

Datenreifegrad sozialer Organisationen – Ein Modellierungsframework

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	PDFs/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 mangelnde Einschränkungen bei der Dateneingabe entstehen	Es fehlen keine Daten und Fehler bei der Dateneingabe werden durch umfassende Maßnahmen beschränkt
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)
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	Partiell 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