



OPEN GOVERNMENT
ACADEMY

KURZÜBERBLICK

Datenqualität bei Open Data

Datenqualität bei Open Data: Wo anfangen?

- › **Überblick verschaffen!** Öffentliche Einrichtungen verfügen über umfangreiche und heterogene Datenbestände. Ein organisationsweiter Überblick über den vorliegenden Datenbestand fehlt oftmals, da die Daten meist proprietär und isoliert voneinander in „Datensilos“ gespeichert und verwaltet werden. Zum Einstieg in das Thema Datenqualität sollte Sie sich daher zunächst einen Überblick über den Gesamtdatenbestand verschaffen. Die wichtigste Frage zu Beginn:
- › **In welchem Format liegen die Daten vor?**
- › Auch die Identifikation der potenziellen Datennutzenden und ihrer Anforderungen an die Daten ist beim Einstieg ins Thema sinnvoll und wichtig. Auf dieser Basis lassen sich Standards hinsichtlich Datenqualität und Datenstruktur entwickeln.

Woran bemisst sich Datenqualität?

- › **Datenqualität** ist ein vielseitiges Konstrukt, welches – je nach Verwendungskontext der Daten – unterschiedliche Anforderungen an die Daten mit sich bringt.
- › Im Idealfall erfüllt ein Datensatz alle Kriterien der FAIR-Prinzipien. In der Realität kostet die Aufwertung von Daten personelle sowie finanzielle Ressourcen, sodass meist Datensätze in minderer Qualität vorliegen.

Die FAIR-Prinzipien bieten eine Strukturierung für die unterschiedlichen Qualitätsindikatoren. Demnach sollten Daten so auffindbar (**Findability**), zugänglich (**Accessibility**), interoperabel (**Interoperability**) und wiederverwendbar (**Reusability**) wie möglich sein.



Hohe Datenqualität sicherstellen durch ...

- › **Auffindbarkeit:** Über Metadaten werden Daten auffindbar (siehe Leitfaden zur Metadatenerfassung Open.NRW-Portal);
- › **Zugänglichkeit** und **Verfügbarkeit:** Der Zugang und die langlebige Verfügbarkeit der Daten („Broken Links“) sollten gewährleistet werden;
- › **Offenheit:** Die Zugänglichkeit und Wiederverwendung von Daten wird bestärkt, wenn sie in einem offenen Format bereitgestellt werden (die Bereitstellung von Daten in proprietärem Format baut stets eine Hürde auf);
- › **Maschinenlesbarkeit:** Um Daten in vielfältigen Anwendungsfällen weiterverwenden zu können, sollten sie in einem maschinenlesbaren Format bereitgestellt werden, welches eine automatisierte Weiterverarbeitung der Daten ermöglicht;

Hohe Datenqualität sicherstellen durch ...

- › **Konsistenz:** Daten und Metadaten müssen widerspruchsfrei sein;
(Bsp. Änderungsdatum vor Erstellungsdatum)
- › **Aktualität:** Daten sind dann von Nutzen, wenn sie aktuell sind und ein möglichst zeitgemäßes Abbild der Realität liefern;
- › **Konformität:** Bei der Bereitstellung von Daten sollten jeweils relevante Standards berücksichtigt werden (Bsp. Datumsangaben nach ISO 8601);
- › **Verständlichkeit:** Die Struktur der Daten sollte so gewählt sein, dass sie für Außenstehende leicht verständlich ist.

Empfohlene Dateiformate

Die Daten sollen maschinenlesbar und in offenen Datei- oder Schnittstellenformaten und vorrangig über offene Schnittstellen bereitgestellt werden. Die Open Data-Verordnung gibt mit Anlage 1 hier für geeignete Formate vor.



Beratungsstelle Open Data

Ministerium für Heimat, Kommunales, Bau und Digitalisierung
des Landes Nordrhein-Westfalen

Jürgensplatz 1

40219 Düsseldorf

Telefon: 0211 8618-50

Telefax: 0211 8618-54444

info@mhkbd.nrw.de

www.mhkbd.nrw