

Der technische Datenreifegrad sozialer Organisationen – Ein Modellierungsframework

Thema	Noch nicht ausreichend	Ausreichend	Fortgeschritten	Ausgezeichnet
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
Erhebungsfrequenz	Einmalig	Jährlich	Regelmäßig	Durchgängig
Granularität	Gesamtheitlich aggregiert (z.B. absolute Teilnehmerzahl)	Pro Gruppe aggregiert (z.B. pro Geschlecht)	Entitätenebene (z.B. pro Person)	Ereignisebene (z.B. eine Person über mehrere Veranstaltungen)
Qualität	Es fehlen Datenpunkte (Individuen oder Einheiten)	Es fehlen Variablen (Spalten)	Es fehlen keine Daten, allerdings finden sich Fehler, die u.a. durch mangelnde Einschränkungen bei der Dateneingabe entstehen	Es fehlen keine Daten und Fehler bei der Dateneingabe werden durch umfassende Maßnahmen beschränkt
Speicherung	Papier	PDFs/Bilder	Textdateien (CSV, XML, ...)	Datenbanken
Analyse	Keine oder nur rudimentäre Analysen	Einfache manuelle Analysen in Excel	Automatisierte Analysen mithilfe externer Software	Anspruchsvolle, selbst programmierten Analysen und Automatisierungen
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
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
Datenschutz	Nicht implementiert	Partiell implementiert (z.B. Passwortschutz für sensible Dateien und Systeme)	Weitestgehend implementiert (z.B. zuständige Person beauftragt, Datenverarbeitungsverzeichnis angelegt, erste organisatorische und technische Hürden eingeführt)	Vollständig implementiert (z.B. zert. Datenschutzbeauftragte:n ernannt, Datenverarbeitungsverzeichnis und Maßnahmenplan angelegt, hohe organisatorische und technische Hürden implementiert)
Dokumentation	Keine Dokumentation zu Variablen	Codebuch mit Variablen und Kategorien	Codebuch mit Metadaten und Erhebungsmethoden	Codebuch mit Metadaten, Erhebungsmethoden, Annahmen, Ausschlusskriterien und potenziellen Verzerrungen
Historie	Historische Daten werden gelöscht	Historische Daten werden bei Updates überschrieben	Personenbezogene Daten werden bis zu 3J. nach Erlöschung der Zweckbindung gespeichert und bei Updates aller Datentypen neue Datenpunkte angefügt	Personenbezogene Daten werden 3J. nach Zweckbindungserlöschung aggregiert, Analysen abgespeichert und bei Updates lassen sich Änderungen nachverfolgen