

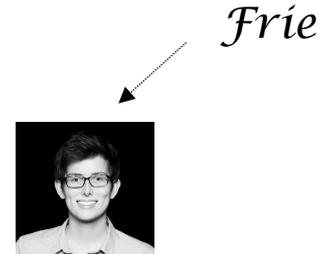


CORRELAID
GOOD CAUSES. BETTER EFFECTS.

Grundlagen Datenmanagement

- 2h -

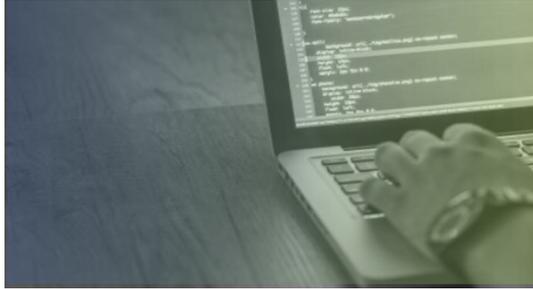
Wer wir sind



Wir sind ein deutschlandweites Netzwerk von über 1,700 Datenwissenschaftler:innen, die die Welt durch und mit Daten verbessern wollen.

#MetaWeltretter

Unsere Mission



PROJEKTE

Wir führen pro-bono Datenanalyseprojekte für gemeinnützige Organisationen durch.



BILDUNG

Wir vernetzen engagierte sozial denkende Datenanalytist:innen und bieten ihnen Möglichkeiten ihr Wissen anzuwenden und zu erweitern.



DIALOG

Wir treten in den Dialog über den Wert und Nutzen von Daten und Datenanalysen für das Gemeinwohl

Inhalte

Themen	Methodik	Zeitaufwand
Grundlagen Datenmanagement	    	45 Min
Q&A	    	15 Min
Übung Excel	    	45 Min
Q&A	    	15 Min

Lernziele

- 1) Teilnehmer:innen kennen typische Datenquellen in gemeinnützigen Organisationen
- 2) Teilnehmer:innen kennen Möglichkeiten der Datenspeicherung
- 3) Teilnehmer:innen kennen grundlegende Konzepte der Datenqualität und Datenstrukturierung
- 4) Teilnehmer:innen können die theoretischen Konzepte in der Praxis mit Excel umsetzen
- 5) Teilnehmer:innen kennen weiterführende Ressourcen



CORRELAID

GOOD CAUSES. BETTER EFFECTS.

Datenquellen in sozialen Organisationen

Manchen Organisationen geht es so...



...anderen so:

?

Wie sehen Daten aus?

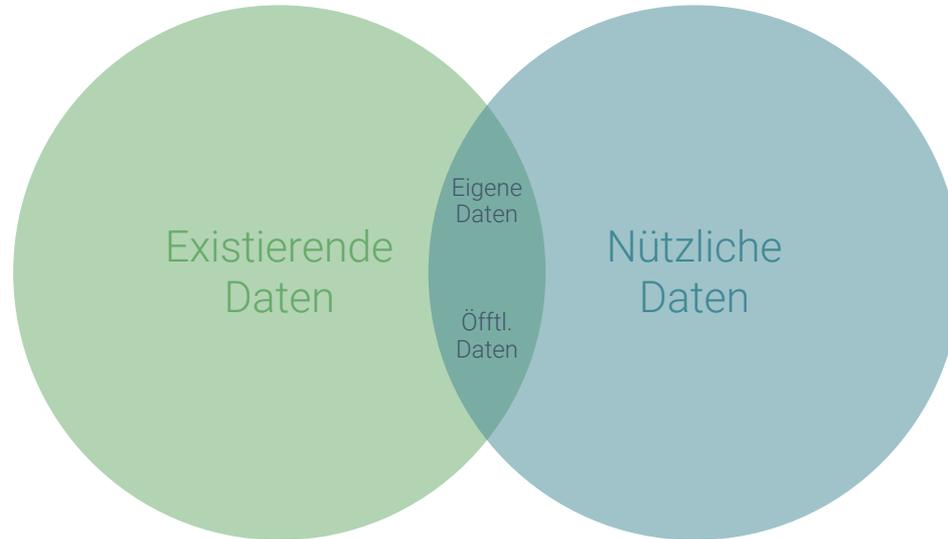
email	kampagne	gelesen
joe@ex.org	a	1
joe@ex.org	a	0
jane@ex.org	b	1
jane@ex.org	b	1

- Beobachtungen → Zeilen
- Variablen: enthalten Werte → Spalten
- Werte → Zellen

Daten können auch unstrukturiert (Bilder, Text, Audio) oder semi-strukturiert (z.B. XML, JSON) vorliegen. Wir fokussieren uns hier auf tabellarische Daten (strukturiert).



Es gibt Daten, die bereits existieren, und solche, die für Euch nützlich sind



Besonders nützlich sind Daten, die bei **operativen Vorgängen** sowieso entstehen – z.B. Kontaktformulare, E-Mails, Chats, ...



Daten können gezielt erhoben, operational gesammelt oder recherchiert werden



Erhebungsdaten

Umfragen,
Beobachtungen,
Registrierungsdaten, ...



Auswahlfreiheit
Flexibilität



Kosten der Erhebung
Zeitintensität



Operationsdaten

Websites (Cookies), E-
Mails, Chats, Social
Media, Portale...



Preiswertigkeit
Unverfälschtheit



Zugriff
Privatsphäre u. DSGVO



Externe Daten

Öffentliche Daten,
Daten anderer NPOs,
Studien, ...



Preiswertigkeit
Neutralität



Generalität
Intransparenz



Datenquellen bei CorrelAid



Operationsdaten

Slack, Newsletter,
Social Media Daten,
Projektbewerbungen,
Projekt-Metadaten, ...



Externe Daten

Daten zum
Digitalreport



Erhebungsdaten

Projektfeedback-
umfrage

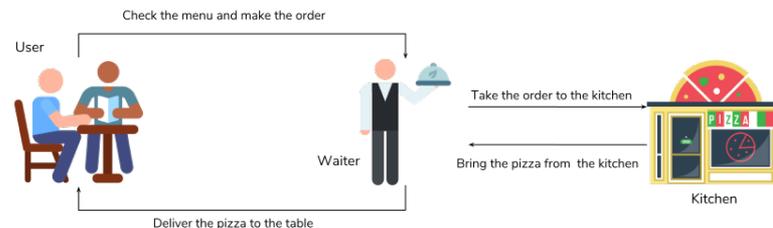
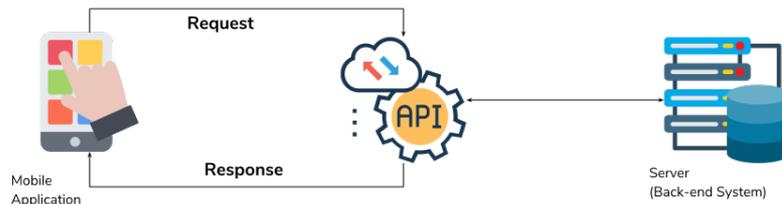
Exkurs: (Web-)APIs

Einfache Erklärung

Eine API ist eine Programmierschnittstelle, die die Kommunikation zwischen verschiedenen Softwarekomponenten ermöglicht. Sie erlaubt Euch beispielsweise Text von einer Applikation in eine andere zu kopieren. Wenn Ihr eine Webseite im Netz öffnet, dann sendet Ihr über eine Web-API eine Anfrage an einen Web-Server. Server sind Computer, die Anfragen verarbeiten und – in diesem Fall - Daten über eine Internet- oder Kabelverbindung an Euren Computer zurücksenden, auf die Ihr dann zugreifen könnt.

Komplexe(re) Erklärung

Eine API (Application Program Interface) ist ein Regelwerk aus Routinen, Protokollen und Tools, das den Bau von Softwareanwendungen ermöglicht. Bei Web-APIs handelt es sich um ein Interface, mit dem über eine HTTP/HTTPS-Abfrage mit einem Authentifizierungssystem GET (anfragende) und POST (schreibende) Befehle über einen Client (Web Browser) auf einem Server ausgeführt werden können. Web-Daten sind normalerweise in JSON oder XML verpackt.



Exkurs: Öffentliche Datenquellen



Startseite → Service → Statistik visualisiert → Regionalatlas



Themen		
1	02	03
01	04	05
06	07	08
09	10	11

Daten der Statistischen Ämter
des Bundes und der Länder
(mit API)



17 Ziele Datenverfi

Indikatoren der UN-Nachhaltigkeitsziele

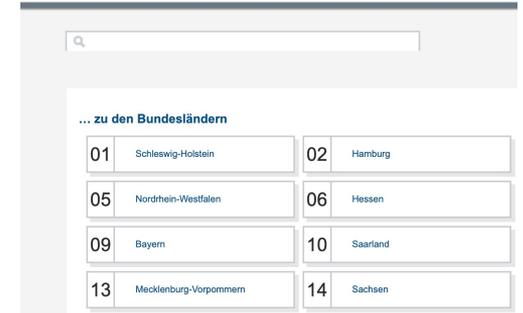
Klicken Sie auf die einzelnen Ziele um zu den Daten zu den Indikatoren der Z Sie gezielt nach spezifischen Themen.



Daten Deutschlands zu den
Zielen für nachhaltige
Entwicklung



Kommunale Bildungsdatenbank



STARTSEITE ERGEBNISSE EINFACH UND SCHNELL

Startseite ▶

Daten zu Bildung und der
Bevölkerung
(Zensus/Mikrozensus)





CORRELAID
GOOD CAUSES. BETTER EFFECTS.

Übung

Welche Datenquellen gibt es in eurer Organisation? Wie liegen sie vor? Wie könnt ihr auf sie zugreifen?

- 2 min -



CORRELAID

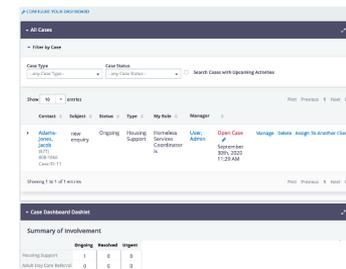
GOOD CAUSES. BETTER EFFECTS.

Datenspeicherung

Exkurs: Was ist eine Datenbank?



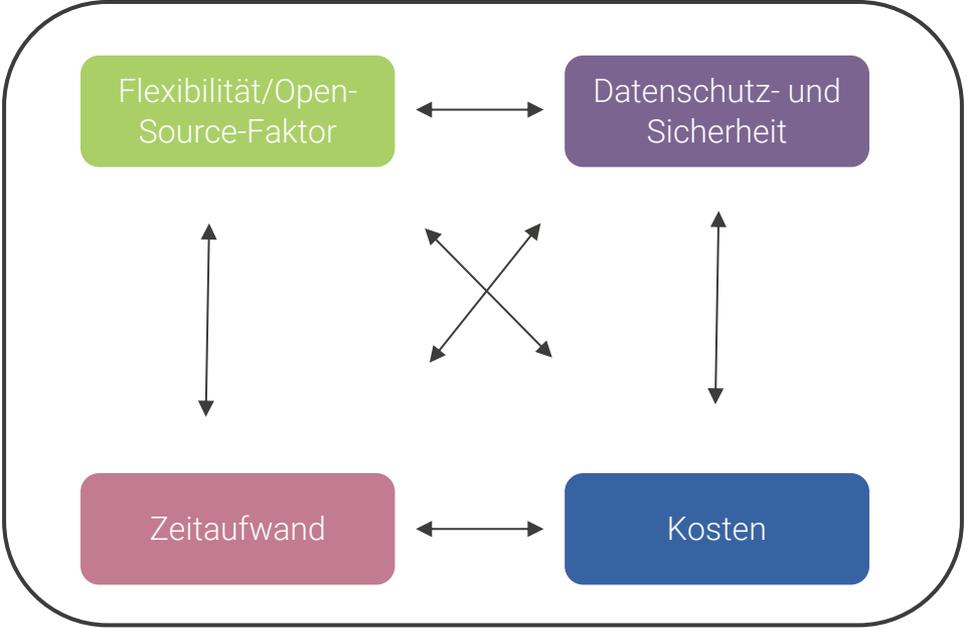
Begriff	Funktion	Beispiel
Datenbankanwendung	Benutzeroberfläche ("GUI") zur Datenpflege	Mitgliederverwaltung, CRM (CiviCRM), Bestellsoftware, (Microsoft Access)
Datenbankmanagementsystem (DBMS)	<ul style="list-style-type: none"> - Schnittstelle zw. Anwendung und Daten - Daten anlegen, überschreiben und löschen - Metadaten ("Schema") - Backup - ... 	PostgreSQL, MySQL, MariaDB, Oracle, (Microsoft Access)
Datenbank (DB)	Speicherung zusammengehöriger Daten im DBMS	Datenbank "projekte" in CorrelAid-PostgreSQL



Optionen zur Datenspeicherung

1. lokale Excel-Dateien
2. selbst gehostete Tools, z.B. CiviCRM, KoboToolbox, SQL Datenbank
3. fremd gehostete Tools, z.B. Google Sheets, Salesforce, Seatable, gehostetes CiviCRM, KoboToolbox, gehostete SQL Datenbank

Evaluationskriterien



Lokale Excel-Dateien*

Vorteile

- einfach zu bedienen, vertraut
- Kombination mit MS Access möglich
- einfach zu teilen, z.B. via Email
- keine zusätzlichen Kosten, wenn Office Lizenzen vorhanden

Nachteile

- keine fest definierten Datenstrukturen
- keine automatisierten Abfragen, da keine Programmierschnittstelle vorhanden
- eignet sich nicht zur Datenweiterverarbeitung in anderen Tools wenn primär zur Dateneingabe / Datendarstellung genutzt
- kollaboratives/gleichzeitiges Arbeiten nicht möglich

*Super Open-Source-Alternative: Libre Office!

Selbst gehostete Tools

Vorteile

- Datenhaltung auf eigenen Servern --> Datenschutz gewahrt
- eigene Anpassungen möglich (v.a. bei Open Source Tools)
- weniger Abhängigkeit von Anbietern (insb. hinsichtlich Preismodellen, Features)
- große Auswahl an spezifischen Open Source Tools
- 😎

Nachteile

- initialer Aufwand hoch / (meistens) technische Expertise benötigt
- Wartungs- und Serverkosten / Zeitaufwand
- selbst verantwortlich für Datensicherheit!
- keine Garantie, dass Open Source Software weiterentwickelt wird
- kein garantierter Support

Beispiel KoBoToolbox

GET STARTED

Get started with KoBoToolbox now, it's free and it takes under a minute. We have two publicly-available instances of KoBoToolbox you can choose from:

Unlimited Use for Humanitarian Organizations

Provided by UN OCHA
Unlimited Submissions
Unlimited Data Storage
Unlimited Projects

[CREATE AN ACCOUNT](#)

or [login](#)

Researchers, Aid Workers & Everyone Else

Provided by KoBoToolbox
10,000 Submissions Per Month
5GB of Data Storage Per Month
Unlimited Projects

[CREATE AN ACCOUNT](#)

or [login](#)

For more information on the differences between the two servers and to help determine which one is right for you, please read our [Which Server Should I Use?](#) support article.

Advanced users can also install KoBoToolbox on their own server (or on a local machine) using Docker. See our [kobo-install](#) repository on GitHub for details.

- Nutzerfreundliches Design
- Diverse Formatfunktionalitäten
- Kollaborationsmöglichkeiten
- Exportformate und API
- SSL Integration
- Online und offline
- Desktop und mobil
- Open Source (Code)



Fremd gehostete Tools

Vorteile

- praktisch: kein Aufwand für Wartung / Ersteinrichtung
- meistens (guter) Support / Updates
- meistens Rabatte für gemeinnützige Organisationen (stifter-helfen oder selbst anfragen)

Nachteile

- Preismodelle (kostenlose Pläne stoßen häufig schnell an Grenzen)
- Datenweitergabe
- technische Abhängigkeit
- Problem des Serverstandorts

Datenschutz und -sicherheit

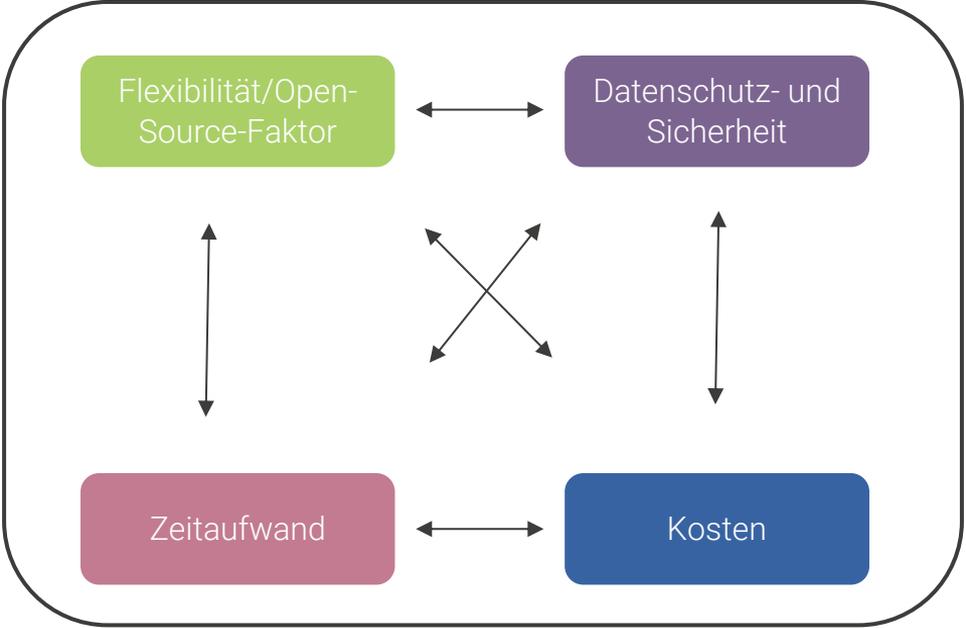
Datenschutz

- Was für Daten werden gespeichert?
- Rechtsgrundlage für die Datenverarbeitung
- Serverstandort
- Auftragsdatenverarbeitung
- Datenschutzerklärung

Datensicherheit

- physische Zugangskontrolle
- Verschlüsselung
- Passwortmanagement
- Benutzerverwaltung

Evaluationskriterien



Evaluation: Für kleine Datensätze, die nur selten Updates benötigen, ist oft auch Excel ausreichend



Excel

Excel bietet als kostengünstiges Office-Tool eine gute Möglichkeit kleine Datensätze, an denen nur eine oder sehr wenige Personen arbeiten, abzubilden.



Onlineversionen ermöglichen Kollaborationen
Geringe Zugangshürden für versch. Nutzer:innen
Integration mit fast allen Tools
Kostengünstige Anschaffung



Rudimentäre Eingabekontrolle
Schutz- und Nutzungsrechte oft nicht ausreichend
Geringes Automatisierungspotenzial für Reports
Funktionalitäten begrenzt



Datenbankmanagementsystem

Bei Datensammlungen, die organisationsweit genutzt werden und regelmäßig angepasst werden, lohnt sich die Anschaffung einer Datenbanklösung.



I.d.R. fortgeschrittene Schutz- und Nutzungsrechte
Komplexe Abfragen möglich
Automatisierung von Reportprozessen
Verschiedene Datenansichten
Umfangreiche Eingabekontrolle



Schnittstellen zur Toolintegration notwendig
Umfangreiche Planung nötig
Hoher Kostenfaktor
Verwaltungs- und Maintenance-Anforderungen hoch



Evaluation: MS Access als Kompromiss



- MS Access* ist eine *dateibasierte* Datenbank- und Entwicklungsumgebung, die eine graphische Benutzeroberfläche mit relationalen Datenbanksystem kombiniert

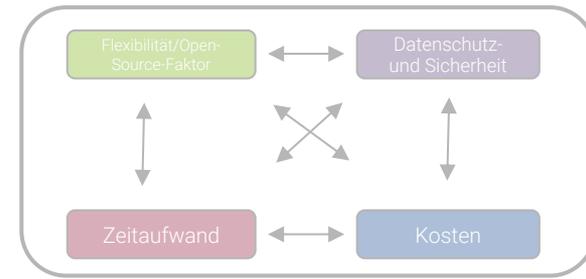


Graphische Benutzeroberfläche
Benutzer:innenrollen können vergeben werden
Relationale Datenbanksysteme



Dateibasiert -> für viele gleichzeitige Nutzer:innen
nicht geeignet
Lizenzen werden benötigt (bis zu zehn für NPOs
kostenlos)
Nur auf Windows PC nutzbar

Übung: 5 Minuten



- Daten: wie komplex sind eure Daten? Wie viele Daten fallen bei euch an und wie oft?
- **Kosten** vs. **Zeitaufwand**: Habt ihr finanzielle und zeitliche Ressourcen um einen eigenen Server zu unterhalten?
- **Kosten** vs. **Flexibilität**: lohnt sich ein Abo-Modell von einem extern gehosteten Tool (insb. in Bezug auf User-Management)? Bieten gehostete Tools alles, was ihr benötigt? Reichen evtl. auch Workarounds mit vorhandenen Tools (z.B. Excel und eine in Deutschland gehostete Cloud)?
- **Datenschutz- und Sicherheit** vs. **Zeitaufwand**: Arbeitet ihr mit personenbezogenen Daten? Könnt ihr wählen, auf welchem Server eure Daten gespeichert werden? Habt ihr personelle Ressourcen um einen eigenen Server zu unterhalten?
- ...



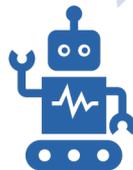
CORRELAID

GOOD CAUSES. BETTER EFFECTS.

Datenqualität

Was ist Datenqualität?

“fit for use”: Daten
haben eine hohe
Qualität, wenn sie für
euch “funktionieren”



Kriterien für Datenqualität

Korrektheit

Daten müssen mit der Realität übereinstimmen

Konsistenz

Ein Datensatz darf in sich und zu anderen Datensätzen keine Widersprüche aufweisen

Vollständigkeit

Datensatz muss alle notwendigen Attribute enthalten.

Redundanzfreiheit

Keine Dubletten („Duplicates“)

Einheitlichkeit

Einheitliche Strukturierung der Informationen eines Datensatzes

Zuverlässigkeit

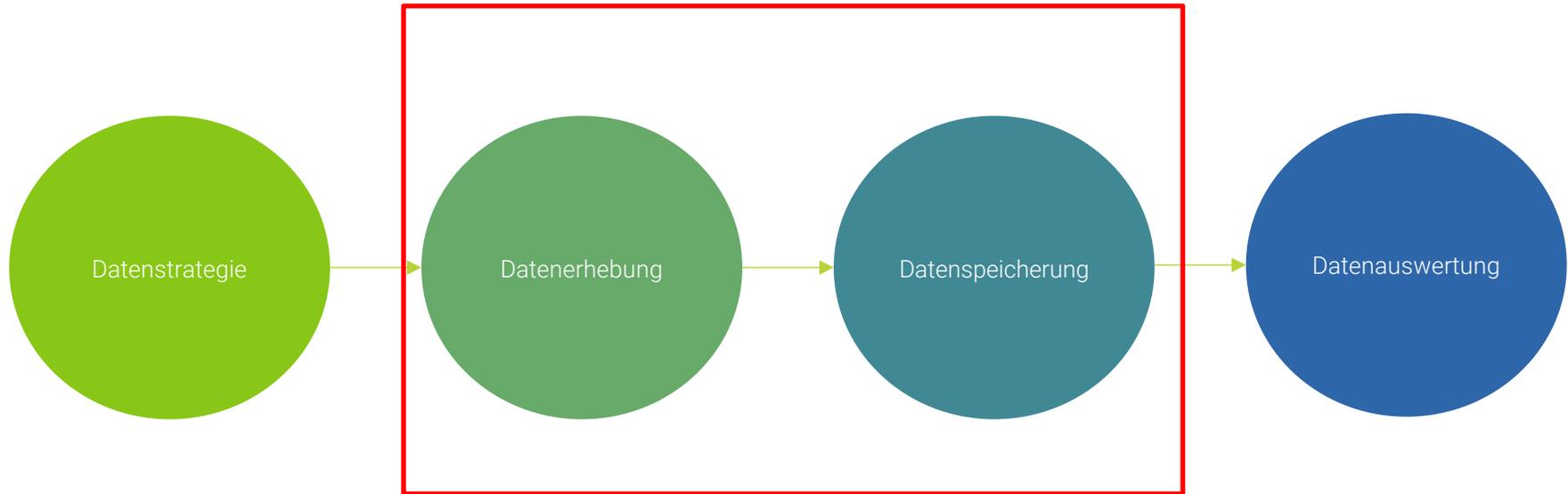
Die Entstehung der Daten muss nachvollziehbar sein

Aktualität

Alle Datensätze müssen jeweils dem aktuellen Zustand der abgebildeten Realität entsprechen.

Zusätzliche Kriterien: Genauigkeit, Relevanz, Verständlichkeit, Eindeutigkeit

Wie könnt ihr eure Datenqualität erhöhen?



Wie könnt ihr eure Datenqualität erhöhen?

Datenerhebung/-generierung

Durch Maßnahmen bei der Datenerhebung/-generierung könnt ihr die Korrektheit, Konsistenz, Zuverlässigkeit und Einheitlichkeit eurer Daten erhöhen.

Datenstrukturierung

Wie und in welcher Struktur ihr eure Daten abspeichert / ablegt, beeinflusst, wofür ihr sie später *ohne großen Aufwand nutzen* könnt

“fit for use“: Daten haben eine hohe Qualität, wenn sie für euch “funktionieren“



User Entry Control - Umfragen

- Wenn möglich, formuliert Fragen geschlossen (90% / 10% Daumenregel)
- Benutzt einheitliche Skalen

Wie fandest du unsere Veranstaltung? (ohne User Entry Control)

Wie zufrieden warst du mit der Veranstaltung? (mit User Entry Control)

- sehr unzufrieden
- unzufrieden
- teils teils
- zufrieden
- sehr zufrieden

Konsistenz

Einheitlichkeit

Zuverlässigkeit

User Entry Control - Umfragen

Aus welcher Stadt kommst du? (ohne User Entry Control)

Berlin Neukölln

Aus welcher Stadt kommst du? (mit User Entry Control)

Klicke einfach auf das Zentrum der Stadt wo du herkommst - wir wollen nicht deine genaue Adresse wissen!

latitude (x.y °)

52.494487

longitude (x.y °)

13.455504

altitude (m)

accuracy (m)



search for place or address



– Je nach Tool gibt es viele Möglichkeiten zur User Entry Control

Konsistenz

Zuverlässigkeit

Einheitlichkeit

User Entry Control - Umfragen

– Datenvalidierungen verhindern unrealistische / objektiv unmögliche Eingaben

Wie alt bist du? (ohne User Entry Control)

-23

Wie alt bist du? (mit User Entry Control)

-23

Bitte gib ein Alter zwischen 0 und 120 Jahren an

Konsistenz

Korrektheit

Zuverlässigkeit

User Entry Control - Umfragen

Die offene Frage wird nur bei
"sehr unzufrieden" oder
"unzufrieden" angezeigt

Wie zufrieden warst du mit der Veranstaltung? (mit User Entry Control)

- sehr unzufrieden
- unzufrieden
- teils teils
- zufrieden
- sehr zufrieden

Bitte gib uns Feedback zur Verbesserung unserer Veranstaltung

Konsistenz

Korrektheit

Zuverlässigkeit

–Fragelogik: Zeigt Fragen nur an,
wenn sie auch relevant für die
beantwortende Person sind

Mehr zum Thema User
Entry Control in Excel
in der Übung!





CORRELAID

GOOD CAUSES. BETTER EFFECTS.

Übung