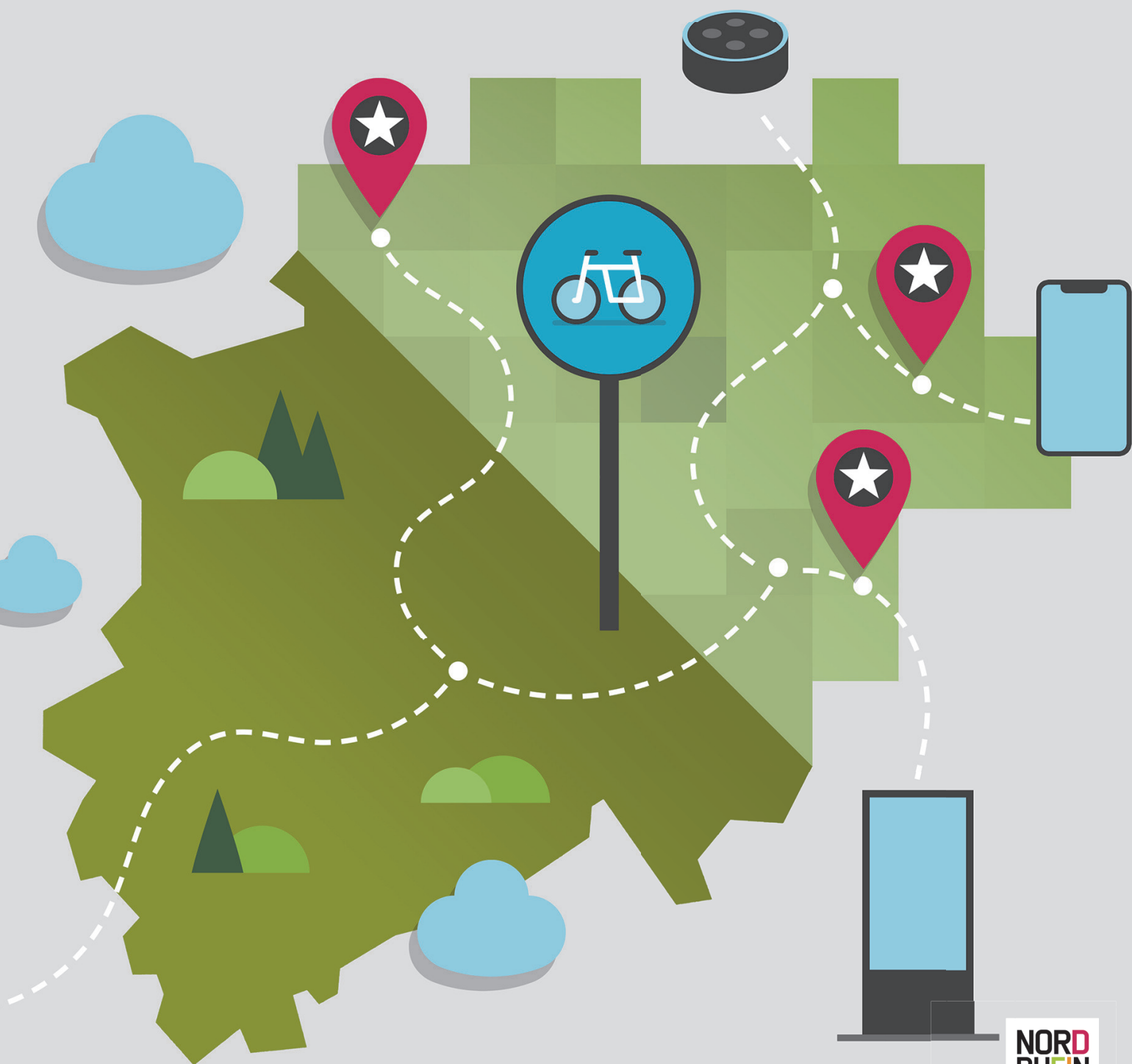


DATAHUB NRW

DATENPFLEGE



**NORD
RHEIN
WEST
FALEN**

Tourismus NRW e.V.

GEFÖRDERT DURCH

Ministerium für Wirtschaft, Innovation,
Digitalisierung und Energie
des Landes Nordrhein-Westfalen



EUROPÄISCHE UNION
Investition in unsere Zukunft
Europäischer Fonds
für regionale Entwicklung

Data Hub NRW

Datenpflege

- Welche Datensätze fließen in den Data Hub NRW?
- Worauf legen wir mit dem Data Hub NRW wert?
- Worauf ist bei der Pflege der Daten zu achten?
- Was bringt der Data Hub NRW für Vorteile mit sich?



Datensatz

Ein Datensatz besteht aus einer Sammlung von miteinander verknüpften, aber eigenständigen Informationen, die in unterschiedliche Datentypen kategorisiert werden können. Die unterschiedlichen Datentypen eines Datensatzes können entweder einzeln oder auch gemeinsam aufgerufen und weiterverarbeitet werden. Jeder Datensatz ist in einer bestimmten Datenstruktur organisiert. Die Deutsche Zentrale für Tourismus e.V. befindet für den Tourismus vor allem drei Arten als wichtig:

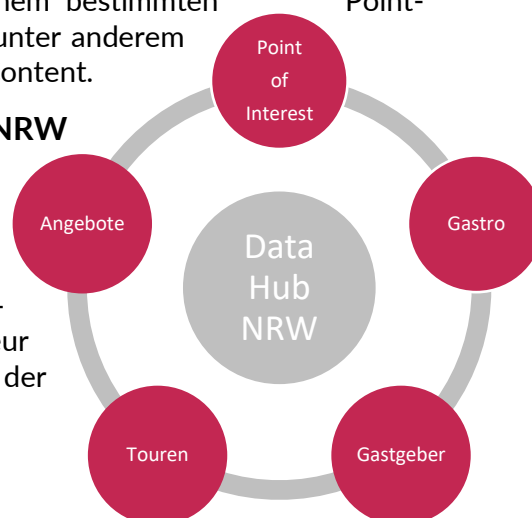
Bei sogenannten **statischen Daten** handelt es sich um dauerhaft bestehende Daten. Änderungen dieses Datentyps fallen vergleichsweise selten an. Beispiele für statische Daten sind unter anderem Geo- oder Adressdaten.

Die zweite Art der Daten stellen **dynamische Daten** dar. Im Gegensatz zu statischen Daten ändern sich diese häufig. Beispiele für dynamische Daten sind Auslastungskennzahlen wie ein Besucher- oder Verkehrsaufkommen. Oftmals handelt es sich bei dynamischen Daten um Echtzeitdaten.

Die dritte Datenart bilden die **redaktionellen Daten**. Hierbei handelt es sich um schützenswerte Daten, welche zu einem bestimmten Point-of-Interest erstellt wurden. Beispiele sind unter anderem Beschreibungstexte oder Bild- und Video-Content.

Datentypen des Data Hub NRW

In Bezug auf das TDM-Projekt sowie die Umsetzung des Data Hub NRW unterscheiden wir folgende Datentypen (siehe Abbildung). Wir haben uns für eine Schnittstellentechnologie entschieden. Kurzum bedeutet das für unsere Partner: Jeder Akteur kann seine Daten weiterhin in dem System der eigenen Region pflegen.



Attribute zum Messen der Datenqualität

Durch die zentrale Steuerung und einheitliche Qualitätskriterien werden die Datenqualität und die Nutzbarkeit der Daten in Nordrhein-Westfalen stark erhöht. Datenabnehmer bekommen damit ein verlässliches und vergleichbares digitales Abbild von ganz NRW. Der Tourismus NRW e.V. sowie seine Partner achten dabei insbesondere auf folgende Kriterien.

1. Aktualität
2. Vollständigkeit
3. Korrektheit
4. Zuverlässigkeit
5. Einheitlichkeit

Eine hohe Datenqualität schafft Vertrauen in die Daten. Für Datenabnehmer und nicht zuletzt für den Nutzer der digitalen Services. Gleichzeitig wird eine Grundlage für zukünftige Analysen (z.B. für die Verbesserung des digitalen Angebotes in Nordrhein-Westfalen) geschaffen. Qualitativ-hochwertige Daten machen die Abnahmen von Daten aus dem Data Hub gegenüber externen Kooperationspartnern noch attraktiver und sorgen somit für mehr Reichweite und Relevanz. So wird das Reiseland Nordrhein-Westfalen fit für die Zukunft gemacht.

Datenqualitätskriterien

Die Datenqualität eines Datensatzes wird erst bei einer erneuten Speicherung oder einem erneuten Import aktualisiert.

Die Basisanforderungen sind, gemeinsam mit mindestens einer CC-BY-SA Lizenzierung des gesamten Datensatzes (nicht der Mediendatei), die Basis-Hürde, um in den Hub zu gelangen. Sind diese erfüllt, werden die Daten mit einem roten Daumen (DQ-Wert bis 25%) ausgezeichnet. Datensätze mit guter Qualität (DQ-Wert zwischen 25% und 75%) erhalten einen gelben Daumen, solche mit sehr guter Qualität (DQ-Wert ab 75%) einen grünen Daumen. Die Kriterien unterscheiden sich je nach Datentyp (siehe nächste Seite).



Datensätze in destination.data, die **rot markiert** sind, erfüllen nicht die Kriterien des Importfilters. Bei ihnen sind nicht alle Pflichtfelder befüllt, z.B. keine Kategorie vorhanden.

Sie werden aktuell in destination.data zur Info angezeigt, um zu sehen, was im Hub landen könnte und um die Datenanalyse zu erleichtern. Sie werden aber NIE ausgespielt - auch wenn sie in der NRW-Experience enthalten sein können und eine offene-Lizenz haben.

	POI	Touren	Gastro
Basisanforderungen	Titel	Titel	Titel
	Kategorie	Kategorie	Betriebsart
		Tour (Polyline)	Geo-Information
	Geo-Information	Tourdaten (Länge)	
Anforderungen für gute Qualität	Öffnungszeiten	Schwierigkeitsgrad	Ort
			Straße
	Teaser-Text UND/ODER Beschreibungstext vorhanden	Teaser-Text vorhanden	Telefon UND/ODER E-Mail
			Webseite UND/ODER Social Media
Beschreibungstext vorhanden	Beschreibungstext vorhanden	Öffnungszeiten	
		Teaser-Text UND/ODER Beschreibungstext vorhanden	
Anforderungen für sehr gute Qualität	Erreichbarkeit	Tourdaten (Dauer)	Erreichbarkeit
	Preisinformation		Eignung
	Webseite UND/ODER Social Media		
	Straße	Ausgangspunkt (Beschreibung)	
	Ort	Zielpunkt	
	Mind. 1 Bild mit mind. CC-BY-SA		
		Events	Gastgeber
Basisanforderungen	Titel	Titel	
	Kategorie	Betriebsart	
	Geo-Information	Ort	
	Ort		
	Datum (inkl. Uhrzeit)		
Anforderungen für gute Qualität	Straße	Straße	
	Webseite UND/ODER Social Media	Webseite UND/ODER Social Media	
	Teaser-Text UND/ODER Beschreibungstext vorhanden	Telefon UND/ODER E-Mail	
		Teaser-Text UND/ODER Beschreibungstext vorhanden	
Anforderungen für sehr gute Qualität	Telefon UND/ODER Buchungslink	Beschreibungstext vorhanden	
	Preisinformation	Merkmale	
	Mind. 1 Bild mit mind. CC-BY-SA*		

Leitlinien für die inhaltliche Pflege von Datensätzen für den Data Hub NRW

Um ein einheitliches Bild und auch inhaltlich eine landesweite Qualität gewährleisten zu können gibt es einige Do's and Dont's für Beschreibungstexte, Bildunterschriften und weitere Textfelder. Letztendlich werden die Texte nicht nur für die eigene Webseite genutzt, sondern dienen unzähligen Ausspielungskanälen als Input. Je universeller die Texte somit formuliert werden, desto mehr Datenabnehmer finden sich am Ende.

Bezeichnung/ Titel 1 - 5 Wörter	Bezeichnung/ Titel Sonderzeichen Detaillierte Beschreibung Hinweis auf Barrierefreiheit
Einleitung 1-2 Sätze	Einleitung Hinweis auf Ausspielungskanal
Beschreibungstexte 150-250 Wörter	Beschreibungstexte Rechtschreibfehler viele Absätze Referenz zu anderen Datensätzen Fremd- und Fachwörter vermeiden
Bildtitel max. 55 Zeichen (inkl. Leerzeichen) Kurze Information, was auf dem Bild zu sehen ist inkl. geografischer Referenz	Bildtitel reine Zahlen unaussagekräftige Beschreibung Sonderzeichen keine reine Dateibeschreibung (Medien/Image usw.)
Bildbeschreibung ca. 180 Zeichen (inkl. Leerzeichen) Präsens Deskriptiv	Bildbeschreibung Absätze
Kategorie IST / HAT-Regel (siehe unten)	

Pflege von Kategorien

Die Kategorie ist ein wichtiger Indikator für Ausspielung in Datensätzen in aller Arten von Anwendungen. Durch die Auswahl einer oder mehrerer Kategorien wird ein Datensatz in den Kategoriebaum des Data Hub NRW eingruppiert.

Bei der Vergabe von Kategorien ist auf die IST / HAT-Regel zu achten. Diese kann am besten an einem Beispiel erläutert werden.

Beispieldatensatz: „NRW-Museum“

Das „NRW-Museum“ IST ein Museum. Aus diesem Grund wird die Kategorie „Museum / Ausstellung“ vergeben.

Das „NRW-Museum“ HAT einen Parkplatz und HAT einen Spielplatz. Dies sind jedoch keine Kategorien, sondern Merkmale die gesondert gepflegt werden können.

Folglich kann mit der IST / HAT-Regel einfach bestimmt werden, ob eine vermeintliche Kategorie wirklich eine Kategorie ist oder es sich viel mehr um ein Merkmal des Datensatzes handelt. Vor allem für Fremdanwendungen ist die korrekte Pflege der Kategorien notwendig, da sonst die Datensätze unbrauchbar sind.

Warum Data Hub NRW?

Wir arbeiten daran möglichst viele, qualitativ hochwertige, offen lizenzierte Daten zu sammeln und somit an Relevanz bei ausspielenden Kanälen zu gewinnen. Leistungsträger können über die Regionen weiterhin im eigenen System pflegen, aber mehr Reichweite für die eigenen Daten und somit ihr Angebot generieren - auch (inter-)national.

