



Aktivitäten und Stand der GDI-NW

GDI-Forum Nordrhein-Westfalen

4. Dezember 2018



Geokatalog

Metadaten

Der GEOkatalog ist das Metadateninformationssystem des Landes NRW und stellt den zentralen Bestandteil des GEOportal.NRW dar.

Im GEOkatalog können Geodaten, Geodienste und Anwendungen registriert, erfasst, gesucht, gefunden und ausgewählt werden.

Bereits existierende andere Metadatenkataloge werden mittels genormter Schnittstellen an den GEOkatalog angebunden. Auf der anderen Seite ist der GEOkatalog in gleicher Weise auch an den Geodatenkatalog der GDI-DE angebunden, der wiederum die Metadaten für die EU bereitstellt. Geodatenhaltende Stellen im Land NRW, die über kein eigenes Metadateninformationssystem verfügen, können die Metadaten online im GEOkatalog erfassen und öffentlich verfügbar machen.

Um die Qualität und die Einheitlichkeit der Metadaten zu steigern, wurde durch die Geschäftsstelle des IMA GDI.NRW ein Leitfaden zur Erfassung der Metadaten erstellt. Die Qualität der Metadaten ist entscheidend für die Auffindbarkeit und Aussagekraft von Informationen zu Daten und Diensten in einer Geodateninfrastruktur.

Nur was dort richtig drinnen steht kann auch gefunden werden!



Aus dem INSPIRE-Monitoring unsere GDI-Aktivitäten in DE [NW]:

Gemäß Artikel 21 Abs. 1 der INSPIRE-Richtlinie überwachen die Mitgliedstaaten die Schaffung und Nutzung ihrer Geodateninfrastrukturen (Monitoring). Sie stellen die Ergebnisse dieser Überwachung der Kommission und der Öffentlichkeit auf Dauer zur Verfügung.

2017 Deutschland gesamt:

- Gemeldet wurden 22.369 Datensätze, damit 50 % mehr als 2016 [NW 218] (NI starker Zuwachs: kommunale B-Pläne)

- Konformität von Metadaten 98-100%

- Zugänglichkeit:

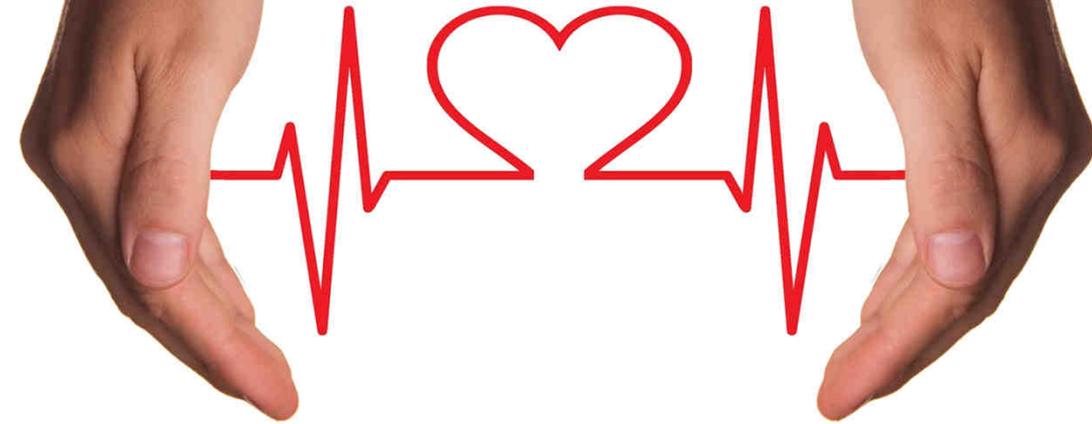
- 57 % der Datensätze über Darstellungsdienste [NW 85 %]
- bei Downloaddiensten sind es 57 % [NW 60%, fehlende Daten-Dienste-Koppelung?], aber => 5 Länder erreichen 100 % Zugänglichkeit der Datensätze über Dienste

Interoperabilität Anhang I

- von 974 gemeldeten Datensätzen wurden nur 24 % interoperabel bereitgestellt [NW nur 60 %, weil auch Meldung von nicht-interoperablen Ausgangsdatsätzen erfolgte].



Metadatengesundheit



Metadaten sind der Schlüssel zu Ihren Geo-Daten und -Diensten! Daher ist bei der Erfassung besonderes Augenmerk auf Qualität und Aussagekraft zu legen!

Metadaten zur Beschreibung von Daten, Diensten und Anwendungen sind nur sinnvoll, wenn sie den potentiellen Nutzer zu diesen Ressourcen hinführen. Die Informationen müssen gefunden und verstanden werden. Dies ist in einem fachlichen und techniklastigen Umfeld, das durch viele - für Insider selbstverständliche - Fachbegriffe geprägt ist, die eigentliche Herausforderung: Eine Sprache und Wortwahl zu finden, die einerseits die Ressource aus fachlicher Sicht korrekt beschreibt, andererseits aber nicht überladen ist mit Fachbegriffen, die der Laie nicht versteht und nicht einsortieren kann. Letztlich sollen Metadaten dazu dienen, eine Entscheidung zu fällen: „Ist die hier beschriebene Ressource das, was ich gesucht habe?“

Metadaten müssen fachlich aktuell und stets inhaltlich richtig und zutreffend sein! Sie bedürfen einer laufenden Überprüfung und müssen bei relevanten Änderungen aktualisiert werden.

Siehe auch Vortrag von Peter Kochmann zur Daten-Dienste-Koppelung



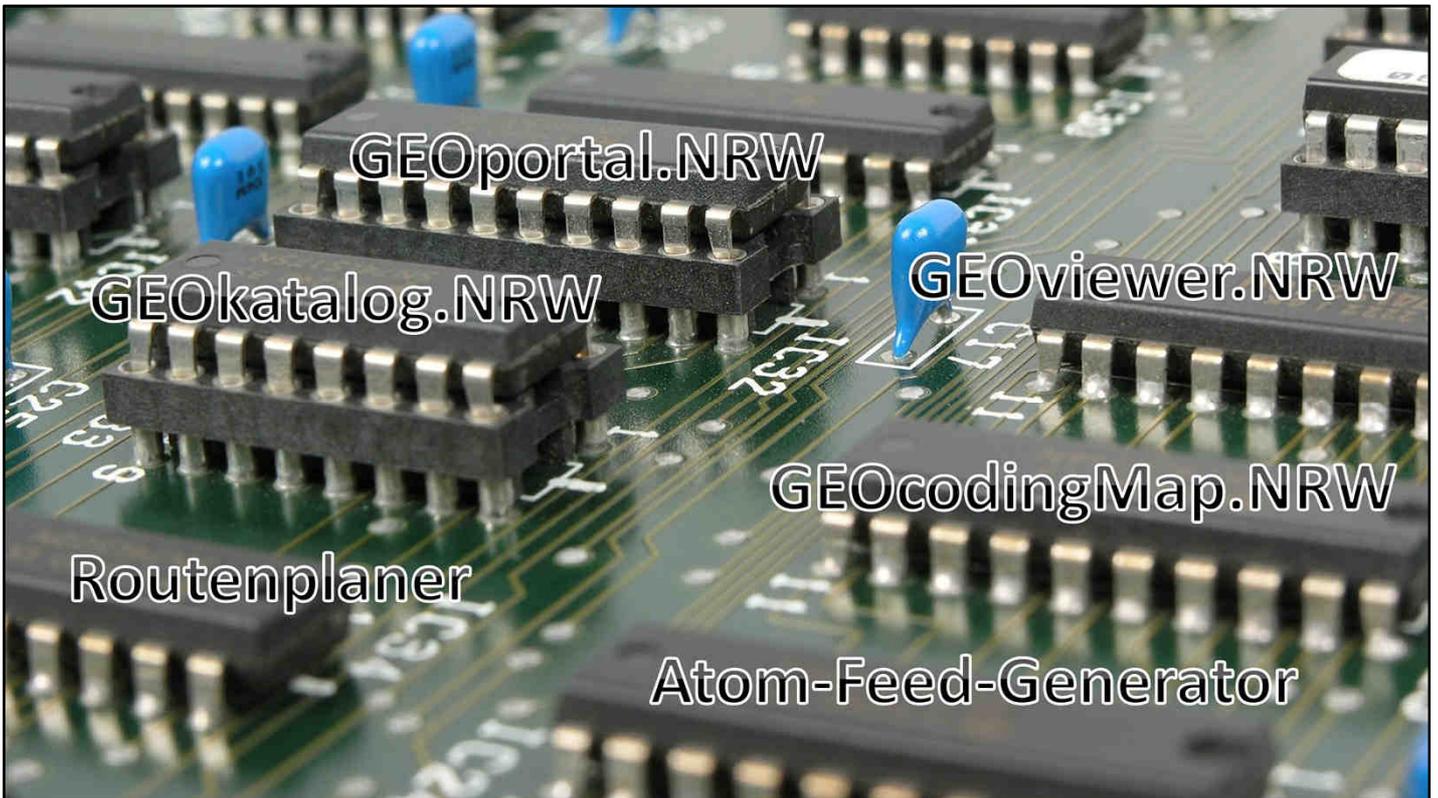
Qualität ergibt sich aus dem **Überdeckungsgrad der drei Zielgrößen „Nutzerforderungen“ (Sollen), „Ausrichtung der Datenbereitsteller“ (Wollen) und „Fähigkeit der Datenbereitsteller“ (Können).** Nutzeranforderungen sind z. B. Forderungen nach spezifischen Funktionalitäten oder Eigenschaften, die z. B. ein Produkt (Produktqualität) oder eine Dienstleistung (Dienstleistungsqualität) aus Sicht des Nutzers erfüllen soll. In der Ausrichtung der Datenbereitsteller spiegeln sich die Ziele und die strategische Ausrichtung des Datenbereitstellers wider. Die Ausrichtung der Datenbereitsteller definiert, wie ein Datenbereitsteller den Markt und damit die Nutzer bedienen möchte. Dabei orientiert sich der Datenbereitsteller an den gesellschaftlichen sowie selbst auferlegten Werten. Unter der Fähigkeit des Datenbereitstellers werden die Kompetenzen verstanden, die der Datenbereitsteller im Rahmen des Qualitätsmanagements besitzt, um die gesetzten Ziele zu erreichen und die Nutzerforderungen umzusetzen.

2019 NRW Planung: Zur Erhöhung des Deckungsgrades von Nutzeranforderungen (Sollen), Ausrichtung der Datenbereitsteller (Wollen) und Fähigkeit der Datenbereitsteller (Können) werden **Handreichungen zum Qualitätsmanagement für NRW erstellt.** Der Einsatz der neuen GDI-DE Testsuite wird dabei eine Rolle spielen, ebenso ein anzustrebender hoher Automationsgrad.

zentrale Komponenten GDI NW



Unterstützung durch die **zentrale Komponenten der GDI NW**: sie sind **robust, einfach, und wirkungsvoll**. Die zentralen technischen Komponenten der GDI-NW sind Serviceangebote des Landes und werden im Auftrag des Innenministeriums durch IT.NRW in Zusammenarbeit mit der GSt. des IMA GDI.NRW betrieben. **Sie können von allen geodatenhaltenden Stellen bei der Bereitstellung eigener Geodaten, Dienste und Anwendungen genutzt werden.**



Zu der GDI-NW Komponenten gehören:

Das **GEOportal.NRW** als zentrale Vermittlungsstelle zwischen Nutzern und Anbietern von Geodaten in NRW. Weiterhin dient es als Informationsplattform hinsichtlich des aktuellen Ausbaus der GDI-NW und der Umsetzung der europäischen INSPIRE Richtlinie.

Der **GEOviewer.NRW** ist die zentrale Komponente, um die bereitgestellten, öffentlich verfügbaren Daten und Dienste zu visualisieren. Der GEOviewer.NRW kann in eigene Anwendungen integriert werden. Der Geoviewer kann mittels Parameter (u.a. Hintergrundkarten, Kartenausmaße) individualisiert aufgerufen werden.

Der **GEOkatalog.NRW** s. Folie 2

Die **GEOcodingMap.NRW** ist eine einfache Anwendung, um in eigenen Webseiten grundlegende Kartendienste (meist topographische Karten als „klassischer Hintergrund“) der GDI-NW und GDI-DE einzubinden. Es genügen wenige Zeilen vorgegebener und bei Bedarf anzupassender Html-Quellcodes. Die GEOcodingMap.NRW kommt für alle möglichen einfachen Standort-, Dienststellen- oder etwa Gebietsübersichten in Frage. Sie hat definierte Schnittstellen zum Aufruf der Map, zur Steuerung der Map und zur Rückgabe von Parametern an die einbettende Anwendung.

Der **Routenplaner** ist die Navigationskomponente der GDI-NW. Die Funktion generiert eine Routenbeschreibung mit Straßennamen und Navigationshinweisen und stellt die Ergebnisse im Kartenfenster dar. Die Technik basiert auf einer Anwendung von Straßen.NRW. Technisch müssen lediglich die Koordinaten für Start- und Ende des Routings gesendet und anschließend die als Webservice gelieferte Route in der eigenen Anwendung verarbeitet werden können. Um die Routingkomponente in eigene Anwendungen integrieren zu können, muss eine Nutzungsvereinbarung mit Straßen.NRW getroffen werden.

Mit dem **Atom-Feed-Generator** steht eine einfache Möglichkeit bereit, einen Downloaddienst für Geodaten (und/oder Sachdaten) zu erzeugen. Ein Atom-Feed kann sowohl gemäß der INSPIRE-Richtlinie als alternativer Downloaddienst als auch für die Bereitstellung offener Daten (Open Data) verwendet werden. Ein Atom-Feed enthält direkte Downloadlinks zu vordefinierten Datensätzen bzw. Downloadpaketen.

NEU:

Open Data DownloadClient

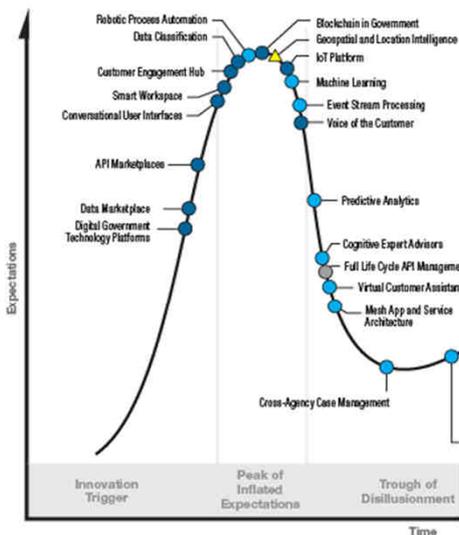
DOWNLOADING...

Die GDI-NW Komponenten haben Nachwuchs erhalten:

Den Open Data DownloadClient

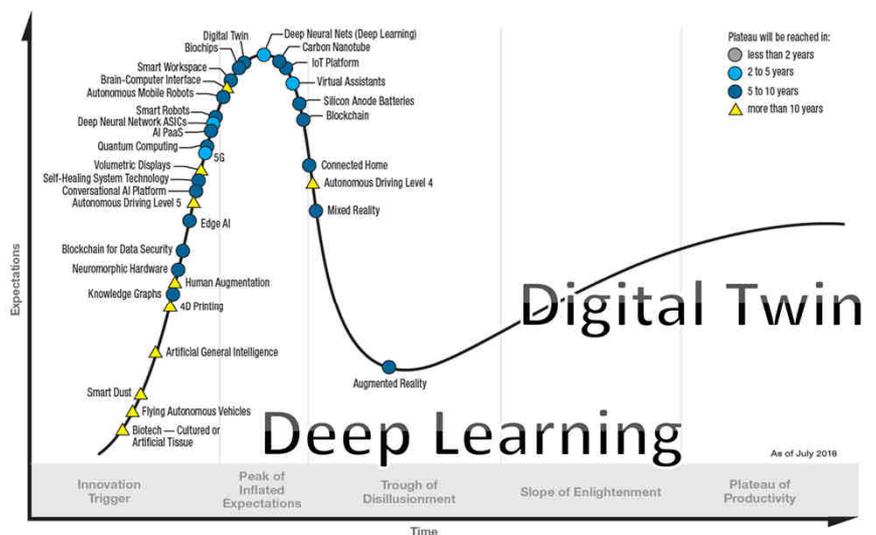
Mehr dazu im anschließenden Vortrag, den Annika Nockert halten wird.

Hype Cycle for Digital Government



gartner.com/SmarterWithGartner

Hype Cycle for Emerging Technologies, 2018



gartner.com/SmarterWithGartner

Location Intelligence

8) Its affiliates. All rights reserved.

Gartner

Auf dem Erreichten kann die GDI NW sich nicht ausruhen. Denn das Umfeld entwickelt sich ständig. Frage: Welcher Hype bekommt langfristige Bedeutung? 3D-Fernseher (Nein)? Augmented Reality (eher ja)? Autonomes Fahren (ja)? Aber welcher Entwicklung muss die GDI NW folgen? Die Gartner-Hype-Curve for Digital Government and for Emerging Technologies gibt Ausblick auf die Entwicklung von Trends. Die Kurve ist zu beobachten und abzuschätzen, welcher Trend besteht und Bedeutung für die GDI haben wird. Evtl. gewinnt künftig an Bedeutung:

a) Ein **Digitaler Zwilling** ist eine digitale Repräsentanz eines materiellen oder immateriellen Objekts aus der realen Welt. Es ist unerheblich, ob das Gegenstück in der realen Welt bereits existiert oder zukünftig erst existieren wird. Digitale Zwillinge ermöglichen einen übergreifenden Datenaustausch. (Beispiel: Gebäude/BIM) => *AUSBAUEN, da Basisinformationen schon in GDI vorhanden.*

b) Eine der häufigsten Techniken in der künstlichen Intelligenz ist [maschinelles Lernen](#). Maschinelles Lernen ist ein selbstadaptiver Algorithmus. **Deep Learning**, eine Teilmenge des maschinellen Lernens, nutzt eine Reihe hierarchischer Schichten bzw. eine Hierarchie von Konzepten, um den Prozess des maschinellen Lernens durchzuführen. Künstliche neuronale Netze verarbeiten dabei eine Rohdateneingabe, wie beispielsweise die einzelnen Pixel eines Bildes => für Bild-/Objekterkennung. *BEOBACHTEN, da Auswirkung auf unsere Community zu erwarten sind:*

Location Intelligence nutzt Open Source bzw. Open Data und beruht auf Unternehmen, die eine größere Vielfalt dynamischer Datenströme für ihre Visualisierung und Analyse verwenden als ausschließlich statische Informationen zu einem bestimmten Zeitpunkt. GIS stützt sich in erster Linie auf proprietäre geografische Datensätze, die Unternehmen/Institutionen gehören.

Dynamische Daten: Standort des mobilen Geräts, Daten des IoT-Geräts, Fußverkehr, Fahrzeug / Transport, Live-Transit, Aufnahme / Absetzung von Taxireisenden, aktuelle Wetterbedingungen

Statische Daten: Topographie, Historische Wettertrends, Bevölkerungsstatistik, Wirtschaft

Beispiele: a) Versicherer nutzen Standortinformationen und offene Wetterdaten, um in Echtzeit

Entscheidungen zu treffen, wie sie ihre Versicherungsnehmer bei Naturkatastrophen unterstützen können

b) Immobilieninvestoren nutzen Standortinformationen und offene Transitdaten, um neue

Wachstumschancen zu evaluieren und c) Banken verwenden Kreditkartentransaktionen und demografische Daten, um die Dynamik der Stadt zu verstehen.



zeitgemäße Architektur

konsistent zum WWW

moderne Technologien

Fokus Web-Entwickler

Web-APIs

(ohne) Schemadefinition

Die GSt wird die Weiterentwicklung der GDI unterstützen und dazu erkannte Defizite analysieren (GDI nur für Experten, Geo-Objekte lassen sich nicht im Web verlinken, ungewohnte Formate, komplexe Abfragen, keine Indexierung durch Suchmaschinen) .

Alle bisherigen Versionen von WFS verwenden eine Architektur, die vor 20 Jahren dem Stand der Technik entsprach, heute aber als veraltet angesehen wird (dasselbe gilt auch für WMS, CSW, usw.).

Ein Standard wird benötigt, der

- eine zeitgemäße Architektur verwendet,
- konsistent mit der aktuellen Architektur des World Wide Web und den (Spatial) Data on the Web Best Practices ist,
- auf modernen Technologien (wie der OpenAPI Specification) aufsetzt,
- einen Fokus auf Software-Entwickler setzt,
- modulare Bausteine für den feingranularen Zugriff auf Geoobjekte für die Verwendung in Web-APIs definiert.
- (Authentifizierung und Zugriffsmechanismen über die Verwendung von OpenAPI bereits im Kern berücksichtigt und)
- nicht mehr zwingend die Definition von Schemata für die Geodaten fordert,

In der Konsequenz werden Implementierungen von WFS 3.0 nicht rückwärtskompatibel mit WFS-2.0-Implementierungen sein. Eine funktionale Kompatibilität ist jedoch sehr wohl ein Designziel von WFS 3.0. D.h., auch die neuen Implementierungen sollen die heutigen Anforderungen unterstützen.



Ein Baustein auf diesen Weg ist der Standard WFS 3.0.

Der Standard wird – mit Unterstützung des OGC Board of Directors – in einem etwas anderen Prozess als bislang entwickelt. Die Entwicklung des Standards erfolgt von Anfang an mit öffentlicher Konsensbildung unter Einbeziehung von Nicht-OGC-Mitgliedern. Außerdem wird die Zeitspanne zwischen dem Vorliegen eines vollständigen Entwurfs und einer Verabschiedung als Standard bewusst verlängert, um ausreichend Zeit für die Validierung durch Implementierungen und die Nutzung in der Praxis zu schaffen.

Im Fall von **WFS 3.0 Teil 1, Core**, wurde die erste Version von **September 2017 bis April 2018** erarbeitet, begleitet von mehreren Implementierungen der Entwürfe, unter anderem in einem Hackathon an dem über 30 Entwickler aus vier Kontinenten teilnahmen. Seitdem wird der Entwurf der Spezifikation in einer ganzen Reihe von Aktivitäten innerhalb und außerhalb des OGC implementiert und getestet. Es wird aktuell erwartet, dass der Entwurf Mitte 2019 vorliegt und anschließend in den formalen Abstimmungsprozess in OGC und ISO geht.

- **Unterstützung für weitere Koordinatenreferenzsysteme neben WGS**
- Beschränkung der zurückgelieferten Attribute pro Objekt, **bessere Selektionsmöglichkeiten der Objekte**
- Unterstützung weiterer Datenformate
- einfache und komplexe Transaktionen
- geometrische Vereinfachungen, Unterstützung für 3D
- **verbesserte Behandlung von Links zu anderen Daten**
- Unterstützung für PubSub (Subskription von Änderungsmitteilungen)
- Erweiterung der grundlegenden Architektur zur Unterstützung von weiteren Formen der Aufbereitung von Geoobjekten, z.B. **Karten** (ähnlich wie WMS), **Vector Tiles** (ähnlich wie WMTS, Mapbox Vector Tiles), 3D-Szenen (ähnlich wie 3DPS)

Geobasis NRW 2019:

5 ★ Open Data

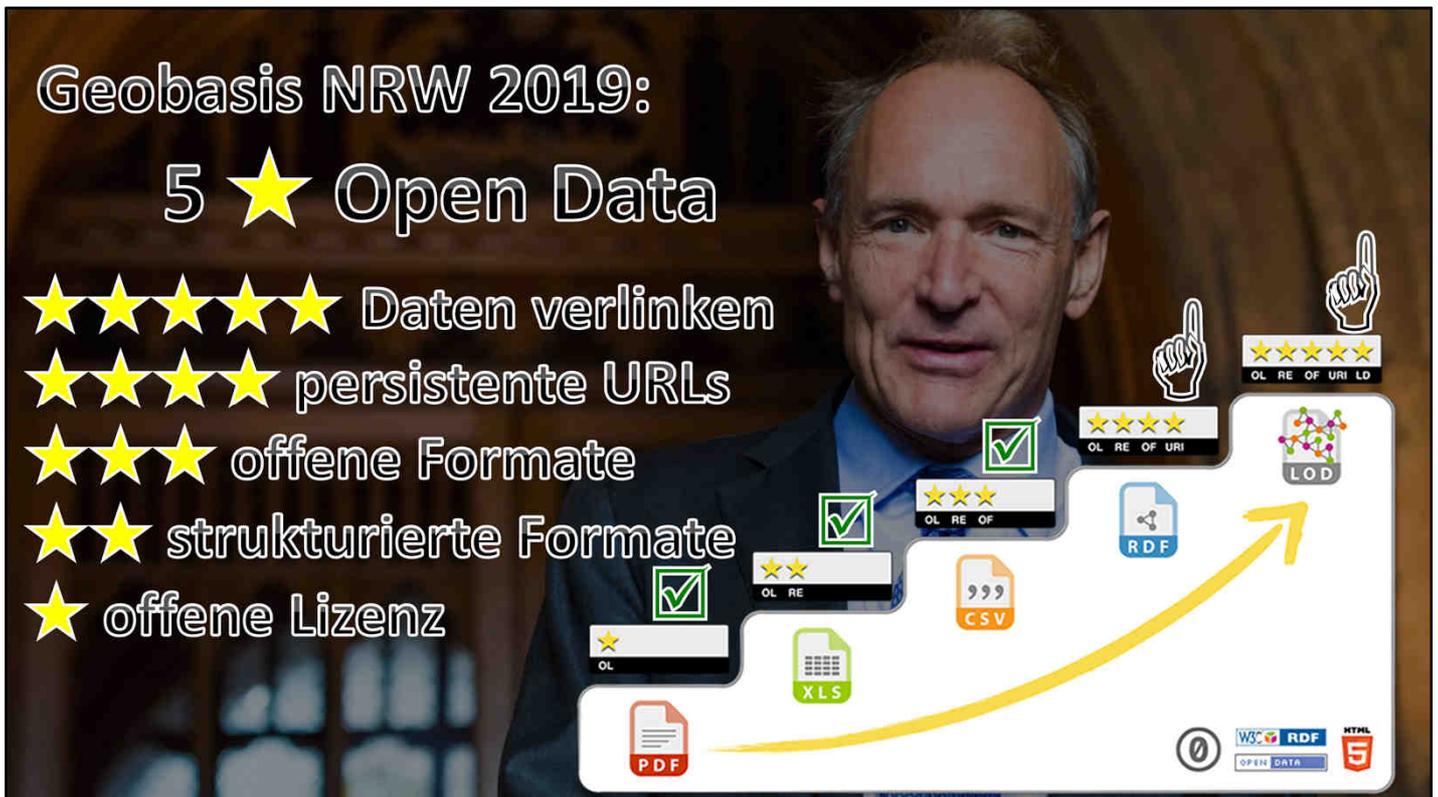
★★★★★ Daten verlinken

★★★★★ persistente URLs

★★★★ offene Formate

★★★ strukturierte Formate

★ offene Lizenz



Tim Berners-Lee, der Erfinder des World Wide Webs und Initiator von Linked Data schlug ein 5-Sterne-Modell für Offene Daten (Open Data) vor.

Die **Geobasisdaten in NRW sollen 2019 zu 5 star open data** (nach Tim Berners Lee, s. <https://5stardata.info/de/>) werden.

Realisiert sind schon ★ - ★★★★★:

- ★ die Daten im *Web* stehen unter einer offenen Lizenz bereit. (Das Format ist dabei egal)
 - ★★ die Daten stehen in einem strukturierten Format bereit (z. B. Excel anstelle eines eingescannten Bildes einer Tabelle)
 - ★★★ es werden offene, nicht proprietäre Formate verwendet (z. B. CSV statt Excel oder GML statt shape)
- In 2019 im Rahmen des Projektes Spatial Data on the Web werden die beiden fehlenden Sterne für die Geobasisdaten erarbeitet:

- ★★★★ [persistente URLs] (URIs) um Dinge zu bezeichnen, damit die Daten verlinkt werden können
- ★★★★★ Daten werden mit anderen Daten verlinkt um Kontexte herzustellen

Das wird zu Netzwerkeffekten führen, von denen Nutzer und Bereitsteller profitieren.

Was sind die Vorteile einer Veröffentlichung von Daten mit ★★★★★?

- ✓ während der Nutzung der Daten können weitere (ähnliche) Daten entdeckt werden (z.B. den links in Wikipedia folgen) und
 - ✓ der Nutzer kann direkt etwas über das Datenschema lernen.
 - ✓ als Bereitsteller macht man die Daten auffindbar,
 - ✓ erhöht den Wert der Daten und
 - ✓ erhält dessen Organisation die gleichen Vorteile von den Verlinkungen wie die Nutzer
- evtl. Nachteile:

! Man muss sich ggf. mit toten Links (z. B. 404-Fehler in Webseiten) befassen.

! Daten von einem beliebigen Link als Tatsache zu präsentieren ist so riskant, wie Personen zu erlauben, Inhalt von jeder Webseite in die eigene Seite einzufügen. Achtung! Vertrauen und ein gesunder Menschenverstand sind nach wie vor notwendig.

! in die Verlinkung der Daten zu anderen Daten im *Web* ist Zeit zu investieren und

! gegebenenfalls muss man tote oder fehlerhafte Links reparieren.



Der Begriff Schnitzelmitkartoffelsalat ist ein Testbegriff von Webmastern, um die Eigenschaften von Suchmaschinen zu erforschen. Ursprünglich sollte damit die Zeitdauer vom Besuch des Robots bis zur Aufnahme in den Index der Suchmaschine gemessen werden. Schnitzelmitkartoffelsalat war dabei ein gewählter Begriff, der zu diesem Zeitpunkt bei Google noch nicht im Index vertreten war.

Das Ziel wurde erweitert zur Untersuchung, wie Google verschiedene HTML-Konstruktionen verarbeitet und welche Ergebnisse geliefert werden. Für die ursprüngliche Zielsetzung ist der Begriff mittlerweile ungeeignet, da er auf Hunderten von Seiten zu finden ist. **Schnitzelmitkartoffelsalat ist inzwischen ein Synonym für Zeichenketten, die noch nicht in Suchmaschinen zu finden sind.**

Von besonderem Interesse ist dabei, wie die verschiedenen Algorithmen der Suchmaschinen funktionieren und wie die Robots die Seiten indexieren. Die Erkenntnisse daraus können helfen, Webseiten für solche Suchmaschinen zu optimieren.

Vor knapp 2 Jahren hatte Google bekannt gegeben, dass sie an der Indizierung von Datensätzen arbeiten, auf der Basis der schema.org/DCAT-Spezifikationen. Seit kurzem bietet Google nun den „Google Dataset Search“ an, sodass die indizierten Datensätze auch gefunden werden können. Etwaige ergänzende Metadatenelemente aus den DCAT-Erweiterungen DCAT-AP, DCAT-AP.de oder GeoDCAT-AP werden hierbei aktuell – soweit bekannt – nicht verarbeitet.

Sucht man zum Beispiel nach allen Datensätzen aus der Internetdomain „nrw.de“, dann werden die Datensätze aus SDW-Projekten gefunden, da die verwendete Software die Datensätze in HTML mit schema.org-Annotationen beschreibt.

HTML enthalten heute nicht mehr nur Text, sondern Daten, die von Suchmaschinen und anderen Werkzeugen verarbeitet werden.

NRW wird sich an der Suchmaschineoptimierung mit dem Ziel beteiligen, die Datensätze und auch einzelne für Nutzer bedeutsame Objektinstanzen „google-bar“ zu machen.

Raus aus dem darknet mit unseren Geodaten!

5 ★ Open Data

Datenqualität

Suchmaschinenoptimierung

Raus aus dem darknet mit unseren Geodaten.

- 5-star open data
- Datenqualität
- Suchmaschineoptimierung



Vielen Dank für Ihre Aufmerksamkeit!

Burkhard Schlegel

Gst. IMA GDI.NRW

Tel.: 0221 / 147 3541

E-Mail: burkhard.schlegel@bezreg-koeln.nrw.de