unterscheidet unterschiedliche Datentypen.

Erhebungsdaten werden gezielt erhoben werden, beispielsweise durch Umfragen, Registrierungen, Beobachtungen.

Außerdem gibt es **Operationsdaten**, die durch Anwendungen gesammelt werden. Dazu gehören Internetseiten (Cookies), Social Media, Portale, E-Mails, Chats etc.

Externe Daten werden recherchiert. Darunter fallen öffentliche Daten, Daten aus anderen Organisationen oder Studien.

Jeder von diesen Datentypen hat Vorteile und Nachteile:

Erhebungsdaten	Operationsdaten	Externe Daten
+	+	+
Auswahlfreiheit Flexibilität	Preiswertigkeit Unverfälschtheit	Preiswertigkeit Neutralität
-	-	-
Kosten der Erhebung Zeitintensivität	Zugriff Privatsphäre und DSGVO	Generalität Intransparenz

Grafik: CorrelAid e. V.

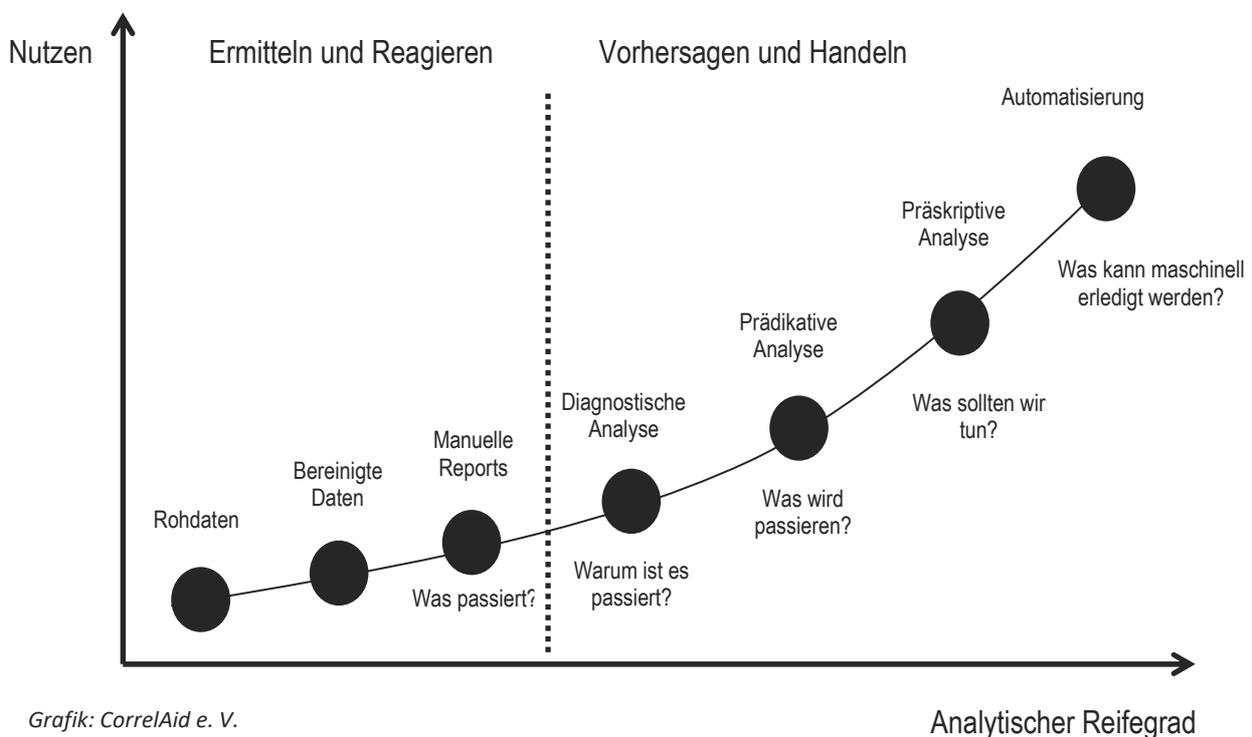
Bitte beachten Sie, dass der Wert von Daten nicht nur in der Modellierung selbst, sondern in der Nutzung von Resultaten liegt! D. h. dass Sie eher darüber nachdenken sollen, wie Sie die Daten, die Ihnen zur Verfügung stehen, nutzen wollen und nicht, wie gut die Methoden der Datensammlung bzw. Datenerhebung sind.

Fragen Sie sich:

- Wie würden Sie Daten definieren?
- Aus welchen Prozessen in Ihren Organisationen ergeben sich Daten, die Sie nutzen können?
- Auf welche Datentypen können Sie zugreifen?

Ebenen, die mit Datenanalysen erreicht werden können

Haben Sie sich schon gefragt, wo Ihre Organisation in Bezug auf Daten steht? Bevor wir uns näher damit beschäftigen, schauen wir uns näher die folgende Grafik an. Sie zeigt Ebenen an, die mit Datenanalysen erreicht werden können. Dabei werden das *Nutzen* und der *analytische Reifegrad* berücksichtigt. Je größer (weiter fortgeschrittener) der analytische Reifegrad, umso größer das Nutzen für eine Organisation.

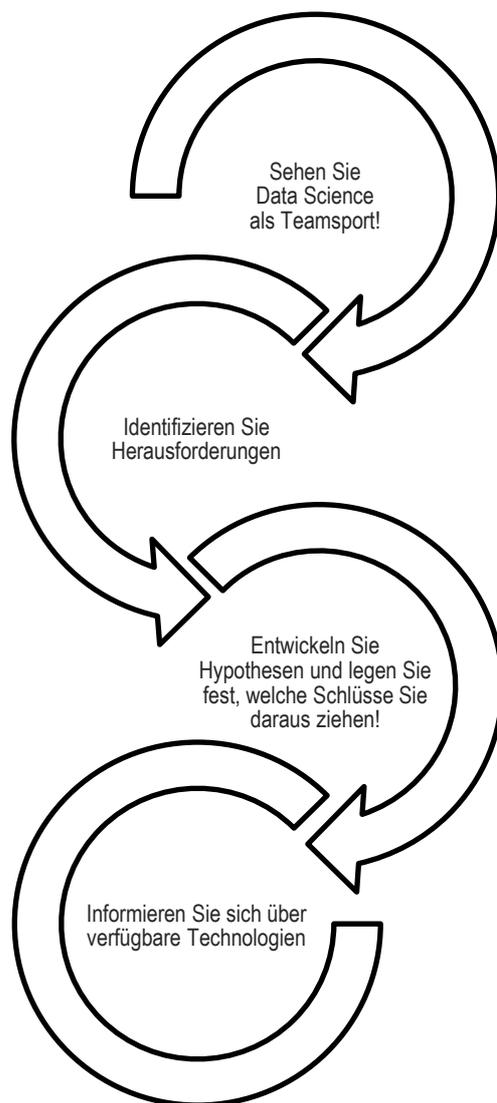


Grafik: CorrelAid e. V.

Die meisten Organisationen sehen sich im hellgrünen Bereich. Sie schaffen gerade noch einen manuellen Report (z. B. schriftlichen und rechnerischen Verwendungsnachweis) an die öffentlichen Förderer zu schreiben. In dieser Stufe wissen Sie, was in Ihrer Organisation passiert. Sie ermitteln und reagieren.

Spannend wird es, wenn Sie und Ihre Organisation Ihr Tun vorhersagen und handeln können! Also wissen, warum bestimmte Dinge passieren, was passieren wird und wenn Sie sagen können, was Sie und Ihre Organisation tun sollten. Die höchste Stufe ist erreicht, wenn Sie Prozesse automatisieren können. Automatisierung geschieht dann, wenn Sie wissen, was eine Maschine für Ihre Organisation tun kann und dies umsetzen.

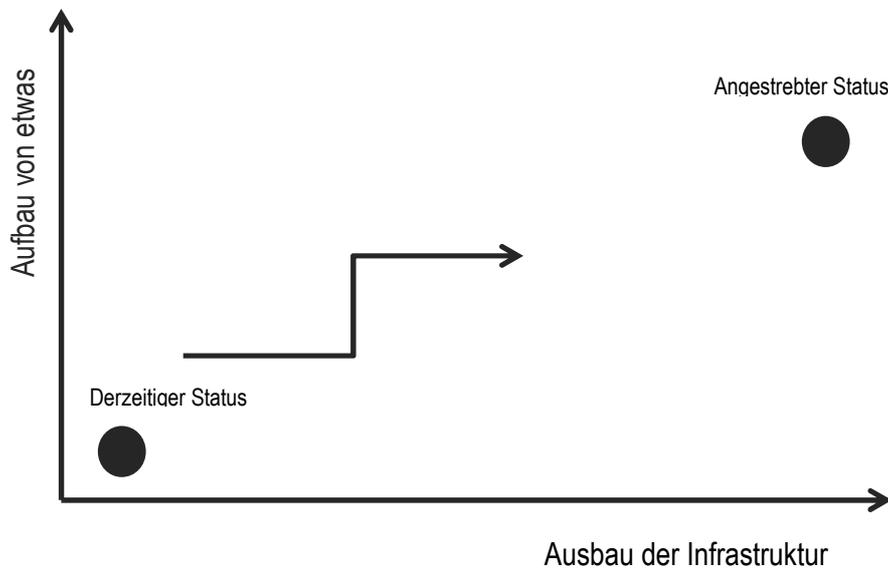
Sie und die Mitglieder Ihrer Organisation sollen aktiv an der Gestaltung der Datenstrategie mitwirken. Bilden Sie ein Datenprojekt-Team! Ein gutes Datenprojekt-Team besteht aus Data Engineers, Data Scientists und Organisationsvertreter*innen, die sich mit der Strategie und Organisation bestens auskennen.



Beachten Sie dabei die folgenden Stufen:

1. Einblicke und Vorhersagen
2. Teilautomatisierung
3. Vollautomatisierung von Prozessen

Wenn Sie ein neues Projekt anfangen, erklimmen Sie von der Vision bis zur Umsetzung verschiedene Stufen:



Grafik: CorrelAid e. V.

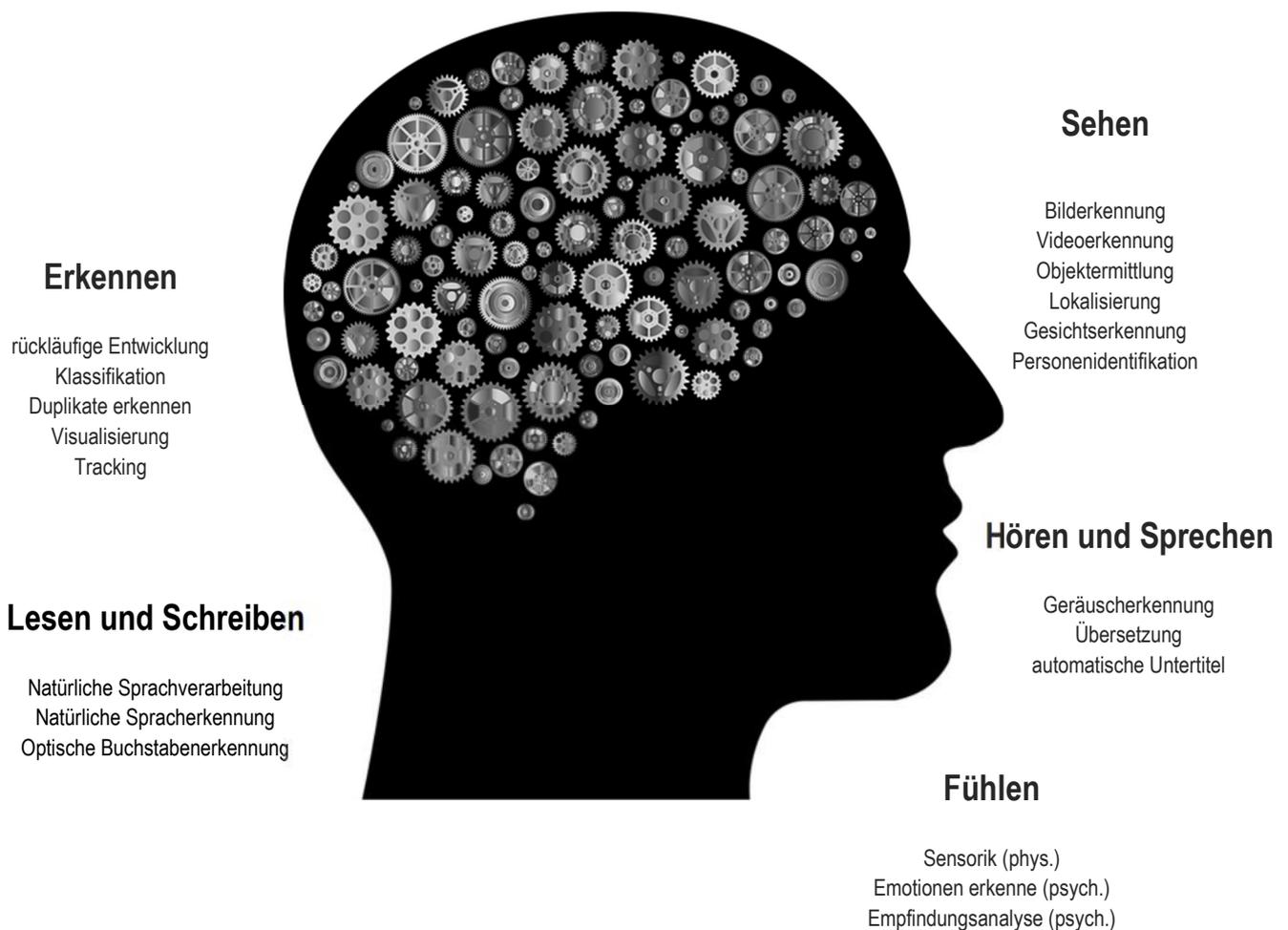
Dabei gilt: Entweder hat man ein Ziel, das man erreichen möchte, oder man hat Daten und versucht Neues zu entwickeln!

Fragen Sie sich:

- Wo steht dabei meine Organisation in Bezug auf die Datenerhebung?
- Haben wir Potenzial für Automatisierungsprozesse?
- Welches Ziel können wir uns setzen, das sich mit Data Science erreichen lässt?

Technologien

Alle Technologien, die entwickelt werden, verfolgen das Ziel menschliche Fähigkeiten nachzubilden:



Grafik: Pixabay

Mit diesen Technologien können Sie die Prozesse in Ihrer Organisation neu denken!

Wirkung der Daten

In Deutschland wurde das Potenzial der Daten im NGO-Sektor noch nicht erkannt. Zudem ist vielen unklar, wie man mit Daten umgeht. Die Wirkung und die Effekte der eigenen Arbeit werden in 19 von 100 Organisationen nicht gemessen, bzw. 33 von 100 Organisationen messen die Wirkung nur qualitativ. 29 von 100 messen ihre Arbeit auch quantitativ und nur 19 von 100 nutzen feste Indikatoren zur Messung ihrer Arbeit. Meistens konzentrieren sich die Organisationen auf die Output-Ebene, indem sie Aktivitäten oder Leistungen nachweisen. Die Ebene des Outcomes oder Impacts, also die Wirkung auf Ebene der Zielgruppe oder die Effekte auf gesellschaftlicher Ebene werden gar nicht berücksichtigt. Dabei sind dies die Wirkungsebenen, die Organisationen eigentlich erreichen wollen.

Wirkungskette:



Die Stufen der Wirkung:



Quelle: Von Stefan Schultze - Kursbuch Wirkung, Praxisratgeber für Alle, die Gutes besser tun wollen, CC BY-SA 3.0 de, <https://commons.wikimedia.org/w/index.php?curid=64336227>

Datenreifegrad

Wie Sie am besten Daten nutzen können, ergibt sich aus dem heutigen Status Ihrer Organisation und der Organisationszielen. Um herauszufinden, wo Sie als Organisation stehen, sollen Sie und Ihre Organisation sich mit den folgenden Themen befassen:

- Ermittlung des Datenreifegrads
- Ermittlung des Organisationsreifegrads
- Dokumentation der Datenlandschaft
- Erarbeitung von Prozessdarstellungen

Weiterhin sollen Sie und Ihre Organisation sich damit auseinandersetzen, wo Sie hinwollen:

- Erarbeitung der Wirkungslogik und Datenstrategie
- Schaffung notwendiger Infrastruktur
- Betrachtung verfügbarer Technologien
- Implementation passender Technologien

Mit dem folgenden *Modellierungsframework für NGOs (I)* finden Sie heraus, wo Ihre Organisation im Rahmen des Datenreifegrades steht – zunächst in Bezug auf **Technische Voraussetzungen**:

Thema	Noch nicht ausreichend	Ausreichend	Fortgeschritten	Ausgezeichnet
Zugang	Nur in der Anwendung verfügbar, in der die Daten erhoben werden	Daten können in einfachen Formaten (z. B. PDF) extrahiert werden	Daten können in maschinell lesbaren Formaten extrahiert werden (CSV, JSON, XML, Datenbankextrakt)	Alle maschinell lesbaren Formate sind verfügbar. Zudem ist die Datenextraktion über eine API* möglich
Speicherung	Papier	PDF/Bilder	Textdateien (CSV, XML, ...)	Datenbanken
Integration	Daten sind in individuellen Anwendungen verfügbar	Daten werden extrahiert und anschließend manuell importiert.	Zentrales Datawarehouse mit automatisierter Aggregation und Verknüpfung	Zusätzlich werden externe Daten abgerufen und integriert
Relevanz und Vollständigkeit	Zwischen tatsächlicher Datenlage und Datenbedarf liegt eine hohe Diskrepanz vor: Zum Beispiel fehlt bei der Analyse von Schulabbruchsquoten die Messgröße der Schulabbrecher*innen	Einige Datenpunkte sind relevant aber zentrale Datendimensionen fehlen: Beispielsweise fehlen Anwesenheitslisten der betroffenen Schüler*innen	Datenpunkte geben bereits Auskunft über zentrale Fragen der Programmlogik, aber für eine qualifizierte Problemlösung fehlen noch Informationen: Zum Beispiel gibt es keine Informationen über Verhaltensauffälligkeiten von Schüler*innen	Daten geben vollständig Auskunft über relevante Fragestellungen und können gezielt eingesetzt werden, um Lösungsvorschläge zu entwickeln, Programme zu überwachen und diese zu evaluieren
Qualität	Es fehlen Datenpunkte (Individuen oder Einheiten)	Es fehlen Variablen (Spalten)	Es fehlen keine Daten, allerdings finden sich Fehler, die u. a. durch	Es fehlen keine Daten und Fehler bei der Dateneingabe werden

			mangelnde Einschränkungen bei der Dateneingabe entstehen	durch umfassende Maßnahmen beschränkt
Erhebungsfrequenz	Einmalig	Jährlich	Regelmäßig	Durchgängig
Granularität (wie kleinteilig die Daten sind)	Gesamtheitlich zusammengefasst (z. B. absolute Teilnehmerzahl)	Pro Gruppe zusammengefasst (z. B. pro Geschlecht)	Entitätenebene (z. B. pro Person)	Ereignisebene (z. B. eine Person über mehrere Veranstaltungen)
Historie	Alte Daten werden gelöscht	Historische Daten werden gespeichert, aber bei Updates überschrieben	Historische Daten werden gespeichert und neue Datenpunkte werden angefügt	Historische Daten werden gespeichert und Updates erfolgen über ein komplexes System aus Data Mapping Prozessen
Datenschutz	Nicht implementiert	teilweise implementiert (Daten stehen der Organisation ganzheitlich zur Verfügung)	Vollständig implementiert (Daten stehen pro Projekt zur Verfügung)	Automatisiert (Softwarelösungen erstellen User-Gruppen und stellen Datenanalysen unter Anwendung der gesetzlichen Vorschriften zur Verfügung)
Dokumentation	Keine Dokumentation zu Variablen	Codebuch, API-Dokumentation o. ä. mit Variablen und Kategorien	Codebuch, API-Dokumentation o. ä. mit Metadaten und Erhebungsmethoden	Codebuch, API-Dokumentation o. ä. mit Metadaten, Erhebungsmethoden, Annahmen, Ausschlusskriterien und potenziellen Verzerrungen

Quelle: CorrelAid e. V., Center for Data Science & Public Policy, Data Maturity Framework, University of Chicago

* API ist eine Programmierschnittstelle (kurz API), ein Programmteil, der von einem Softwaresystem anderen Programmen zur Anbindung an das System zur Verfügung gestellt wird. Im Gegensatz zu einer Binärschnittstelle definiert eine Programmierschnittstelle nur die Programmbindung auf Quelltext-Ebene (Quelle: Wikipedia)

Jedoch nicht nur technische Voraussetzungen, sondern auch (oder vor allem) die **Organisationsstruktur** spielt dabei eine bedeutende Rolle. Mit dem folgenden *Modellierungsframework für NGOs (II)* erfahren Sie mehr über den datenstrategischen Reifegrad Ihrer Organisation:

Bereich	Noch nicht ausreichend	Ausreichend	Fortgeschritten	Ausgezeichnet
Organisation	Arbeitnehmer*innen wissen, dass Daten existieren, aber noch nicht, was diese enthalten und wozu diese genutzt werden können	Daten werden von einzelnen Arbeitnehmer*innen bedarfsorientiert genutzt, um Förderberichte zu erstellen	Daten werden bereits von einigen Abteilungen gezielt genutzt, um Qualitätsmanagement zu betreiben, programmatische Entscheidungen zu treffen, Berichtsicherungen	Die Organisation verfügt über eine gesamtheitliche Datenstrategie, die sowohl in der Unternehmenskultur verankert ist als auch als Basis für sämtliche Managemententscheidungen dient

			von Fördern gerecht zu werden und Skalierungen zu planen	
Programmatisches Personal	Arbeitnehmer*innen, die für die Erhebung von Daten zuständig sind, sehen diese als notwendige Verpflichtung, der sie nur unregelmäßig nachkommen	Arbeitnehmer*innen, die für die Erhebung von Daten zuständig sind, erfüllen diese Anforderungen pflichtbewusst	Arbeitnehmer*innen, die für die Erhebung von Daten zuständig sind, erfüllen diese Anforderungen pflichtbewusst und erhalten in fixierten Zeitabständen Analysen, die für ihre programmatische Arbeit relevant sind	Arbeitnehmer*innen, die für die Erhebung von Daten zuständig sind, erheben diese in Echtzeit, erhalten durch Analysetools direkte Einblicke in ihre programmatische Arbeit, aus denen sie Handlungsempfehlungen ableiten können, und machen eigenständig Vorschläge zur Datenstrategie
Analytisches/technisches Personal	Der Organisation steht kein Personal zur Verfügung, das mit erhobenen Daten arbeiten kann	Ehrenamtliche Mitarbeitende unterstützen die Organisation bei der Nutzung ihrer Daten durch Analysen oder Bereitstellung einiger Infrastruktur	Hauptamtliches Personal übernimmt wichtige Aufgaben der Analyse und ist fähig, genutzte Anwendungen und Tools in Stand zu halten	Hauptamtliches Personal führt Analysen selbstständig durch und managt notwendige Anwendungen und Tools selbst
Management	Dem Management ist der Nutzen von Daten noch unklar	Das Management sieht die Relevanz von Daten insbesondere im Bereich von Förderungen	Das Management überlässt es Abteilungen, inwieweit sie neben den Basisanforderungen der Berichterstattung Daten nutzen und toleriert den Einsatz von kleineren Projektbudgets für solche Vorhaben	Das Management unterstützt Abteilungen bei der Durchführung ihrer Datenerhebungen und -analyse, stellt notwendige Ressourcen bereit und fordert bei programmatischen Entscheidungen datenorientierte Begründungen ein
Datenschutzrechte	Es gibt kein Konzept zur Datennutzung und zu Zugangsrechten	Es gibt partielle Konzepte zur Datennutzung und zu Zugangsrechten, die sich noch nicht über die gesamte Datenlandschaft erstrecken	Es gibt interne Konzepte für Datennutzung und Zugangsrechte, die in standardisierten Prozessen vergeben werden	Es gibt neben einer internen Datennutzungsstrategie auch Konzepte, ob und wie Daten mit externen Organisationen geteilt werden dürfen
Partnerschaften	Es gibt keine Partnerschaften	Partnerschaften existieren lediglich auf programmatischer Ebene	In Partnerschaften werden manuell Daten geteilt	Der Austausch von Daten und Analysen ist Bestandteil von Partnerschaften
Fördernde	Fördernde verlangen lediglich qualitative Berichte und Zahlen zu oberflächlichen Indikatoren (z. B. Teilnehmerzahl)	Fördernde verlangen eine Analyse von Indikatoren bis hin zur Outcome-Ebene	Fördernde stellen für tiefgreifende Analysen Mittel bereit	Fördernde setzen Datenorientierung für Qualitätsmanagement und Programmlenkung voraus und stellen dafür ausreichend Mittel bereit

Quelle: CorrelAid e. V., Center for Data Science & Public Policy, Data Maturity Framework, University of Chicago

Fragen Sie sich:

- Wo stehen wir als Organisation?
- Wo wollen wir hin?
- An welchen Stellen in den Frameworks finden wir unsere Organisation wieder?
- Was können wir tun, um unsere Ausgangsposition zu verbessern?

Datenfluss

Denken Sie bei jeder Aufgabe *prozessorientiert*. D. h. fragen Sie sich und Ihr Team, wo können Schritte automatisiert werden? Denken Sie dabei an die Stufen Einblicke/Vorhersagen, Halb- und Vollautomatisierung! Mit geeigneten Rechnungsprogrammen können beispielsweise Belege und Verwendungsnachweise automatisch erfasst werden. Einige E-Mail-Programme legen Adressbücher automatisch an. E-Mail-Verteiler erleichtern den Versand von Rundnachrichten.

Datenlandschaft

Mit der folgenden Vorlage verschaffen Sie sich einen Überblick über die Daten in Ihrer Organisation:

Nummer		1.	2.	3.
Titel	Wählen Sie eine eindeutige Bezeichnung für den Datensatz.			
Inhalt und Datentyp	Welche Datenpunkte werden in dem Datensatz erfasst? Um was für einen Datentyp, z. B. Zahlen, Text, Bild, Ton (etc.) handelt es sich?			
Zweck	Warum erheben und wozu nutzen Sie diese Daten? Wer benötigt die Daten, trifft darauf basierend Entscheidungen und wer ist davon betroffen? Gibt es Ausgangs-/Zielwerte und Annahmen und Risiken, auf denen diese basieren? Beispiel für Zwecke sind: Operationale Notwendigkeit, Qualitätsmanagement,			

	Projektsteuerung, Legitimation, Skalierung			
Verantwortliche	Wer ist für Daten zuständig, sachlich und technisch? Ändert sich das über den Zeitablauf?			
DSGVO	Enthält der Datensatz personenbezogene Daten?			
Erhebung	Wie entstehen bzw. werden die Daten erhoben? Gibt es Maßnahmen zur Qualitätskontrolle? In welcher Frequenz werden sie erhoben?			
Speicherort	Wo werden die Daten gespeichert und in welchem Format?			
Zugang	Wer hat zu den Daten Zugang? Wie wird Zugang erteilt?			
Archivierungsort	Wo werden Daten archiviert? In welchem Zeitintervall geschieht das?			
Löschung	Werden historische Daten, die nicht mehr gebraucht werden, gelöscht?			
Verarbeitung	Wie werden Datensätze für die Analyse vorbereitet? Werden Datensätze kombiniert?			

Quelle: CorrelAid e. V., Template - Überblick zur Datenlandschaft

Datentypen:

Je nachdem, was für Datentypen in Ihrem Datensatz enthalten sind, können verschiedene Technologien benutzt werden.

Fokus Zwecke:

Was keinen Zweck hat, sollte auch nicht erhoben werden. Mögliche Gründe für Datenerhebungen sind: Operationale Notwendigkeit, Qualitätsmanagement, Projektsteuerung, Legitimation und Skalierung.

Fokus Erhebung:

Da Data Science so stark von der Datenqualität abhängt, sollte hier auf qualitätssichernde Strukturen gesetzt werden.



Fokus Speicherung:

Wenn in Ihrer Organisation wenige oder keine IT-Ressourcen vorhanden sind, dann ist hier die Auswahl von einfachen, externen Programmen, ggf. mit einem Betreuungsangebot für migrantische Organisationen sinnvoll.

Fokus Dokumentation:

Zu jedem Datensatz sollte eine Beschreibung vorliegen, die alle Kernelemente enthält und erläutert, Indikatoren genau beschreibt und Fragestellungen Indikatoren zugeordnet.

2. Datensicherheit und Datenschutz

Wer mit Daten hantiert, sollte auch immer Sorge dafür tragen, dass diese Daten vor fremden Zugriffen sicher sind. Mit gesetzlichen Regelungen sorgt etwa der Gesetzgeber für gewisse Maßstäbe, die in der Datenverarbeitung einzuhalten sind.

Richtlinien

Art. 1 DSGVO gibt Auskunft darüber, was Gegenstand, Ziele und Grenzen der Verordnung sind:

- *Gegenstand:*

Diese Verordnung enthält Vorschriften zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Verkehr solcher Daten.

- *Ziele:*

Diese Verordnung schützt die Grundrechte und Grundfreiheiten natürlicher Personen und insbesondere deren Recht auf Schutz personenbezogener Daten.

- *Grenzen:*

Der freie Verkehr personenbezogener Daten in der Europäischen Union darf aus Gründen des Schutzes natürlicher Personen bei der Verarbeitung personenbezogener Daten weder eingeschränkt noch verboten werden.

Die DSGVO definiert viele Begrifflichkeiten direkt im Gesetzestext. So fallen z. B. auch (teilweise) automatisierte Prozesse unter den Datenschutz:

Personenbezogene Daten sind alle Informationen, durch die eine natürliche Person, also Menschen, direkt oder indirekt identifiziert werden oder identifizierbar sind. Dazu gehören Merkmale, die



Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind.

Verarbeitung heißt Erheben, Erfassen, Organisation, Ordnen, Speicherung, Anpassung, Veränderung, Auslesen, Abfragen, Verwendung, Offenlegung, Abgleich, Verknüpfung, Einschränkung, Löschung oder Vernichtung in automatisierten und nichtautomatisierten Vorgängen.

Sieben Kernpflichten

Mit der DSGVO hat Ihre Organisation *sieben Kernpflichten*:

- 1. Rechtmäßigkeit**
Daten dürfen nur so verarbeitet werden, wie bei der Erhebung angegeben. Dies muss für Betroffene nachvollziehbar sein.
- 2. Zweckbindung**
Nur Daten, die einen klaren Erhebungszweck haben, dürfen erhoben werden.
- 3. Datenminimierung**
Daten, die für spätere Analysen nicht relevant sind, sollten nicht erhoben werden.
- 4. Richtigkeit**
Daten müssen sachlich richtig und aktuell sein. Falsche Daten müssen gelöscht werden.
- 5. Datenreduktion**
Personenbezogene Daten dürfen nur so lange gespeichert werden, wie sie auch noch einen Zweck erfüllen.
- 6. Vertraulichkeit**
Daten müssen angemessen gesichert sein und vertraulich behandelt werden.
- 7. Rechenschaft**
Entscheidungssträger*innen sind für die Einhaltung dieser Pflichten verantwortlich und müssen dies nachweisen können.

Natürliche Personen und ihre Rechte

Alle natürlichen Personen, deren Daten Sie und Ihre Organisation verarbeiten, haben eine Vielzahl von Rechten (Vor und nach der Datenerhebung):

Transparenz

- Rechtlich geprüfte Einwilligungserklärung einsammeln
- Unterschrift von Betroffenen und ggf. Erziehungsberechtigten einsammeln (online möglich)
- Erklärung beinhaltet u.a. den Zweck der Datenerhebung, das Verfahren, sowie die Speicherdauer - Besondere Kategorien personenbezogener Daten (Art. 9 DSGVO) beachten

Information/Auskunft

- Hinterfragen der Daten auf Zweck und Minimierung
- Klares Verständnis welche Daten für welchen Zweck genutzt werden sicherstellen
- Regelmäßiges Hinterfragen welche Daten ggf. nicht abgefragt werden müssen, insb. wenn es sich um persönliche Merkmale handelt, die eventuell keinen Mehrwert bedeuten.

Berichtigung/Löschung

- Erhobene Daten pseudonymisieren
- Namen der Personen durch Pseudonym (z.B. Teilnehmernummer) ersetzen, um Feststellung der Identität zu erschweren - Bezüge zwischen verschiedenen Datensätzen sind so möglich - im Gegensatz zur Anonymisierung (anonymisierte Daten unterliegen nicht der DSGVO)

Widerspruch

- Angemessene Datensicherheit gewährleisten
- Speichermedien mit Wirkungsdaten physisch und technisch sichern (z. B. Speicherort in Deutschland, USB-Sticks verschlüsseln)
- Mitarbeiter*innen schriftlich zur Einhaltung der Datenschutzrichtlinien verpflichten

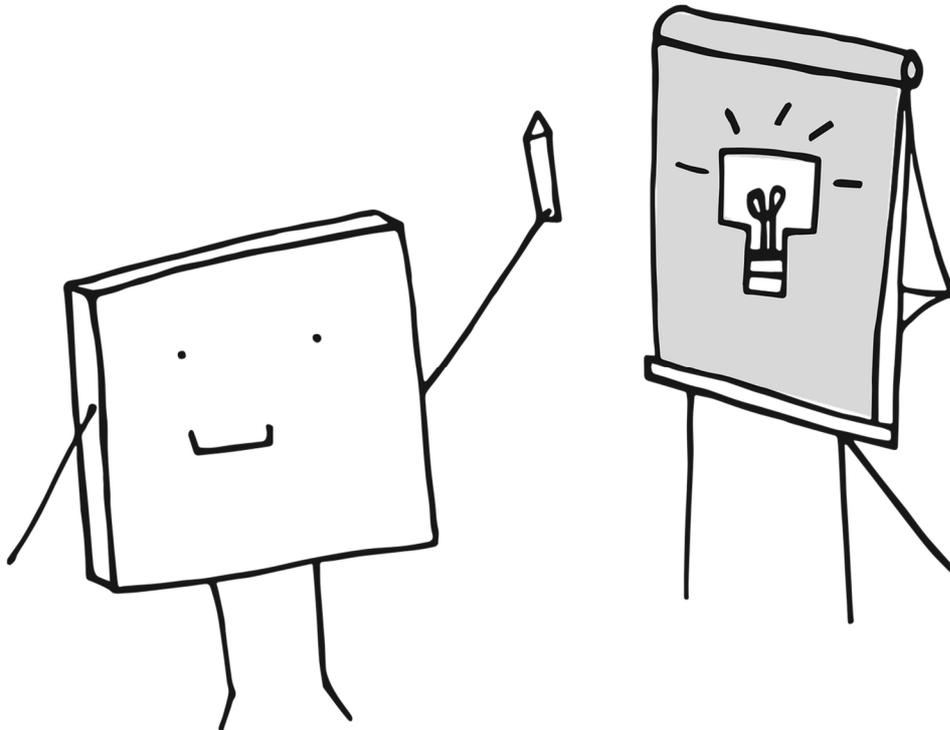
Stellen Sie den Datenschutz in der gesamten Organisation sicher:

- Erstellen Sie einen Umsetzungsplan
- Legen Sie klare Verantwortlichkeiten fest
- Führen Sie ein IT-Sicherheitskonzept ein
- Erstellen Sie Verarbeitungsverzeichnisse
- Beantworten Sie Anfragen zu personenbezogenen Daten innerhalb von 30 Tagen (Nachweispflicht)

3. Visualisierung von Daten mit Tableau

Daten visualisieren

Mit der fortschreitenden Technologisierung können Visualisierungen immer mehr Wirkung entfalten. D. h. dass die neusten Programme oft recht einfach in der Bedienung sind und sich sogar für Marketingzwecke und Reportingformate eignen.



Vorteile der Visualisierungen:

Übersichtlichkeit

Visualisierungen helfen Ihnen und Ihrer Organisation Zusammenhänge zwischen Datenpunkten zu erkennen.

Detailliertheit

Visualisierungen ermöglichen Ihnen und Ihrer Organisation Datenpunkte im Zeitablauf zu betrachten.

Vereinfachung

Visualisierungen erleichtern die Kommunikation von trockenen und komplexen Botschaften.

Aktualität

Daten können in viele Programme nun live eingepflegt werden – damit erhalten Sie in Echtzeit Updates.

Interaktivität

Visualisierungen können interaktiv sein, sodass Empfangende Daten selbst erkunden können.

Attraktivität

Durch die Ausweitung von Gestaltungsmöglichkeiten werden Visualisierungen anschaulicher.

Transparenz

Durch Hosting auf öffentlich zugänglichen Servern können Sie und Ihre Organisation ein größeres Publikum erreichen.

Visualisierungs-Tool Tableau

Ein der Programme, die für Visualisierung u. a. von Data Scientist genutzt werden, ist *Tableau*. Tableau ist eine der bekanntesten Visualisierungssoftwares: <https://www.tableau.com/de-de>



Für Tableau sprechen die Einfachheit der Bedienung, die Vielfalt der Funktionalitäten, gute Infrastruktur, Service und Dokumentation. Leider ist das Automatisierungspotenzial beschränkt. Schauen Sie bei Stifter Helfen <https://www.stifter-helfen.de/> nach einer Lizenz für Ihre Organisation!

Tableau bietet eine weitgehende Infrastruktur zur Datenbereinigung, Visualisierung und Hosting:

Daten bereinigen:

Daten für Tableau liegen idealerweise als Tabelle vor, wobei mit Tableau Prep - und in Tableau Desktop - selbst eine umfassende Bereinigung und Verarbeitung von Daten möglich ist.



Daten laden:

Tableau verfügt über zahlreiche Importfunktionalitäten und Schnittstellen (APIs), mit denen Datensätze in die Anwendung geladen werden können.

Visualisieren:

Bei der Gestaltung von Visualisierungen ist fast alles möglich. Über Filter, Visualisierungsdimensionen und logische Ausdrücke werden Visualisierungen verfeinert.

Layout gestalten:

Abgeschlossene Visualisierungen können in ein Layout eingebettet werden, das auch Text- und Bilddateien enthalten kann.

Veröffentlichen:

Visualisierungen können entweder (z. B. als PDF) exportiert oder auch über den Tableau Server, in dem man auch Zugangsrechte festlegen kann, veröffentlicht werden.

Welche Visualisierungen für welchen Zweck in Frage kommen, können Sie hiermit entscheiden: Tabellen (mit Farbgebung), Heat Maps, Karten, Kreisdiagramme, Balken-/Säulendiagramme, Baumkarten, Linien- und Bereichsdiagramme, Streudiagramme, Histogramme/Box-WhiskerPlots, Gant-/Bullet-Diagramme oder Bubble Charts.

Inspirationen finden Sie bei Tableau Public – dort gibt es eine Galerie:

<https://public.tableau.com/de-de/s/>

Quelle: youvo e. V., Nina Hauser von CorrelAid e. V.

Seminarreihe Digitalverbunden 2020, Modul 4 Daten – mit freundlicher Unterstützung von youvo e. V. und Bundesverband Netzwerke von Migrantenorganisationen (NeMO) e. V.



ARBEITSMATERIAL FÜR MIGRANTEN- ORGANISATIONEN